

TDK-dolgozat

Dalotti Ágoston
2023.



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

Depresszió súlyosságát becslő beszédalapú modell optimalizálása többféle érzelmi állapotot imitált beszédminták felhasználásával

Készítette:

Dalotti Ágoston
IV. évfolyam

Konzulens:

Dr. Kiss Gábor,
egyetemi tanár

2023.

Tartalomjegyzék

Kivonat	i
Abstract	ii
1. Bevezetés	1
2. Depresszió	3
2.1. Meghatározás	3
2.2. Diagnosztizálás	4
2.3. Beszéd, mint objektív biomarker	4
2.4. BDI-II - Beck Depression Inventory-II	5
3. Érzeti beszédjellemzők számszerű leírása - akusztikai-fonetikai jellemzők	6
3.1. Beszédjellemzők megváltozása depresszió hatására	9
4. Statisztikai fogalmak, és gépi tanuló modellek	10
4.1. Párosított t-próba	10
4.2. Szupport Vektor Gépek (SVM)	11
4.2.1. Lineárisan szeparálható eset	11
4.2.2. Lineárisan nem szeparálható eset	12
4.3. Szupport Vektor Regresszió (SVR)	13
4.4. Regresszió pontosságának leírására alkalmazott jellemzők	14
4.5. Keresztvalidáció	15
4.6. Fast Forward Selection	16
5. Felhasznált és létrehozott adatbázisok	17
5.1. Magyar Depressziós Beszédatadtbázis	17
5.1.1. Releváns információk az MDB adatbázisról	18
5.2. Magyar Depressziós Imitációs Beszédatadtbázis	19
5.2.1. Releváns információk az MDIB adatbázisról	20
5.2.2. Az érzelmek imitálásáról	20
6. Depression State Assessment	22
6.1. DSA jellemzők	22
6.2. DSA tesztelése az MDIB mintáival	23
6.2.1. Saját BDI értékek párosítása a depressziót imitált mintákhoz .	23
6.2.2. Kontroll minták párosítása a depressziót imitált mintákhoz . .	25

6.2.3. Kontroll minták párosítása különböző hangulatokat imitált mintákhoz	26
7. Modellépítés	27
7.1. Jellemzőkinyerés és előfeldolgozás	27
7.2. Modell realizálása	28
8. A modell tesztelése és eredmények	29
8.1. Modell tesztelése az MDB segítségével	29
8.2. A modell tesztelése az MDIB segítségével	30
8.2.1. Saját BDI értékek párosítása a depressziót imitált mintákhoz .	31
8.2.2. Kontroll minták párosítása a depressziót imitált mintákhoz . .	33
8.2.3. Kontroll minták párosítása különböző hangulatokat imitált mintákhoz	34
9. Összefoglalás	36
Köszönetnyilvánítás	38

Kivonat

Az emberi kommunikáció legáltalánosabb eszköze a beszéd. A legtöbb ember számára nem okoz gondot egy beszédproduktum megalkotása. Viszont a beszéd, és a hangkeltés működésének megértésével belátható, hogy egy rendkívül összetett folyamatról van szó. Az összetettség olyan magasfokú, hogy kijelenthető: minden beszédproduktum egyedi. Éppen ezért a beszéd az ember személyiségjegye, amely akár beszélőazonosításra is alkalmas.

Ennek ellenére megfigyelhető, hogy mind érzelmi hatások, mind fiziológiai és pszichiátriai tényezők hatással lehetnek a keletkezett beszédproduktumra.

A depresszió korunk egyik legmeghatározóbb gyógyítható betegsége, amely jelentősen befolyásolja a benne szenvedő betegek életminőségét, szomatikus panaszok mellett komoly kognitív, viselkedésbeli és motivációs tünetek jelentkeznek.

A modern pszichológia kezdeteitől találhatóak feljegyzések, miszerint a depresszió hatására az egyén beszédproduktuma megváltozik. A technika fejlődésével komoly kutatási területté nőtte ki magát a depresszió beszéd alapján történő, automatizált felismerése.

Dolgozatomban kifejtem, hogy bizonyos érzelmek mekkora hatással lehetnek a beszédalapú depresszió súlyosságát becslő modellek döntésére. Továbbá megvizsgálom, hogy szándékos depresszió imitálással létrehozható-e olyan beszédproduktum, amely jól közelíti a valós depresszió hatása alatt keletkezett beszédproduktumot.

Magyar nyelv esetében még nem vizsgálták, hogy a depresszió, vagy érzelmek imitálása mennyiben befolyásolja az automatikus felismerés pontosságát, így a feladatom részét képezte egy - a vizsgálati céljaimnak megfelelő - adatbázis létrehozása is.

Dolgozatomban bemutatott eredmények oda vezettek, hogy saját - optimalizált - modellt fejlesszek, amely beszédelemzés által képes az imitált és a valós depresszió különválasztására, továbbá ami képes csökkenteni a beszéd alapú depresszió súlyosságát becslő modellek döntésének hangulatfüggését.

Abstract

The most common means of human communication is speech. For most people, creating a speech product is not a problem. However, by understanding the functioning of speech and vocalization, it can be seen that it is an extremely complex process. The complexity is so high that it can be stated that every speech product is unique. Therefore, speech is a characteristic of a person, which can even be used for speaker identification.

However, it can be observed that both emotional influences and physiological and psychiatric factors can affect the resulting speech product.

Depression is one of the most prevalent curable diseases of our time, which significantly impacting the quality of life of those suffering from it. Alongside somatic complaints, it presents with serious cognitive, behavioral, and motivational symptoms.

There have been records since the beginnings of modern psychology indicating that depression leads to changes in an individual's speech production. With the advancement of technology, the automated recognition of depression based on speech has grown into a serious field of research.

In my thesis, I elaborate on how certain emotions can significantly influence the decisions of speech-based models estimating the severity of depression. Furthermore, I examine whether intentional imitation of depression can create a speech product that closely approximates the effects of speech produced under genuine depressive conditions.

In the case of the Hungarian language, it has not yet been examined how the imitation of depression or emotions affects the accuracy of automatic recognition. Therefore, creating a database that aligns with my research objectives was part of my task.

The results presented in my thesis led me to develop my own - optimized - model, which is capable of separating simulated and real depression through speech analysis, and which is also able to reduce the mood dependence of the decision of speech-based depression severity models.

1. fejezet

Bevezetés

A depresszió egy komplex, rengeteg embert sújtó betegség. A WHO (World Health Organization) statisztikái szerint a felnőttek közül minden 25 ember között van egy depressziós[1]. Ez gyakorlatilag azt jelenti, hogy minden felnőtt - nem túl távoli - környezetében van olyan személy, aki depresszióban szenved. A depresszió hatással van az általa sújtott egyén életminőségére, munkavégző képességére. Így az emberi, empátikus hozzáálláson túl, a társadalmi felelősségvállalás is szükséges, hiszen a betegség globális szinten súlyos gazdasági károkat okozhat[2].

A betegség kialakulása során a korai stádiumban való reagálás kulcsfontosságú lehet, súlyos következmények elkerülhetők általa[3].

Az empirikus kutatások alapján bebizonyosodott, hogy a depresszió szignifikánsan befolyásolja az egyén beszédproduktumát. Ezen változásokat informatikai eszközökkel mérni és számszerűsíteni tudjuk. A különféle beszédleíró jellemzők elemzésével különböző beszélők beszédproduktumai összehasonlíthatókká válnak. Ezáltal lehetőség nyílik a depresszió gépi alapú automatikus felismerésére, a depresszióval küzdő személyek korai azonosítására[4].

A beszédfeldolgozás, jellemzőkinyerés és osztályozás során gyakran használnak gépi tanuló algoritmusokat. Az ilyen modellek alkalmazása jelentősen megkönnyíti és hatékonyabbá teszi a depresszió felismerését.

Az elmúlt években számos kutatás foglalkozott a témával[4], de ezek közül egy sem tért ki az algoritmusok döntésének hangulatfüggésére, illetve becslésszerűségére. Így kutatásom során célkitűzésemmé vált, hogy körüljárjam ezt a kérdést.

Kutatásom kezdetén az egyetem Beszédakusztikai Laboratóriuma rendelkezésemre bocsátotta a magyar nyelvű depressziós beszédadatbázisát, melyben depressziós és egészséges személyektől származó felvételek is vannak. Úgyszintén rendelkezésemre bocsátottak egy beszéd alapján depresszió súlyosságát prediktáló alkalmazást, és

egy beszédjellemező kinyerésre felhasználható alkalmazást.

Dolgozatomban bemutatott eredményekkel a depresszió beszédre gyakorolt hatásának mélyebb megértését szeretném segíteni. Kifejtem a téma megértéséhez szükséges elméleti háttérrel, bemutatok egy új, általam rögzített beszédadatbázist, amely segítségével rávilágítok a jelenlegi modellek hiányosságaira. Továbbá bemutatok egy saját, optimalizált modellt, amely képes a döntés hangulatfüggését minimalizálni, illetve képes megkülönböztetni az imitált és a valós depresszió során keletkezett beszédproduktumot.

Dolgozatomban az alábbi hipotézisem volt: *Az imitáció nincs hatással a depresszió súlyosságát beszéd alapján becsülő modellek működésére.*

2. fejezet

Depresszió

A depresszió¹ korunk egyik legmeghatározóbb gyógyítható betegsége. A WHO (World Health Organization) statisztikái szerint a teljes populáció 3,8% -át, míg a felnőttek 5% -át érinti[1]. Továbbá predikciójuk szerint az unipoláris depresszió 2030-ra benne lesz a 3, világszinten legtöbb embert érintő betegségben. (A HIV/AIDS és a szívproblémák mellett)[5].

A depresszió jelentősen befolyásolja az ebben szenvedő betegek életminőségét, szomatikus panaszok mellett komoly kognitív, viselkedésbeli és motivációs tünetek jelentkeznek. A korai tünetek viszont annyira különfélék lehetnek, hogy a családorvoshoz (házi orvoshoz) forduló betegek ritkán kapnak megfelelő diagnózist. (Amennyiben az orvos a szomatikus tünetek mentén kezdi feltárni a problémát ez az arány 10%[6].) A depresszió diagnosztizálására csak egy szűk, magasan képzett orvosi-pszichológusi réteg képes[7][8].

2.1. Meghatározás

A depresszió kialakulását, és lefutását tekintve igen komplex betegség. Általánosan elfogadott meghatározását a The Diagnostic and Statistical Manual of Mental Disorders (DSM)-ben olvashatjuk[9].

Depresszió akkor áll fent, hogyha az általános kedvtelenségen, levertségen túl a felsorolt tünetek közül négy (vagy több) legalább két hétig fenn áll:

- Testi tünetek: pszichomotoros gátlás, vagy remegés, jelentős hízás vagy fogyás, alvászavar (insomnia, vagy hypersomnia), fejfájás, szédülés, ájulás, fáradékonyság

¹Major Depressive Disorder - továbbiakban depresszió

- Kognitív tünetek: Negatív én- és világkép, csökkent kognitív képességek, pesszimizmus, alaptalan túlzott bűntudat, gyakori halál gondolatok, öngyilkos gondolatok
- Viselkedés tünetek: alacsony aktivitás, és produktivitás, görnyedt testtartás, és halk beszéd
- Motivációs tünetek: érdeklődés, és kedv csökkenése
- Érzelmi tünetek: szomorúság, rosszkedv, örömgýűlölet, szeretetképtelenség

A felsorolt tünetekbe belevettem a Semmelweis Egyetem Pszichoterápiás Osztály által közölt tüneteket is[10].

2.2. Diagnosztizálás

A depresszió diagnosztizálására jelenleg csak olyan objektív biomarkereket ismerünk (alacsony szerotonin szint, neurotranszmitter diszfunkció, genetikai rendellenességek), amelyek kinyerésére jelenleg igen körülményes, invazív beavatkozást igényelő módszerek léteznek. Ezek a beteg számára igen kellemetlenek lehetnek[11][12].

A 21. században egyre populárisabbá vált a nem-invazív beavatkozást igényelő módszer, ebbe tartozik például az arcon megfigyelhető változások vizsgálata illetve a jelenleg egyre elterjedtebb beszéd alapú diagnosztizálás[13].

Már a 90'-es években is voltak korai kutatások, amelyek igazolták, hogy a páciens beszéde mint nonverbális kifejező eszköz használható a depresszió felismerésére[14][15]. Itt már említik a hosszabb szüneteket, kisebb artikulációt, lassabb beszédet, a monotonitást. Valamint a kevesebb szemkontaktust mind beszéd, mind hallgatás esetében.

A depressziós beszéd-adatbázisok megjelenésével ezen, és további beszédjellemzők összehasonlítására nyílt lehetőség a mérnöki eszköztár módszereivel.

2.3. Beszéd, mint objektív biomarker

A beszédkeltés egy igen komplex folyamat, rengeteg izmot kell megfelelően kontrollálni, hogy összefüggő beszédproduktum keletkezessen, amely a nyelvi kód továbbítására alkalmas. Többek között ezen izmok közé tartoznak a légző, a gége, az artikulációs izomcsoportok[16].

Depresszió hatással van az izmok feszítettségére és kontrollálhatóságára, így közvetlen hatással bír a megalkotott beszédproduktumra. Magas korreláció mutatható ki bizonyos beszédjellemzők és a depresszió súlyossága között [13][17].

Továbbá, ahogy a 2.1 fejezetben olvasható, a depresszió hatással van a beteg kognitív képességeire is, ezáltal sérül a fonológiai hurok, ami az artikulációs rendszer kontrollálásáért felel. A beteg artikulációs és fonációs hibákat ejt[18][19]. Emellett megfigyelhető, hogy a depresszióban szenvedő emberek beszéde lassabb, és hosszabb szüneteket tartanak, ami arra enged következtetni, hogy nehezebbé válik számukra a beszéd megtervezése is[13][20].

2.4. BDI-II - Beck Depression Inventory-II

A BDI-II (továbbiakban BDI) olyan tünetek klinikai megfigyelésén alapul, amelyek depressziós személyeknél gyakran, míg egészséges személyeknél ritkábban jelentkeznek.

A 21 tünet: a) hangulat, b) pesszimizmus, c) kudarc érzése, d) elégedettség hiánya, e) bűntudat, f) büntetés érzése, g) önutálat, h) önvád, i) öngyilkos gondolatok, j) sírás, k) ingerlékenység, l) szociális visszahúzódás, m) döntésképtelenség, n) testkép torzulása, o) munkagátlás, p) alvászavar, q) fáradékonyság, r) étvágytalanság, s) fogyás, t) szomatikus lefoglaltság, u) szexuális vágy elvesztése

Minden tünetet 0-3-ig kell értékelni, az intenzitás függvényében[21].

Az egyes tünetekre adott értékek összegzéséből alakul ki a BDI érték.

A határokat irodalomtól függően egy kicsit máshol találjuk, én az alábbiak szerint dolgoztam[53]:

- 0-13: minimális depresszió (egészséges)
- 14-19: enyhe depresszió
- 20-28: mérsékelt depresszió
- 29-63: súlyos depresszió

A BDI egy elfogadható, és nemzetközileg használt skála a depresszió súlyosságának becslésére. Kimutatható, hogy felhasználásával a gépi tanuló modellek pontosabb becslést képesek adni a bemondó betegségének² súlyosságát illetően. A teszt egyszerű felvétele lehetővé teszi a depressziós beszédatabázisok gyors növelését, ami elengedhetetlen a jelenlegi modellek pontosságának fejlesztéséhez[23].

Azt Beck is megemlíti, hogy a BDI teszt elcsalható, a kitöltő lelkiismeretességétől függhet[24].

²Depresszió

3. fejezet

Érzeti beszédjellemzők számszerű leírása - akusztikai-fonetikai jellemzők

A 2 fejezetben említett jellemzők számszerű leírása már komoly mérnöki feladatnak bizonyul. Ebben a fejezetben azon jellemzőket mutatom be, amelyek (a feltárt szakirodalom szerint) megváltoznak depresszió hatására, és egymással nem, vagy kevésbé korrelálnak.

Beszédkeltés során a tüdő biztosítja a kifelé irányuló levegőáramot, majd a gége működés során dől el, hogy zöngés, vagy nem zöngés hangot ejtünk ki. Zöngéhang esetén egy kváziperiódikus levegőnyomás-változásról beszélünk, amely a hallási rendszerünknek hallható frekvenciával rendelkezik[25].

Alapfrekvencia f_0 [Hz]:

A zöngés hang periódusidejének reciproka, ez határozza meg a beszélő hangfekvését. Értéke a beszédkeltés során permanensen változik, ez adja a beszéd dallamosságát. A beszélő széles tartományban tudja változtatni az alapfrekvenciát, és a (beszélő) neme jó közelítéssel meghatározható az átlagértékéből. Férfiak esetében ez 100 Hz, nők esetében 180 Hz, míg gyerekek esetében akár több száz Hz is lehet[25].

Jitter [%]:

Az alapfrekvencia rövid elemzési ablakban számított ingadozása a hang rekedtes érzetével van összefüggésben. Ezt az értéket (az f_0 kis elemzési ablakban számított ingadozását) nevezzük Jitter-nek[29].

A Praat program¹ "Get Jitter" funkciója az alábbi képlettel dolgozik[30]:

$$jitter_{ddp}[\%] = \frac{\frac{\sum_{i=2}^N |(T_{i+1}-T_i)-(T_i-T_{i-1})|}{N-2}}{\frac{\sum_{i=1}^N T_i}{N}} \cdot 100, \quad (3.1)$$

ahol a T_i az i -dik periódus hosszát jelöli (másodpercben), az N pedig az egymást követő vizsgált periódusok számát jelöli ($N > 2$).

Shimmer [%]:

Az alapfrekvencia egymást követő periódusainak százalékos amplitúdóingadozása. A Jitter-hez hasonlóan ez az ingadozás is a beszéd rekedtes érzetével van összefüggésben[29].

A Praat program "Get Shimmer" funkciója az alábbi képlettel dolgozik[13][32]:

$$shimmer_{ddp}[\%] = \frac{\frac{\sum_{i=2}^N |(A_{i+1}-A_i)-(A_i-A_{i-1})|}{N-2}}{\frac{\sum_{i=1}^N A_i}{N}} \cdot 100, \quad (3.2)$$

ahol a A_i az i -dik periódus amplitúdóját jelöli, az N pedig az egymást követő vizsgált periódusok számát jelöli ($N > 2$).

Formánsfrekvenciák F1, F2, ... [Hz]:

Fonáció, vagyis hangszalag rezgés esetén az előállított zöngé gerjeszti az artikulációs csatornát. A zöngé felbontható alapfrekvenciára, és annak egész számú többszöröseire, a felharmónikusokra, egészséges esetben akár 5000 Hz-ig. A zöngé spektrumában, ezen felhangok amplitúdó értékeiben általában 12 dB/oktáv csökkenés figyelhető meg. Legnagyobb amplitúdóval az alapfrekvencia rendelkezik.

Ez a felharmónikus szerkezet az artikulációs csatornába jut, ahol a felharmónikusok bizonyos csoportjában (és környékén) erősítés figyelhető meg. A spektrumban mérhető ilyen felerősödött felhang csoportokat (a fonetikai szakirodalomban) formánsnak nevezik. Az adott formáns frekvenciáján a rezonanciafrekvencia tartomány burkológörbéjének amplitúdó-maximumát értjük. Minden zöngés hang egyértelműen meghatározott a formánsai által²[33].

Formánsfrekvenciák sáv szélessége B1, B2, ... [Hz]:

A formánsokhoz definiálnak formánssáv szélességet is, amelyek a rezonanciafrekvencia burkoló-görbe maximumától visszaszámolt -3 dB-es pontok közötti frekvenciatávolság. Ezeket B1, B2, stb-vel jelölik. Minél magasabb a formáns sorszáma, annál távolabb van az alapfrekvenciától, és annál szélesebb a sáv szélessége[33].

A normális beszédhez viszonyított megnövekedett formáns sáv szélesség értékek akkor alakulnak ki, amikor a formánsfrekvenciák erősítése nem megfelelő. Ez artikulá-

¹nemzetközileg elterjedt beszédelemző program

²általában 4-5 formáns mérhető

ciós beszédhibának számít.

Intenzitás [dB]:

A hangintenzitás (I) definíció szerint egységnyi felületen merőlegesen időegység alatt átáramlott hangenergia [W/m²], ezzel szemben az intenzitást az alábbi képlet határozza meg ($I_{ref} = 10^{-12}$ [W/m²] - hallásküszöb)[13]:

$$intenzitas[dB] = 10 \cdot \log_{10}\left(\frac{I}{I_{ref}}\right) \quad (3.3)$$

Az intenzitás értéke a hangerősség érzetével áll szoros kapcsolatban, és első sorban a zöngés szakaszokra határozható meg[33].

Mel-sávós energiaértékek [dB]:

Az érzeti hangmagasság mértékegysége a mel. Ez egy pszichofizikai hangmagasság-skála. Létezősége kísérleti alapokon fekszik. Kimutatható, hogy az 1000 Hz érzésre 0 Hz és 4000 Hz között félúton van. Ennek következtében merült fel az igény egy olyan skálára, amely értékeinek duplázódása a hangmagasságérzetet is duplázza[34].

A 131 Hz megfelel 131 melnek. A frekvencia értékek mel-skálába történő átszámolása a következő képlettel számítható[13]:

$$x[mel] = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (3.4)$$

Mel frekvenciás kepsztrum együtthatók - MFCC:

Egy beszédjelből a beszédjelet leíró mel-sávós energia értékeket megkapjuk, ha a beszédjel egy rövid szeletét Fourier-transzformáljuk, majd rajta szűrősoros elemzést végzünk el, azaz az összetevőket mel-sávok szerint összegezzük. Az így kapott összegekből áll össze a beszédjel mel sprektuma. Ezen összegek logaritmusának diszkrét koszinusz transzformációját véve kapjuk meg az MFC-együtthatókat[35][36]:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot \cos\left(\frac{\pi}{N} \cdot (n + 0,5) \cdot k\right) \quad (3.5)$$

ahol X_k a k -adik transzformált MFC-együttható ($k = 0, 1, \dots$), x_n az n -edik eredeti együttható, N a szűrők száma.

Artikulációs sebesség [hang/s]:

Ez a fogalom azt fejezi ki, hogy milyen gyors az artikuláció, időegység alatt hány beszédhangot ejtünk. Értéke nagyban függ a beszéd típusától, illetve a beszédtervezés sebességétől. Mivel kifejezetten az artikulációra fordított időre vonatkozik, mérése során a beszédben tartott szüneteket nem veszik figyelembe. A magyar beszéd átlagos artikulációs sebessége 12-14 hang/s[37].

Jellemző	Változás iránya
Alapfrekvencia	↓
Első formánsfrekvencia átlagos értéke	↓
Második formánsfrekvencia átlagos értéke	↓
Formánsfrekvenciák sávszélessége	↑
Intenzitás dinamikája	↓
Artikulációs sebesség	↓
Szünetek átlagos hossza	↑
Jitter	↑
Shimmer	↑
RoT	↓
MFCC	↓

3.1. táblázat. Kutatásom kezdetekor ismert változási tendenciák (depresszió hatására a beszéd akusztikai-fonetikai jellemzőiben)

Relatív szünethossz [%]:

Az emberi beszéd természetes részét képezik a szünetek. A szünetek lehetőséget biztosítanak a levegővételre, illetve a beszédtervezésben is fontos szerepük van. A megfelelően használt szünet segíthet a mondanivaló megértésében, akár nyomatékosíthat is. Relatív szünethosszon szünetek hosszának és a teljes beszédminta hosszának a százalékos arányát értjük a beszédmintákban található kezdő és záró szünetszakaszt leszámítva[37].

Tranziens arány (RoT) [%]:

Az emberi beszédproduktum felosztható közel állandó és változó (tranziens) szakaszok egymás utáni sorozatára. Az RoT jellemző alatt a beszédmintában található tranziens szakaszok hosszának (szüneteket elhagyva) százalékos arányát értjük a beszédmintában található beszédszakaszok hosszához képest (szüneteket elhagyva)[13].

3.1. Beszédjellemzők megváltozása depresszió hatására

Az egyes beszédjellemzők depresszióra gyakorolt hatásáról a szakirodalom nem minden esetben ért egyet, azt viszont napjainkban már kijelenthetjük, hogy a depressziós beszéd akusztikai-fonetikai jellemzői az egészségeshez képest megváltoznak. Az egyet-nem-értés egyik oka a jellemzők nyelvfüggősége lehet, illetve további problémát okozhat, hogy a felhasznált adatbázisok nem egy egységes keretrendszer szerinti mérés alapján készültek el; tehát a mérés körülményei különböznek.

A témában végzett irodalomkutatásom eredményeit a 3.1 táblázatban foglaltam össze[4][13][14][20][26][31][38][39].

4. fejezet

Statisztikai fogalmak, és gépi tanuló modellek

Kutatásom során statisztikai eszközökkel értékeltem ki eredményeimet. Az így kapott számszerű adatok segítenek a kapcsolatok megértésében. Továbbá a statisztikai modellekkel, gépi tanuló eljárások esetén, gyors, és egyszerű jellemzőkiválasztást végezhetünk. Ebben a fejezetben összefoglalom a felhasznált modellek elméleti hátterét.

4.1. Párosított t-próba

A páros mintás t-próba célkitűzése, hogy ugyanannak a sokaságnak két összetartozó mintáját hasonlítsa össze. Ezek az összetartozó minták jellemzően ugyanazon az egyedeken mért elő és- utóvizsgálat, esetleg valamilyen módon összepárosítható adatok lesznek. Ehhez mintákat szükséges vennünk a populációból (két darabot), ezen kívül az ott kapott értékek különbségének nullához szükséges közelítenie, esetleg pont nullának kell lennie. Így lehet kifejezni, hogy a páros mérés eredményei között nulla (azaz nincs) a különbség. A próba feltétele, hogy a különbséget adó minta normális eloszlású legyen.

- Nullhipotézis: a populáció átlagának a különbsége nulla
- Alternatív hipotézis: a populáció átlagának különbsége eltér nullától

A párosított t-próba próbastatisztikája:

$$t = \frac{\bar{x} - m}{\frac{s^*}{\sqrt{n}}} \quad (4.1)$$

A nullhipotézist elegendően nagy 1-p érték alapján tudjuk elutasítani, ahol a p értékét a Student-féle t-eloszlás táblázat alapján kaphatjuk meg[40][41]. Vizsgálataim során a statisztikában szokásos szignifikancia szintek mellett vizsgáltam:

- $p > 0,05$: nem szignifikáns
- $0,05 > p > 0,01$: *
- $0,01 > p > 0,001$: **
- $0,001 > p$: ***

4.2. Szupport Vektor Gépek (SVM)

Napjainkban relatív kisméretű depressziós adatbázisok érhetők el, ezekre az osztályozási feladatot leghatékonyabban a szupport vektor gépek valósítják meg.

Az SVM az osztályok közötti maximális távolságú hipersík keresésére törekszik. Ezt a hipersíkot nevezzük döntési határnak. Az alapötlet az, hogy egy olyan hipersíkot keresünk, amelyre az összes pozitív (tanító)minta a pozitív oldalra, és az összes negatív (tanító)minta a negatív oldalra kerül.

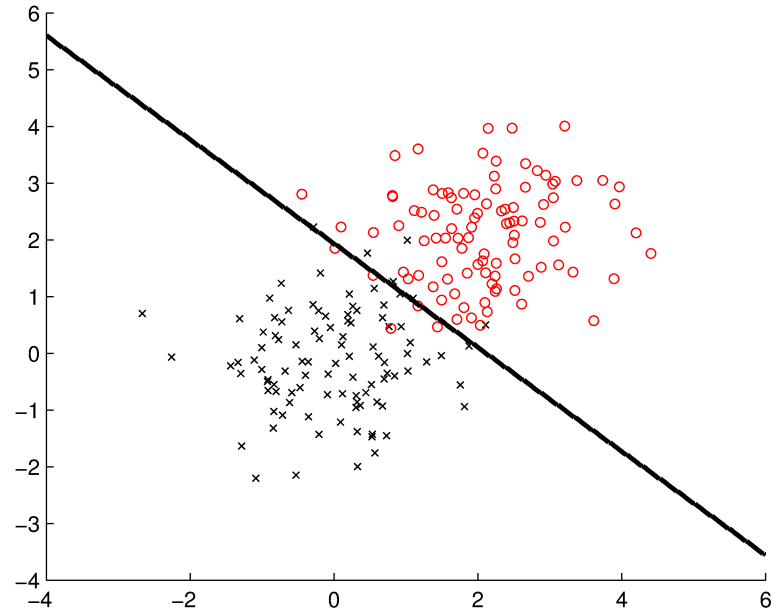
4.2.1. Lineárisan szeparálható eset

Ebben az esetben az algoritmus célja azon döntési határ megtalálása, amely a legnagyobb távolságra van a két osztály közötti pontoktól.

A döntési határ egy olyan hipersík, amely pontosan a két osztály között halad. Matematikailag a hipersíkot a $w \cdot x + b = 0$ egyenlet írja le, ahol a w a hipersíkra merőleges vektor, és $\frac{b}{\|w\|}$ a hipersík és az origó távolsága. Az SVM a hipersík meghatározásához a tanító vektorok közül azokat a x -vektorokat választja ki, amelyek legközelebb vannak a hipersíkhhoz, ezeket szupport vektoroknak nevezik. Az SVM feladata, hogy a tanítóminták alapján meghatározza w és b optimális értékét, úgy, hogy teljesüljön az alábbi képlet[42]:

$$y(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \in I \quad (4.2)$$

ahol y a két osztály reprezentációja -1-gyel és +1-gyel, az I pedig a szupport vektorok index halmaza, miközben az $\frac{1}{\|w\|}$ értéket maximalizálja[13].



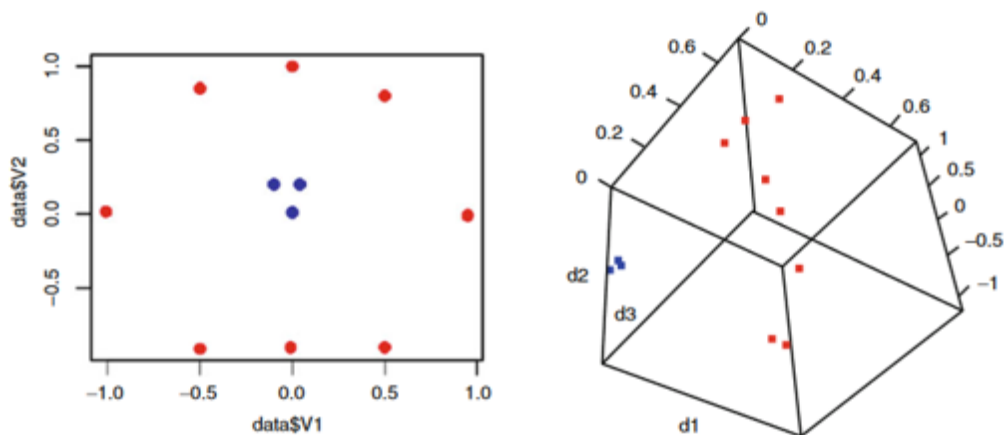
4.1. ábra. SVM működésének szemléltetése 2 dimenzió esetén.
Forrás: [43]

4.2.2. Lineárisan nem szeparálható eset

Gyakran találkozunk olyan adathalmazokkal, ahol a két osztály nem húzható be egyetlen síkkal vagy hipersíkkal.

Az SVM a lineárisan nem szeparálható esetekben történő alkalmazásához gyakran az úgynevezett kernel függvényt használja. Ez a technika lehetővé teszi, hogy az algoritmus olyan magasabb dimenziós térbe képezze az adatokat, ahol már lehetséges a lineáris szeparáció. A kernel függvények segítségével az SVM képes olyan transzformációt végezni az adatokon, amelyek lineárisan szeparálhatóvá teszik őket[42].

A legelterjedtebb kernel függvények[13]:



4.2. ábra. Magasabb dimenzióba képzés. Forrás: [42]

$$\text{Lineáris: } K(\theta(x_i), \theta(x_j)) = x_i^T \cdot x_j \quad (4.3)$$

$$\text{Polinomiális: } K(\theta(x_i), \theta(x_j)) = (\gamma \cdot x_i^T \cdot x_j + r)^d, \gamma > 0 \quad (4.4)$$

$$\text{Radiális bázisfüggvény: } K(\theta(x_i), \theta(x_j)) = \exp(-\gamma \cdot \|x_i - x_j\|^2), \gamma > 0 \quad (4.5)$$

$$\text{Sigmoid: } K(\theta(x_i), \theta(x_j)) = \tanh(\gamma \cdot x_i^T \cdot x_j + r) \quad (4.6)$$

ahol a d , γ és az r az adott kernel függvények hiperparaméterei.

Fontos megjegyezni azonban, hogy elronthatja a modell általánosító képességét az, ha túl magas dimenzióba transzformál a kernel függvény, így bevett módszer egy ξ gyengítő paraméter bevezetése, amely értéke a következő tartományokba eshet:

- ha az adott vektor a hipersík megfelelő oldalán helyezkedik el a margón kívül: $\xi = 0$
- ha az adott vektor a hipersík nem megfelelő oldalán helyezkedik el: $\xi = 1$
- ha az adott vektor a hipersík megfelelő oldalán helyezkedik el, de a margón belül: $0 < \xi < 1$

Ennek megfelelően mind a feltételek, mind a minimalizálni kívánt kifejezés módosul az alábbiak szerint:

$$y(\theta(x_i) \cdot w + b) - 1 + \xi_i \geq 0, \forall i \in I \quad (4.7)$$

$$\min_{w,b,\delta} \left\{ \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^k \xi_i \right\} \quad (4.8)$$

ahol k a tanítóminták száma, és a $C \geq 0$ konstans hiperparaméter szabályozza, hogy a hipersík margójához képest mekkora legyen az ξ_i által okozott büntetés mértéke. $C = 0$ esetében visszakapjuk a lineárisan szeparálható esetet[13].

4.3. Szupport Vektor Regresszió (SVR)

Az SVR esetében a tanítómintákhoz nem egy osztálycímke van rendelve (SVM-el ellentétben), hanem egy valós szám (y). Az SVR feladata, hogy kiválasszon egy olyan $f(x)$ függvényt, amely optimális módon közelíti az y célértékeket. Az optimális mód az SVR esetében azt jelenti, hogy olyan $f(x)$ függvényt keres, amely maximum

ε hibával követi az y értékeket, miközben minimális a függvény változása (vagyis nincsenek benne hirtelen nagy változások), ezáltal kompromisszumot teremtve a függvény pontossága és általánosító képessége között[13].

Az SVR döntési függvénye:

$$f(x) = w \cdot x + b \quad (4.9)$$

Ahhoz, hogy a függvény változása minél kisebb legyen, minimalizálnunk kell a $\frac{1}{2}||w||^2$ kifejezést az alábbi feltételek mellett:

$$|y_j - f(x_j)| \leq \varepsilon, \forall j \in J \quad (4.10)$$

Hasonlóan a SVM-hez, ebben az esetben is előfordulhat, hogy ezek a feltételek nem teljesülnek. Ilyenkor alkalmazhatunk egy kernel függvényt, amely segítségével a probléma egy magasabb dimenziójú térbe kerül át, és bevezethetünk egy gyengítő ξ paramétert. Ennek eredményeként a feltételrendszer és a feladat a következő módon módosul:

$$|y_j - f(x_j)| \leq \varepsilon + \xi_j, \forall j \in J \quad (4.11)$$

$$\min_{w,b} \left\{ \frac{1}{2}||w||^2 + C \cdot \sum_{i=1}^n \xi_i \right\} \quad (4.12)$$

ahol, (az SVM-hez hasonlóan) $C > 0$ konstanssal lehet befolyásolni a függvény változásának minimalizálását, az ε -nál nagyobb eltérések tolerálásának módosításával[13].

4.4. Regresszió pontosságának leírására alkalmazott jellemzők

A regressziós eljárás pontosságának becslésére három jellemzőt használtam. A Mean Absolute Error (MAE) az átlagos hibaértéket, a Root Mean Square Error (RMSE) az átlagos négyzetes hiba értékének a gyökét jelenti. Ezen értékek a következő képletekkel számíthatóak:

$$MAE = \frac{\sum_{i=1}^n |e_i|}{n} \quad (4.13)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (e_i^2)}{n}} \quad (4.14)$$

ahol az e_i az a i -dik minta becslésének a hibáját jelöli, az n a minták számát.

Továbbá a Pearson korrelációt, mint jellemzőt is használtam, a korrelációs együttható értéke az alábbi képlettel számítható.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{(n - 1) \cdot s_x^* \cdot s_y^*} \quad (4.15)$$

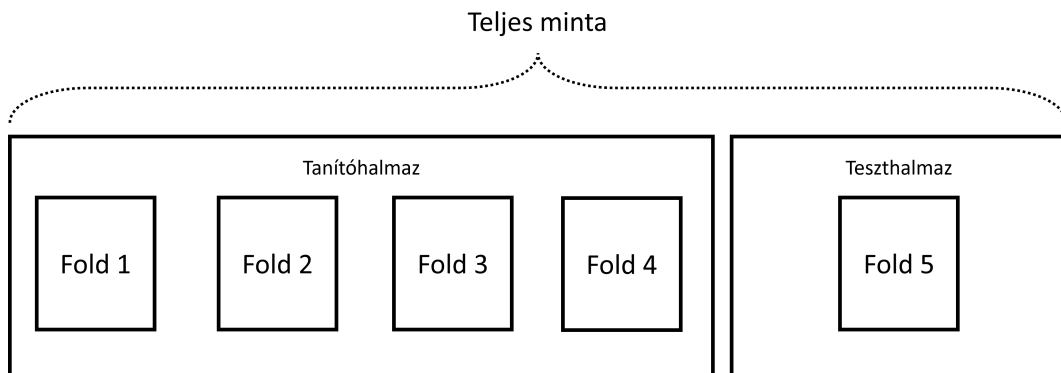
Az x_i és y_i az i -dik mintapár a mintasorban, míg az $\bar{x}$ és $\bar{y}$ ezek átlagát és az s_x^* és s_y^* ezek korrigált szórását jelöli.

Ha a korrelációs együttható 0 (H_0), akkor a két változó korrelálatlan, egyébként van közöttük lineáris kapcsolat (H_1)[13][44].

4.5. Keresztvalidáció

Gépi tanuló eljárások alkalmazása esetén, amennyiben az adatbázis viszonylag alacsony mintaszámmal rendelkezik, gyakori a keresztvalidáció technika alkalmazása. k -fold CV esetén a keresztvalidáció azt jelenti, hogy a folyamat elején nem egy statikus tanuló- és teszhalmazra, hanem k darab diszjunkt halmazra bontjuk a rendelkezésre álló mintákat. A modell alkalmazása során k iterációt hajtunk végre, melyben minden egyes iterációban $k - 1$ darab halmaz unióját használjuk tanításra, és a kimaradt halmazt tesztelésre[47]. (Lsd. 4.3 ábra.)

Beágyazott keresztvalidáció során egy olyan kiértékelési módszerről beszélünk, amely során a keresztvalidáción belül egy másik keresztvalidációt is alkalmazunk. A belső ciklus klasszikus keresztvalidációként működik, a külső ciklus pedig a teszhalmazán teszteli a belső ciklusban optimalizált hiperparaméterekkel rendelkező modellt[48].



4.3. ábra. Egy 5 fold keresztvalidáció egy iterációja

4.6. Fast Forward Selection

Mind az SVM és SVR működését befolyásolja a megfelelő kvantitású és kvalitású jellemzővektor kiválasztása[45]. A kiválasztás komoly jelentőséggel bír, hiszen 2^n lehetőségünk van (ahol n a jellemzők száma), és ezt magasabb n -ek esetén a gyakorlatban lehetetlen végigpróbálni. Erre a problémára (többek között[46]) egy megoldás a Fast Forward Selection (FFS) algoritmus.

Az FFS eljárás során $\frac{n \cdot (n+1)}{2}$ különböző hosszúságú jellemzővektort tesztelünk. Az eljárás lényege az, hogy gyorsan kiválassza az algoritmus által optimálisnak ítélt n elemű jellemzővektort az osztályozás vagy regresszió során, figyelembe véve az algoritmus által előállított különböző szuboptimális jellemzővektorokat. Az algoritmus működését a következő pontokban lehet összefoglalni[13]:

- Az eljárás üres jellemzővektorból indul ki.
- Ezután n lépésben mindig egy jellemzőt ad hozzá a jellemzővektorhoz, kiválasztva azt a jellemzőt, amely a legnagyobb pontosságot eredményezi.
- Elmenti az aktuális n hosszúságú jellemzővektort és az elért pontosságot. (Erre a követhetőség miatt van szükség)
- Végül az n lépés során előállított n darab eltárolt jellemzővektor közül azt választja ki, amely a legnagyobb pontosságot eredményezte.

5. fejezet

Felhasznált és létrehozott adatbázisok

Kutatásom során az egyetem Beszédakusztikai Laboratóriuma rendelkezésemre bocsájtotta a magyar nyelvű depressziós beszédadatbázisát. Ezt a modell működésének mélyebb megértésére, illetve az optimalizáláshoz használtam. Továbbá létrehoztam egy saját adatbázist is, amely különböző érzelmi töltetű, illetve depressziót imitált bemondásokat is tartalmaz. Ebben a fejezetben összefoglalom a két adatbázis legfontosabb jellemzőit.

5.1. Magyar Depressziós Beszédadatbázis

A kutatásom kezdetekor elérhető egyetlen, magyar nyelvű felvételeket tartalmazó, depressziós beszédadatbázis, melynek rögzítése során "Az Északi Szél és a Nap" c. akusztikailag és fonetikailag kiegyensúlyozott szöveget olvasta fel minden bemondó. "A felvételek csendes helyiségben kerültek rögzítésre Lexicon Alpha típusú külső usb audio interfész eszköz segítségével, 44,1 kHz mintavételi frekvenciával, 16 bites kvantálással, tömörítetlen PCM kódolással, ATR3350 típusú csiptetős mikrofonnal. A mikrofon a vizsgált személy mellkasához volt rögzítve." [13]

Az adatbázis készítésekor szempont volt, hogy ne csak a depresszió megléte, hanem annak súlyossága is rögzítve legyen, így felvették a bemondók BDI értékét, amelyeket szakorvosok validáltak.

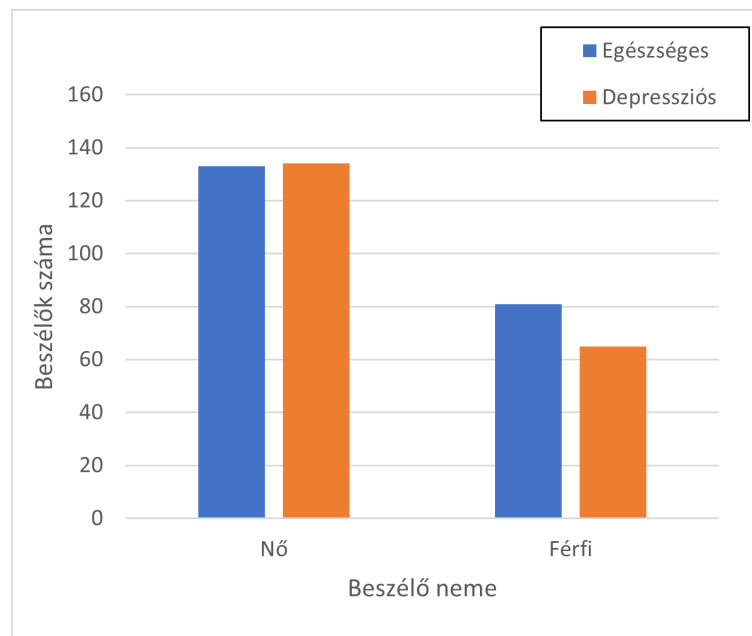
Erre az adatbázisra a későbbiekben "Magyar Depressziós Beszédatabázis (MDB)" néven hivatkozom.

5.1.1. Releváns információk az MDB adatbázisról

Az adatbázisban összesen 413 felvétel van, 214 egészséges és 199 depressziós embertől. A 5.1 táblázatban összefoglaltam a bementőkkel kapcsolatos legfontosabb információkat.

	Bementők száma (nők/férfiak)	Bementők BDI szerinti átlaga és szórása	Bementők életkorának átlaga és szórása
Depressziós bementők:	199 (134/65)	27,1 ($\pm 9,1$)	39,25 ($\pm 14,25$)
Egészséges bementők:	214 (133/81)	5,48 ($\pm 5,1$)	52,47 ($\pm 20,13$)
Összesített:	413 (267/146)	15,75 ($\pm 13,15$)	46 ($\pm 18,8$)

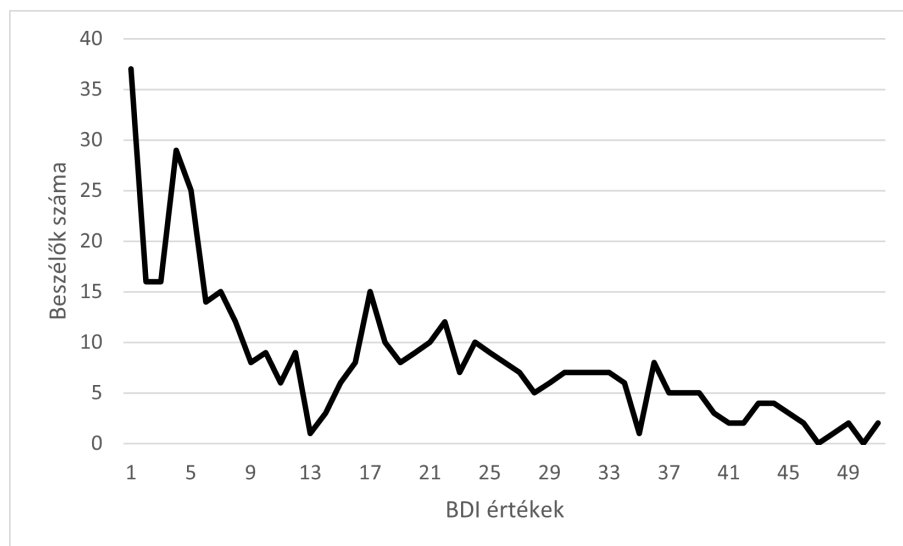
5.1. táblázat. Az MDB bementőinek adatai



5.1. ábra. Az MDB bementőinek nemek szerinti eloszlása a depressziós és az egészséges minták között

Továbbá az adatbázisban rögzítve vannak az alábbi információk is:

- Akut (depressziótól különböző) betegség a felvétel készítése idején, amely hatással lehet a hangképzésre. (Például megfázás, allergia)
- Dohányzik-e.
- Krónikus betegség a felvétel készítése idején, amely hatással lehet a hangképzésre. (Például hangszalag izombénulás, asztma, reflux)



5.2. ábra. AZ MDB bemondóinak BDI skála szerinti eloszlása

- Szed-e antidepresszánt, vagy egyéb gyógyszert.

Ezen információkat a kutatás alatt nem használtam fel, esetleges további kutatások alapjait szolgálhatják.

5.2. Magyar Depressziós Imitációs Beszédatadatbázis

Kutatásom kezdetekor még nem létezett adatbázis, amelyben a depresszió imitálására tett kísérleteket rögzítettek volna. Így lehetőségem nyílt saját adatbázis elkészítésére.

A felvételek elkészítése előtt minden bemondót tájékoztattam az adatkezelési kötelességeimről, és kitölttettem egy nyilatkozatot, miszerint tárolhatom az általuk származó hangfelvételeket[49].

A munka során igyekeztem olyan felvételeket készíteni, amelyek kritikus paraméterei megegyeznek a Magyar Depressziós Beszédatbázis szempontjaival (lsd.: 5.1 fejezet).

A felvételeket egy USB 5.1 külső hangkártya segítségével egy shure SM58 vokál mikrofonnal készítettem el. A mikrofont a vizsgált személy a kezében tartotta, 15-20 cm távolságra a szájától[50][51].

44,1 kHz-es mintavételezést használtam, 16 bites kvantálással.

Minden bemondóval 5 felvételt készítettem, melyben "Az Északi Szél és a Nap" szöveget olvasták fel az alábbiak szerint:

- Kontroll felolvasás: Minden érzelem nélkül történő felolvasás. Erre a továbbiakban "Kontroll minta"-ként hivatkozom.
- Depressziót imitált felolvasás. Fontos megjegyezni, hogy ezen kívül semmilyen instrukciót nem adtam, az adott bemondó döntése volt, hogy ez mit jelent. Erre a továbbiakban "Depressziót imitált minta"-ként hivatkozom.
- Dühös felolvasás. Erre a továbbiakban "Dühöt imitált minta"-ként hivatkozom.
- Szomorú felolvasás. Erre a továbbiakban "Szomorúságot imitált minta"-ként hivatkozom.
- Vidám felolvasás. Erre a továbbiakban "Vidámságot imitált minta"-ként hivatkozom.

Továbbá minden bemondóval kitöltettem egy BDI tesztet, melynek eredményére a továbbiakban "Saját BDI érték"-ként hivatkozom.

Erre az adatbázisra a későbbiekben "Magyar Depressziós Imitációs Beszédatadabázis (MDIB)" néven hivatkozom.

5.2.1. Releváns információk az MDIB adatbázisról

Az adatbázisban összesen 135 felvétel van, 27 különböző embertől. Az adatbázis

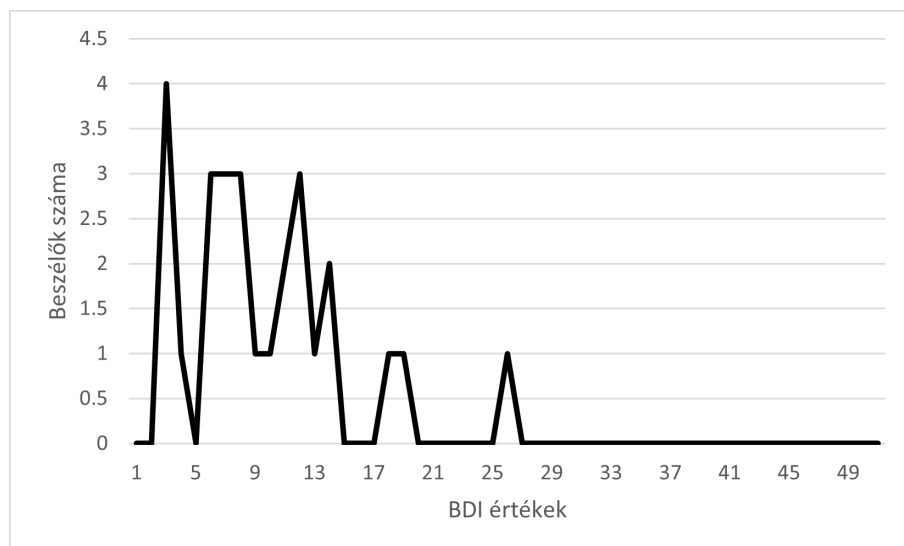
Bemondók száma (nők/férfiak)	Saját BDI értékek átlaga és szórása	Bemondók életkorának átlaga és szórása
27 (13/14)	8,63 ($\pm 5,35$)	31,3 ($\pm 16,57$)

5.2. táblázat. Az MDIB bemondóinak adatai

létrehozása során igyekeztem minél több színészt megkeresni. Végül a minták 48%-a származik színészekről.

5.2.2. Az érzelmek imitálásáról

Fontos megjegyezni hogy az MDIB-ben rögzített felvételek imitált érzelmekkel töltött bemondásokat tartalmaznak. Saját tapasztalatom, és korábbi kutatások is azt igazolják, hogy a színészek által előadott érzelmek gyakran hevesebbek, mint ahogy azt a valóságban tapasztaljuk[52]. Ennek ellenére az érzelmek beszéd alapú detektálásával foglalkozó kutatásokban bevett szokás, hogy színészek bemondásaival dolgoznak.



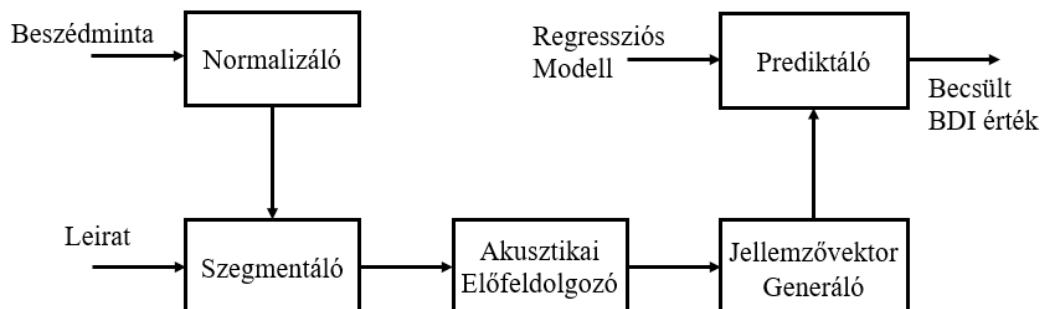
5.3. ábra. AZ MDIB bemondóinak BDI skála szerinti eloszlása

Mivel a kutatásom során kötött szöveggel dolgoztam, az érzelmek spontán megélése, és előadása nem tűnt járható útnak.

6. fejezet

Depression State Assessment

A kutatásom során a Beszédakusztikai Laboratórium rendelkezésemre bocsátott egy, beszédfeldolgozáson alapuló, depresszió súlyosságát prediktáló applikációt, az ún. Depression State Assessment alkalmazást (továbbiakban DSA). Működésének a blokkábrája a 6.1. ábrán látható. A beszédminta szegmentálására a pontosabb akusztikai-fonetikai jellemző kinyerés miatt van szükség[13].



6.1. ábra. A DSA működésének blokkábrája

A DSA szupport vektor regresszió (SVR) modellel működik[53].

6.1. DSA jellemzők

A DSA tanítómintáinak legfontosabb jellemzőit a 6.1 táblázatban foglaltam össze[13][54]. A táblázatból kiolvasható a bemondók BDI szerinti szórása, amit ha a tesztelés során kapott RMSE értékkel összevetünk, belátható, hogy (együttműködő páciensek esetén) a gép rendkívül jó performancia mellett működik.

Az alkalmazásra többnyelvű mintákat használtak. Ezek közül én a magyar nyelvű adatokat emeltem ki, mivel az általam rögzített felvételek is magyar nyelvű bemondásról történtek.

Bemondók száma (nők/férfiak)	Bemondók BDI szerinti átlaga és szórása	Bemondók BDI szerinti mediánja	Bemondók életkorának átlaga és szórása
127 (48/79)	14,2 ($\pm 13,5$)	8	42,2 ($\pm 14,4$)

6.1. táblázat. DSA alkalmazás tanítómintáinak fő jellemzői

A program tesztelésekor kapott értékek[53]:

- Mean Absolute Error: 6,7
- Root Mean Square Error: 8,9
- Pearson Correlation: 0,7

6.2. DSA tesztelése az MDIB mintáival

Minden rögzített felvételt kiértékeltem egyesével és feljegyeztem, hogy a DSA milyen BDI értéket döntött¹. Az így kapott értékeket több módon is összehasonlítottam. Egy szemléletes megoszlás látható a 6.2 ábrán. Megfigyelhető, hogy a DSA által döntött depresszió súlyosságára jellemző értékek átlagosan magasabb értéket vettek fel, mint a bemondók saját depresszió értékei. (BDI értékekből származtatott érték 2.4 alapján.)

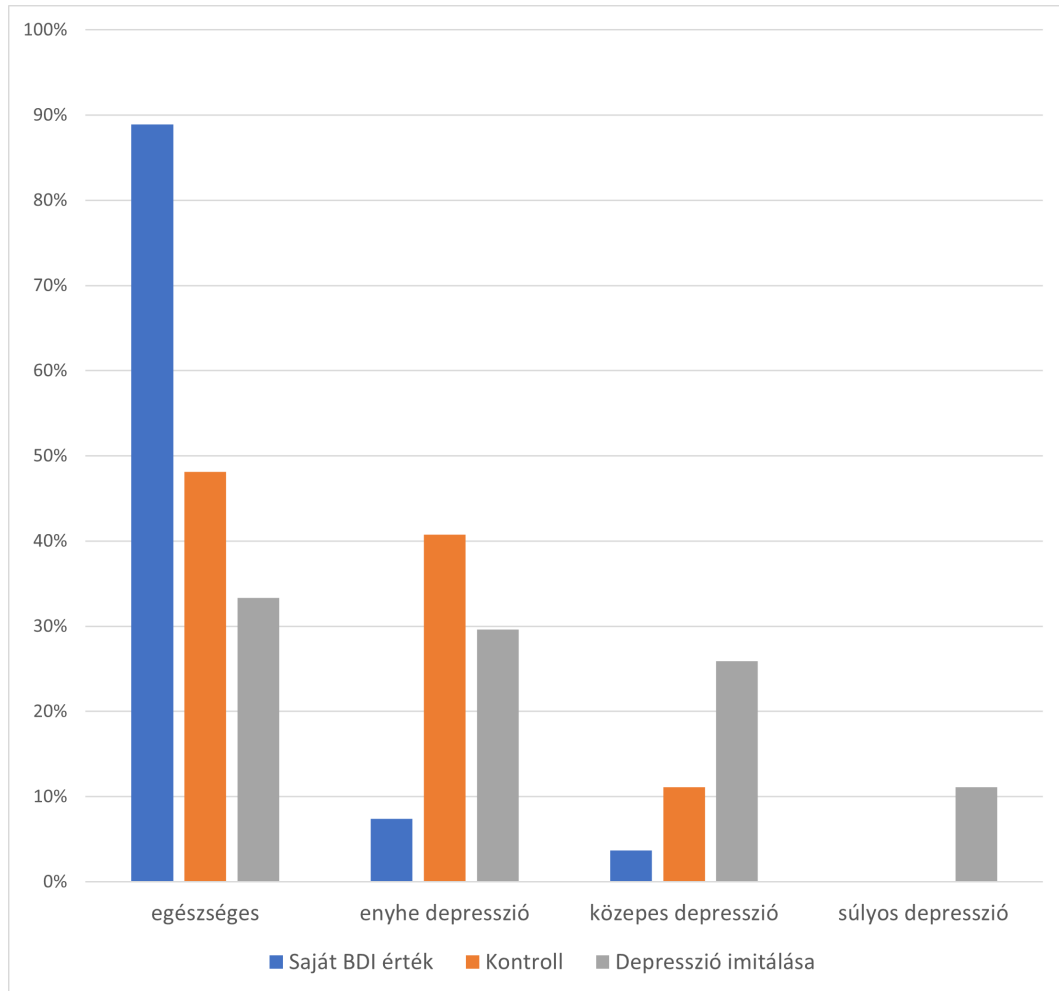
6.2.1. Saját BDI értékek párosítása a depressziót imitált mintákhoz

Bemondók által kitöltött BDI eredményeinek átlaga és szórása	Depressziót imitált mintákra prediktált BDI értékek átlaga	MAE	RMSE	Pearson korreláció (r értéke)
8,6 ($\pm 5,35$)	18,4	12,1	14,9	-0,12

6.2. táblázat. DSA regressziós működésének kiértékelése során kapott eredmények

A tesztelés eredményeit több módon értékeltem ki. Ebben a fejezetben az MDIB bemondói által kitöltött BDI teszt eredményeit és a depressziót imitált mintáikra a

¹Tehát ebben a fejezetben a "Saját BDI értékeket" leszámítva minden eredmény a DSA használatával született.



6.2. ábra. A saját; a kontroll mintákra és a depressziót imitált mintákra prediktált BDI értékek százalékos megoszlása

DSA által prediktált értékeket vettem össze. Az így kapott MAE és RMSE értékek szemléletesen azt mutatják be, hogy a DSA milyen hiba mellett prediktálja azon depresszió súlyossági szintet, amelyet a BDI teszt felvétele során az adott bemondó elért.

A 6.2 táblázatból kiolvashatók a tesztelésem során kapott értékek. Ha ezeket összevetjük a program tesztelésekor (ld.: 6.1) kapott eredményekkel, akkor megfigyelhető, hogy a DSA meglehetősen rosszul teljesített az imitált mintákra. A táblázatban feltüntetett eredményekből kiszámolható a mérés normalizált RMSE hibája.

$$normalizáltRMSE = \frac{RMSE}{\sigma} = \frac{14,9}{5,35} = 2,79 \quad (6.1)$$

Mivel az 1 értéknél lényegesen nagyobb eredmény született, kijelenthető, hogy a DSA lényegében teljesen véletlenszerűen prediktál BDI értéket a depressziót imitált

mintákra. Ennek egy fő oka, hogy a DSA tervezése során a bemondók kooperációját feltételezték.

6.2.2. Kontroll minták párosítása a depressziót imitált mintákhoz

Ebben az esetben az azonos bemondó mindenféle imitálás nélkül történő bemondását (kontroll minta lsd.: 5.2) és depressziót imitált bemondását vetettem össze. Szemléletesen azt vizsgáltam, hogy egy adott bemondó milyen szinten képes a DSA által prediktált értéket befolyásolni.

A mérés során a nullhipotézisem az alábbi volt: *A depresszió imitálása nincs hatással a DSA működésére.*

A prediktált BDI értékekre végzett párosított t-próba esetén $p < 0,001$ (***) értéket kaptam.

Tehát elmondható, hogy szignifikáns eltérés tapasztalható, a nullhipotézis elutasítható. Gyakorlatilag ez azt jelenti, hogy a DSA regressziós eljárása nem véletlen döntött más értéket a kontroll mintára, mint a depressziót imitált mintára. A 6.3 táblázatban bemutatott átlagok összehasonlításával belátható, hogy depressziót imitált bemondásra jellemzően magasabb értéket prediktált a DSA, mint a kontroll bemondásra.

A 2.4 fejezetben bemutatott határok alapján az alkalmazás osztályozó működését is megvizsgáltam. A minimális depressziót 0; az enyhe depressziót 1; a közepes depressziót 2; és a súlyos depressziót 3-as értékre egyszerűsítettem. Az így elvégzett párosított t-próba esetén $0,01 < p < 0,001$ (**) értéket kaptam.

Tehát ebben az esetben is szignifikáns eltérés tapasztalható, a nullhipotézis elutasítható. Gyakorlatilag ez azt jelenti, hogy a DSA által depressziót imitált bemondásra prediktált értékek átlagosan annyival magasabbak, hogy az már az osztályozó működésben is megjelenik. (Ha összeasonlítjuk a 6.3 táblázatban bemutatott átlag értékeket, látható, hogy átlagosan is 2 külön osztályba estek a predikciók: kontroll minták esetén minimális depresszióba, és depressziót imitált minták esetén enyhe depresszióba.)

A kiértékelés során született eredményeket és hibaértékeket összefoglaltam a 6.3 táblázatban. Az így kapott MAE és RMSE értékek azt mutatják meg, hogy azonos bemondó két különböző bemondására prediktált értékek milyen hibával vannak egymáshoz képest.

Kontroll mintákra prediktált BDI értékek átlaga és szórása	Depressziót imitált mintákra prediktált BDI értékek átlaga	MAE	RMSE	Pearson korreláció (r értéke)
12,4 ($\pm 6,61$)	18,4	7,6	9,6	0,61

6.3. táblázat. Kontroll mintákra és depressziót imitált mintákra DSA által prediktált eredmények

6.2.3. Kontroll minták párosítása különböző hangulatokat imitált mintákhoz

Megvizsgáltam, hogy milyen hatással vannak a különböző hangulatok a DSA által prediktált értékekre. Ahogy azt a 5.2.2 fejezetben is kifejtettem, fontos megjegyezni, hogy az MDIB-ban imitált érzelmekkel töltött bemondásokat rögzítettem.

A mérés során az alábbi nullhipotézisem volt: *Különböző hangulatok imitálása nincs hatással a DSA működésére.*

	t-próba (p értéke)	MAE	RMSE	Pearson korreláció (r értéke)
kontroll - düh	$0,05 > p > 0,01$ (*)	5,5	8,2	0,39
kontroll - vidámság	$0,05 > p > 0,01$ (*)	6,5	8,2	0,31
kontroll - szomorúság	$0,001 > p$ (***)	6,1	8,1	0,59

6.4. táblázat. Kiértékelés során született eredmények

A 6.4. táblázat alapján kijelenthető, hogy a pillanatnyi hangulat hatással van a döntött értékre. Az elvégzett párosított t-próba esetén mindhárom esetben elutasítható a nullhipotézis.

- Kontroll mintákra prediktált és dühöt imitált mintákra prediktált értékek összevetése esetén (*) szignifikancia szint mellett.
- Kontroll mintákra prediktált és vidámságot imitált mintákra prediktált értékek összevetése esetén (*) szignifikancia szint mellett.
- Míg kontroll mintákra prediktált és szomorúságot imitált mintákra prediktált értékek összevetése esetén (***) szignifikancia szint mellett.

7. fejezet

Modellépítés

A 6. fejezetben bemutattam egy (rendelkezésemre bocsátott) modellen keresztül, hogy milyen eredmények születnek, ha az MDIB-ben rögzített imitált beszédmintákkal végzem a tesztelést. Megállapítottam, hogy azonos bemondó mindenféle imitálás nélkül történő bemondását és depressziót imitált bemondását összevetve (párosított t-próba esetén) szignifikáns (***) eltérés tapasztalható.

Így felmerül az igény egy olyan modell kiépítésére, amely fel van készítve depressziót imitáló bemondókra, hiszen előfordulhat, hogy a bemondó együttműködése nem biztosított.

Depresszió imitálására több indok is előállhat. Többek között a vizsgálandó személy személyes preferenciája (pl.: hosszabb betegszabadság elérése), vagy akár pszichiai tévképzet is (melyben a vizsgálandó személy azt gondolja magáról, hogy depressziós, miközben nem az).

Továbbá egy olyan igény is támasztható a modell felé, hogy képes legyen szűrni a különböző hangulatú bemondásokat. Gyakorlati példával élve: Képes legyen elkülöníteni, hogy a bemondó éppen pillanatnyilag szomorú, vagy depressziós. Hiszen előfordulhat, hogy a bemondó a felvétel elkészítése során éppen egy külső esemény miatt valamilyen hangulatban van. (Pl.: elpusztult a macskája, és ezért szomorú.)

7.1. Jellemzőkinyerés és előfeldolgozás

Kutatásom során a Beszédakusztikai Laboratórium rendelkezésemre bocsátott egy jellemzőkinyerésre felhasználható alkalmazást. A kinyert jellemzők között elkülönülnek a beszédhangszinten értelmezhető (f0, F1, F2, B1, B2, intenzitás, jitter, shimmer, mel-sávós energiaérték) és beszédhangszinten nem értelmezhető jellemzők (AS, RSzH, RoT).

Előbbieket különböző lépésközökkel, minden lehetséges helyen kimértem.

Utóbbiakból AS-t, és RoT-ot (szüneteket elhagyva) a teljes beszédszakaszon, illetve

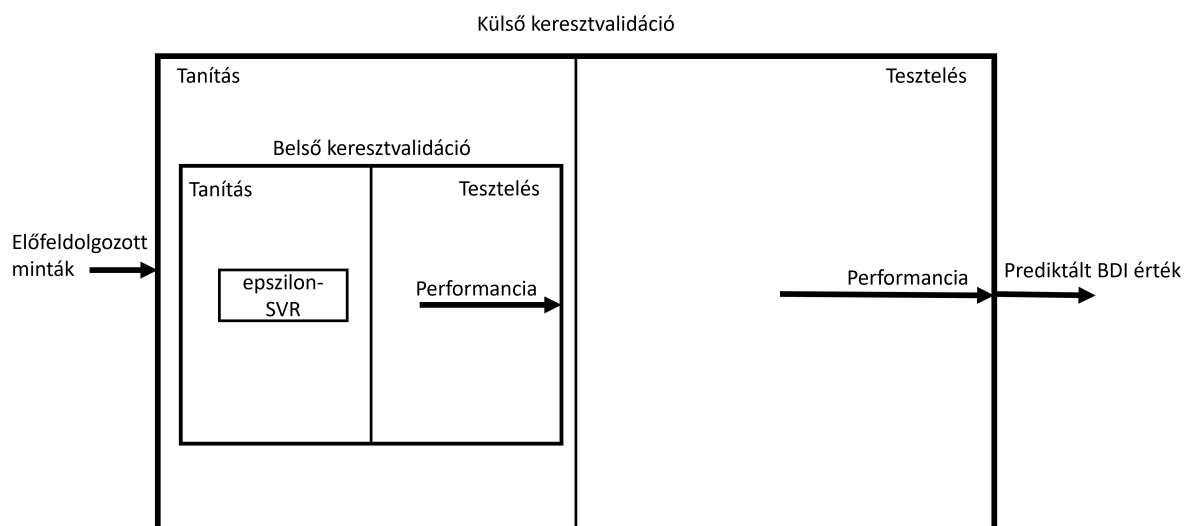
RSZH-t a teljes felvételen (elhagyva a felvétel elején és végén lévő szüneteket) mértem.

Ennek eredményeként minden felvételtől 1420 jellemzőt nyertem ki. Tehát ebben az esetben az összes lehetséges jellemzővektor száma: 2^{1420} . Mivel ennek a mennyiségnek a feldolgozása a gyakorlatban végtelenhez közelítő időbe telne, a jellemzőkön Pearson korreláció alapú előválogatást végeztem. A depressziós állapot súlyosságával szignifikánsan korreláló jellemzőket választottam a jellemzővektorba. A válogatást 3 különböző esetre végeztem el.

- a Magyar Depressziós és Imitációs Beszédatbázis egyesített mintáira
- az egyesített adatbázis női mintáira
- az egyesített adatbázis férfi mintáira

7.2. Modell realizálása

A modellt RapidMiner keretrendszerben valósítottam meg[55]. Az adatbázis relatív kis mérete miatt, az ilyenkor szokásos beágyazott keresztvalidációt használtam ($20fold \cdot 20fold$). Benne egy epsilon-SVR gépet tanítottam rbf kernellel. Mivel beágyazott keresztvalidáció esetén az könnyen automatizálható, a kernel két hiperparaméterének (cost, gamma) értékeit a 2 hatványaira teszteltem. (A kitevő -10 és +10 közötti tartományában.) A modell blokvázlata a 7.1 ábrán látható.



7.1. ábra. A felépített modell blokvázlata

8. fejezet

A modell tesztelése és eredmények

Az optimalizált modell tesztelését, performanciájának vizsgálatát a RapidMiner keretrendszerben végeztem el.

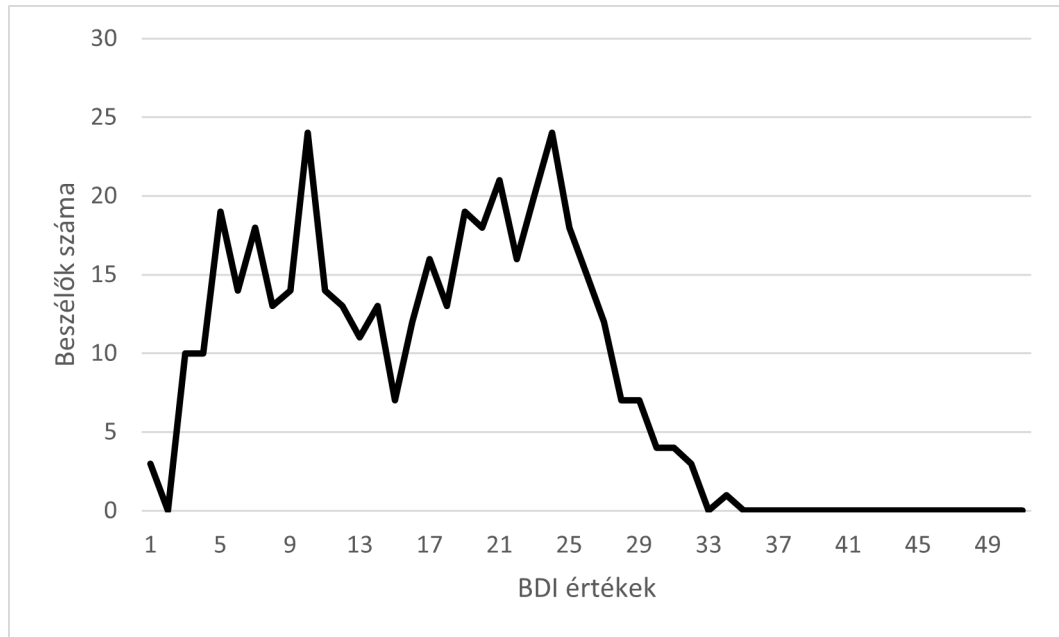
8.1. Modell tesztelése az MDB segítségével

A fejlesztés során fontos szempont volt, hogy kompromisszumot teremtsék céljaim és a "valós depresszió detektálás" között. Tehát, hogy az új modell egyrészt legyen képes különválasztani a különböző hangulatú és depressziót imitált bemondásokat a depressziótól, másrészt viszont képes legyen a valós depressziót felismerni. Ebben a fejezetben bemutatott eredményeim szemléltetik, hogy a valós depresszió súlyosságának prediktálására milyen eredményeket értem el.

MDB bemondóinak BDI szerinti átlaga és szórása	MAE	RMSE	Pearson korreláció (r értéke)
15,75 ($\pm 13,15$)	8,7	11,12	0,54

8.1. táblázat. MDB mintáival történő tesztelés eredményei

A 8.1 táblázatban foglaltam össze az összesített eredményeket (vö. 6.1 fejezetben bemutatott eredményekkel). Fontos megjegyezni, hogy a DSA tanítása kevesebb mintával történt. A modell performanciája némiképp romlott a korábbiakhoz képest. Ez nem meglepő, hiszen az új modell tanítása során az imitált beszédmintákat is felhasználtam. Ezen mintákból kinyert jellemzők olykor lényegesen más értékeket vesznek fel, mint amikkel a korábbi gépeket tanították. De ami itt negatív hatásként jelenik meg, éppen az ad majd lehetőséget a döntés hangulatfüggésére és becsaphatóságára történő optimalizálásra.



8.1. ábra. A saját modellem által prediktált értékek az MDB mintáival történő tesztelés során

A tanítás során megvizsgáltam, hogy milyen hatással van a gép performanciájára, ha kizárólag női/férfi mintákkal tanítom, majd tesztelem a modellt.

	MAE	RMSE	Pearson korreláció (r értéke)
Női bemondók:	9,76	12,14	0,44
Férfi bemondók:	8,26	10,89	0,49
Összesített:	9,23	11,71	0,46

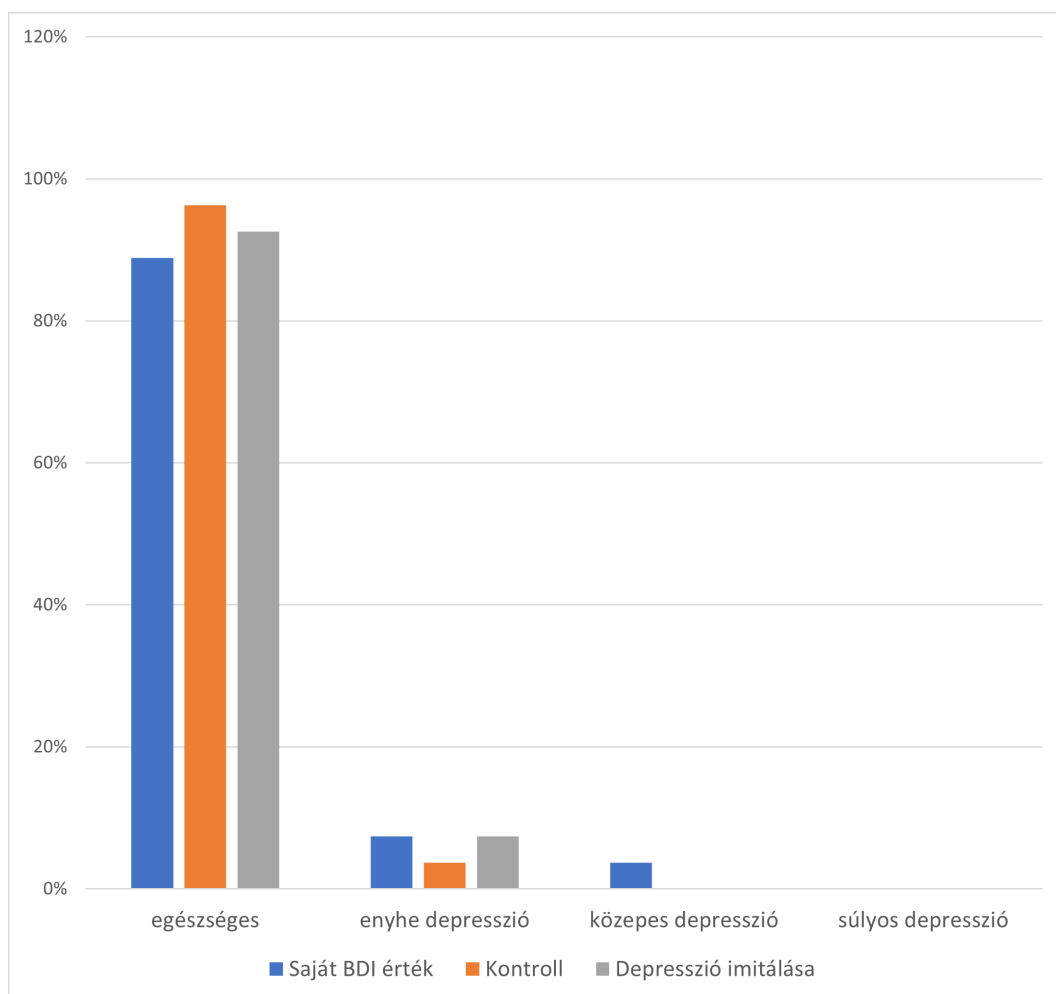
8.2. táblázat. MDB női illetve férfi mintáival való tanítás és tesztelés eredményei

A 8.2 táblázatban foglaltam össze az így kapott eredményeimet. Az összesített eredményeimet összehasonlítva a 8.1 táblázatban bemutatott eredményeimmel beláttam, hogy a modell rosszabbul teljesít, ha külön férfi és női mintákkal történő tanítást végzek.

8.2. A modell tesztelése az MDIB segítségével

Ebben a fejezetben bemutatott eredményeim képesek szemléltetni, hogy az optimalizált modell MDIB-vel történő tesztelése során hogyan teljesített. A követhető összehasonlítást elősegítve eredményeimet a 6.2 fejezet sorrendje szerint mutatom

be. Egy szemléletes megoszlás látható a 8.2 ábrán (vö.: 6.2 ábra). Természetesen ebben a fejezetben az optimalizált modell által prediktált értékekkel dolgoztam.



8.2. ábra. A saját; a kontroll mintákra és a depressziót imitált mintákra prediktált BDI értékek százalékos megoszlása

8.2.1. Saját BDI értékek párosítása a depressziót imitált mintákhoz

Ebben a fejezetben az MDIB bemondói által kitöltött BDI teszt eredményeit és a depressziót imitált mintáikra az optimalizált modell által prediktált értékeket vetettem össze.

A 8.3 és 8.4 táblázatokban számszerűen is összefoglaltam a tesztelésem során kapott értékeket. Az összehasonlíthatóság kedvéért feltüntettem a DSA által prediktált eredményeket is (6.2 táblázatban bemutatott eredmények). Az átlag értékek összevetéséből látszik, hogy az optimalizált modell jellemzően alacsonyabb depresszió

súlyossági értéket prediktált a depressziót imitált mintákra, mint amit az MDIB bemondói a BDI teszt kitöltésével elértek.

	Bemondók által kitöltött BDI eredményeinek átlaga és szórása	Depressziót imitált mintákra prediktált BDI értékek átlaga
DSA:	8,6 ($\pm 5,35$)	18,4
Optimalizált modell:	ua.	7,76

8.3. táblázat. A DSA és az optimalizált modell regressziós működésének kiértékelése során kapott átlagok

	MAE	RMSE	normalizált RMSE	Pearson korreláció (r értéke)
DSA:	12,1	14,9	2,79	-0,12
Optimalizált modell:	4,09	5,76	1,08	0,18

8.4. táblázat. A DSA és az optimalizált modell regressziós működésének kiértékelése során kapott eredmények

A normalizált RMSE értékekből kiderül, hogy az optimalizált modell - habár lényegesen jobban teljesít - még mindig egészen pontatlan becslést ad a depressziót imitált minták BDI értékére. Ez az eredmény vezetett arra, hogy a tesztelést elvégeztem a külön női és férfi mintákkal tanított modellre is. Az így kapott eredményeimet összefoglaltam a 8.5 táblázatban (vö.: 8.3 és 8.4). Az alacsonyabb MAE és RMSE értékből arra következtethetünk, hogy az így tanított modell kicsit pontosabban működik. A normalizált RMSE ebben az esetben már 1 alatti értéket vesz fel ($\frac{RMSE}{\sigma} = 0,85$), tehát a nemenkénti külön tanítás a depressziót imitált minták kiszűrhetőségéhez vezet.

Bemondók által kitöltött BDI eredményeinek átlaga és szórása	Depressziót imitált mintákra prediktált BDI értékek átlaga	MAE	RMSE	Pearson korreláció (r értéke)
8,6 ($\pm 5,35$)	9,08	3,28	4,52	0,57

8.5. táblázat. A külön női és férfi mintákkal tanított modell regressziós működésének kiértékelése során kapott eredmények

8.2.2. Kontroll minták párosítása a depressziót imitált mintákhoz

Ebben az esetben az azonos bemon-dó mindenféle imitálás nélkül történő bemon-dását (kontroll minta, lsd.: 5.2) és depressziót imitált bemon-dását vetettem össze. Szemléletesen azt vizsgáltam, hogy egy adott bemon-dó milyen szinten képes az optimalizált modell által prediktált értéket befolyásolni.

A mérés során az alábbi nullhipotézisem volt: *A depresszió imitálása nincs hatással az optimalizált modell működésére.*

A prediktált BDI értékekre végzett párosított t-próba esetén $p > 0,05$ (**nem szignifikáns**) értéket kaptam.

Tehát a nullhipotézis nem utasítható el.

A 2.4 fejezetben bemutatott határok alapján az alkalmazás osztályozó működését is megvizsgáltam. A minimális depressziót 0; az enyhe depressziót 1; a közepes depressziót 2; és a súlyos depressziót 3-as értékre egyszerűsítettem. Az így elvégzett párosított t-próba esetén $p > 0,05$ (**nem szignifikáns**) értéket kaptam.

Tehát a nullhipotézis ebben az esetben sem utasítható el.

	Kontroll mintákra prediktált BDI értékek átlaga és szórása	Depressziót imitált mintákra prediktált BDI értékek átlaga
DSA:	12,4 ($\pm 6,61$)	18,4
Optimalizált modell:	7,13 ($\pm 2,36$)	7,76

8.6. táblázat. A DSA és az optimalizált modell által prediktált BDI értékek átlaga

	MAE	RMSE	Pearson korreláció (r értéke)
DSA:	7,6	9,6	0,61
Optimalizált modell:	1,98	3,1	0,44

8.7. táblázat. A DSA és az optimalizált modell által kontroll és depressziót imitált mintákra prediktált eredmények kiértékelése

A kiértékelés során született eredményeket és hibaértékeket összefoglaltam a 8.6 és 8.7 táblázatokban. Az összehasonlíthatóság kedvéért feltüntettem a DSA által prediktált eredményeket is (6.3 táblázatban bemutatott eredmények). Az alacsony MAE és RMSE értékek szemléletesen azt mutatják, hogy az optimalizált gép közel hasonló értéket prediktált a két különböző bemon-dás esetére.

A tesztelést elvégeztem a külön női és férfi mintákkal tanított modellre is. A

prediktált BDI értékekre végzett párosított t-próba esetén ebben az esetben is $0,05 > p > 0,01$ (*) értéket kaptam. Tehát a nullhipotézis ebben az esetben elutasítható.

A 2.4 fejezetben bemutatott határok alapján az alkalmazás osztályozó működését is megvizsgáltam. Az így elvégzett párosított t-próba esetén $p > 0,05$ (**nem szignifikáns**) értéket kaptam. Tehát a nullhipotézis ebben az esetben nem utasítható el.

Kontroll mintákra prediktált BDI értékek átlaga	Depressziót imitált mintákra prediktált BDI értékek átlaga	MAE	RMSE	Pearson korreláció (r értéke)
7,44	9,08	2,46	3,74	0,58

8.8. táblázat. A külön női és férfi mintákkal tanított optimalizált modell kiértékelése során született eredmények

Az így kapott eredményeimet összefoglaltam a 8.8 táblázatban (vö.: 8.6 és 8.7).

8.2.3. Kontroll minták párosítása különböző hangulatokat imitált mintákhoz

Megvizsgáltam, hogy milyen hatással vannak a különböző hangulatok az optimalizált modell által prediktált értékekre. Ahogy azt a 5.2.2 fejezetben is kifejtettem, fontos megjegyezni, hogy az MDIB-ban imitált érzelmekkel töltött bemondásokat rögzítettem. *Különböző hangulatok imitálása nincs hatással az optimalizált modell működésére.*

	t-próba (p értéke)	MAE	RMSE	Pearson korreláció (r értéke)
kontroll - düh	$p > 0,05$ (n. sz.)	1,95	2,53	0,3
kontroll - vidámság	$p > 0,05$ (n. sz.)	1,4	1,78	0,75
kontroll - szomorúság	$p > 0,05$ (n. sz.)	1,8	2,26	0,53

8.9. táblázat. Az optimalizált modell kiértékelése során született eredmények

A 8.9 táblázatban összefoglaltam a kiértékelés során született eredményeimet (vö.: 6.4). Az elvégzett párosított t-próba esetén egyik esetben sem utasítható el a nullhipotézis. Az alacsony MAE és RMSE értékek szemléletesen azt mutatják, hogy az optimalizált gép közel hasonló értéket prediktált a két különböző bemondás esetére. Tehát sikerült a modell hangulatfüggését minimalizálni.

Ebben az esetben is elvégeztem a tesztelést a külön női és férfi mintákkal tanított modellre. A 8.10 táblázatban foglaltam össze az így kapott eredményeimet (vö.: 6.4

és 8.9).

	t-próba (p értéke)	MAE	RMSE	Pearson korreláció (r értéke)
kontroll - düh	$p > 0,05$ (n. sz.)	2,18	2,95	0,4
kontroll - vidámság	$p > 0,05$ (n. sz.)	1,8	2,67	0,55
kontroll - szomorúság	$p > 0,05$ (n. sz.)	2,56	4,1	0,53

8.10. táblázat. A külön női és férfi mintákkal tanított modell kiértékelése során született eredmények

9. fejezet

Összefoglalás

A depresszió egy komplex és komoly betegség, amely - súlyosságától függően - nem csak az egyén életére, de a közvetlen szociális környezetére is hatással lehet. Az időben történő diagnosztizálás kiemelt jelentőségű a betegek gyógyulása szempontjából, de mivel az orvosi diagnózis idő- és költségigényes, fontos kutatási terület, hogy hogyan lehet automatizált rendszerekkel támogatni ezt a folyamatot.

Dolgozatomban összefoglaltam a depresszió meghatározását, és diagnosztizálásával kapcsolatos nehézségeket. Bemutattam egy depresszió súlyosságát leíró skálát, a Beck Depression Inventory-II-t.

Továbbá bemutattam az emberi beszéd leírására alkalmas beszédakusztikai jellemzőket, és - a feltárt irodalom szerinti - megváltozásukat a depresszió hatására.

Ismertettem az eredmények megértéséhez szükséges statisztikai fogalmakat. Ugyancsak ismertettem a Szupport Vektor Gépek működésének elvét és a gépi tanuló modellekkel kapcsolatos eszközöket.

Kutatásom során rendelkezésemre állt egy magyar nyelvű depressziós beszédadatbázis, amely paramétereit ismertettem dolgozatomban.

Létrehoztam egy saját beszédadatbázist, amelyben a depresszió imitálására történő kísérleteket, továbbá különböző hangulatú bemondásokat rögzítettem.

Adatbázissal teszteltem egy, már létező, beszéd alapú, depresszió súlyosságának becslésére képes alkalmazás működését, az ezzel kapcsolatos megállapításaim az alábbiak:

Azonos bemondó mindenféle imitálás nélkül történő bemondására DSA által prediktált és depressziót imitált bemondásáára DSA prediktált értékeket összevetve (párosított t-próba esetén) szignifikáns (***) eltérés tapasztalható.

Az alkalmazás osztályozó működésére végzett vizsgálat során, azonos bemondó

mindenféle imitálás nélkül történő bemondására DSA által prediktált és depressziót imitált bemondására DSA prediktált értékeket összevetve (párosított t-próba esetén) szignifikáns (***) eltérés tapasztalható.

Imitálás során keletkezett hiba a Mean Absolute Error és a Root Mean Square Error eredeti (az alkalmazás tesztelésekor kapott) értékéhez képest az eredő hiba értékét megnöveli.

A pillanatnyi hangulat befolyással bír a keletkező beszédproduktumra, és az így keletkező hiba értéke tovább növeli az alkalmazás pontatlanságát.

Ezen eredmények nyújtották a motivációt egy olyan modell elkészítésére, amely alkalmas az imitált és a valós depresszió különválasztására, továbbá képes a pillanatnyi hangulatból fakadó hibát minimalizálni.

Dolgozatomban bemutattam egy optimalizált modellt, amelynek a tesztelése után az alábbi megállapításokat tettem:

Lehetséges az imitált és a valós depresszió különválasztása beszédelemzés alapján. Megfelelő tanítás mellett csökkenthető a beszéd alapú depresszió súlyosságát becslő modellek döntésének hangulatfüggése.

A depresszió beszéd alapú automatizált felismerése egy fontos, de gyakorlatban nem alkalmazott kutatási terület. A modellek által elért eredmények az adatbázisok bővítésével feltehetően javíthatóak.

A depresszió beszéd alapján történő detektálása a továbbiakban is egy fontos, kihívásokkal teli, de arra abszolút érdemes munka.

Köszönetnyilvánítás

Szeretném megköszönni Kiss Gábornak a kitartó és megbízható támogatást.

Soroncz-Szabó Flóriánnak az MDIB rögzítése során felmerülő technikai nehézségek leküzdésében nyújtott segítségét és Szabó Andrásnak a modellem elkészítéséhez adott tanácsait.

Édesanyámnak és édesapámnak a lektorálást, és Dalotti Andrásnak az előválogatásban nyújtott segítségét.

Hálásan köszönöm az Örkény István színház, a Katona József színház, a Vígyszínház, és a Színház- és Filmművészeti egyetem színészeinek a közreműködést, név szerint: Elek Ferenc, Ertl Zsombor Péter, Fullajtár Andrea, Fülöp Tímea, Für Anikó, Juniki Noémi, Koltainé Pogány Judit Anna, Liber Ágoston, Majsai-Nyilas Tünde, Máté Gábor, Nagy Zsolt, Németh Áron Valentin, Szandtner Anna, Tenki Réka.

Továbbá köszönöm az MDIB létrehozásában nyújtott segítséget Bakó Borókának, Csuka Balázsnak, Deák Júliának, Illés Dáriusnak, Koltai Kristófnak, Korompay Ábelnek, Kotmayer Annának, Makovsky Rékának, Nagy Blankának, Osztie Gergelynek, Tibiássy Adalbertnek, Vörös Gergelynek.

Irodalomjegyzék

- [1] Depressive disorder (who) <https://www.who.int/news-room/fact-sheets/detail/depression> megtekintés: 2023.09.26
- [2] World Health Organization. (2017). Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2). World Health Organization.
- [3] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech communication*, 71, 10-49.
- [4] Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope investigative otolaryngology*, 5(1), 96-116.
- [5] Marcus, M., Yasamy, M. T., van Ommeren, M. V., Chisholm, D., & Saxena, S. (2012). Depression: A global public health concern.
- [6] Péter, T., Lilla, S., Klaudia, D., László, K., & Márta, N. (2009). A depresszió diagnosztikája és kezelése a családorvosi gyakorlatban.
- [7] Gilbody, S., Whitty, P., Grimshaw, J., & Thomas, R. (2003). Improving the detection and management of depression in primary care. *Quality & safety in health care*, 12(2), 149.
- [8] Schumann, I., Schneider, A., Kantert, C., Löwe, B., & Linde, K. (2012). Physicians' attitudes, diagnostic process and barriers regarding depression diagnosis in primary care: a systematic review of qualitative studies. *Family practice*, 29(3), 255-263.
- [9] American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). *American Psychiatric Pub.*

- [10] A depresszió tünetei és kezelése, <https://semmelweis.hu/pszichoterapiasosztaly/kiket-gyogytunk/a-depresszio-tunetei-es-kezelese/> megtekintés: 2023.05.20
- [11] Luscher, B., Shen, Q., & Sahir, N. (2011). The GABAergic deficit hypothesis of major depressive disorder. *Molecular psychiatry*, 16(4), 383-406.
- [12] Dwivedi, Y., Rizavi, H. S., Conley, R. R., Roberts, R. C., Tamminga, C. A., & Pandey, G. N. (2003). Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. *Archives of general psychiatry*, 60(8), 804-815.
- [13] Kiss, G. (2019). A depressziós beszéd akusztikai-fonetikai jellemzőinek vizsgálata.
- [14] Hall, J. A., Harrigan, J. A., & Rosenthal, R. (1995). Nonverbal behavior in clinician—patient interaction. *Applied and preventive psychology*, 4(1), 21-37.
- [15] Sobin, C., & Sackeim, H. A. (1997). Psychomotor symptoms of depression. *American Journal of Psychiatry*, 154(1), 4-17.
- [16] Kent, R. D. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication disorders*, 33(5), 391-428.
- [17] Quatieri, T. F., & Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity. In Thirteenth annual conference of the international speech communication association.
- [18] Trevino, A. C., Quatieri, T. F., & Malyska, N. (2011). Phonologically-based biomarkers for major depressive disorder. *EURASIP Journal on Advances in Signal Processing*, 2011, 1-18.
- [19] Brooks, S. J., Nilsson, E. K., Jacobsson, J. A., Stein, D. J., Fredriksson, R., Lind, L., & Schiöth, H. B. (2014). BDNF polymorphisms are linked to poorer working memory performance, reduced cerebellar and hippocampal volumes and differences in prefrontal cortex in a Swedish elderly population. *PloS one*, 9(1), e82707.
- [20] Alpert, M., Pouget, E. R., & Silva, R. R. (2001). Reflections of depression in acoustic measures of the patient’s speech. *Journal of affective disorders*, 66(1), 59-69.

- [21] Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical psychology review*, 8(1), 77-100.
- [22] Kiss, G., Sztahó, D., & Vicsi, K. (2019). Depression State Assessment: Application for Detection of Depression by Speech. *In INTERSPEECH* (pp. 966-967).
- [23] Hajduska-Dér, B., Kiss, G., Sztahó, D., Vicsi, K., & Simon, L. (2022). The applicability of the Beck Depression Inventory and Hamilton Depression Scale in the automatic recognition of depression based on speech signal processing. *Frontiers in Psychiatry*, 13, 879896.
- [24] Beck, A. T., & Beamesderfer, A. (1974). Assessment of depression: the depression inventory. S. Karger.
- [25] Olaszy Gábor (2010), A beszédképzés folyamata. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 19-26.)
- [26] Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of neurolinguistics*, 20(1), 50-64.
- [27] Stassen, H. H., Kuny, S., & Hell, D. (1998). The speech analysis approach to determining onset of improvement under antidepressants. *European Neuropsychopharmacology*, 8(4), 303-310.
- [28] Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological psychiatry*, 72(7), 580-587.
- [29] Farrús, M., Hernando, J., & Ejarque, P. (2007). Jitter and shimmer measurements for speaker recognition. In 8th Annual Conference of the International Speech Communication Association; 2007 Aug. 27-31; Antwerp (Belgium). [place unknown]: ISCA; 2007. p. 778-81.. International Speech Communication Association (ISCA).
- [30] Jitter DDP (Praat's original Get Jitter) https://www.fon.hum.uva.nl/praat/manual/PointProcess__Get_jitter__ddp____.html megtekintés: 2023.10.05.
- [31] Ozdas, A., Shiavi, R. G., Silverman, S. E., Silverman, M. K., & Wilkes, D. M. (2004). Investigation of vocal jitter and glottal flow spectrum as possible

- cues for depression and near-term suicidal risk. *Ieee transactions on Biomedical engineering*, 51(9), 1530-1540.
- [32] Imre, V., Barlangi, R. & Vicsi, K. Kórós beszédhangok statisztikai analízise és osztályozása. *TDK dolgozat*
 - [33] Vicsi Klára (2010), A beszéd fizikai jellemzése, *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. *Akadémia Kiadó* (50-51.o, 42.o)
 - [34] Vicsi Klára (2010), Pszichofizikai tényezők. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. *Akadémia Kiadó* (pp. 56-69.)
 - [35] Kamath, U., Liu, J., & Whitaker, J. (2019). Deep learning for NLP and speech recognition (Vol. 84). Cham, Switzerland: Springer (pp. 372-376.)
 - [36] Szaszák György (2010), Kepsztrum. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. *Akadémia Kiadó* (pp. 239-243.)
 - [37] Olaszy Gábor (2010), A beszéd szupraszegmentális szerkezete, *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. *Akadémia Kiadó* (pp. 171-205.)
 - [38] Quatieri, T. F., & Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity. In Thirteenth annual conference of the international speech communication association.
 - [39] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech communication*, 71, 10-49.
 - [40] Sándor, F. (2007). Valószínűség-számítás és matematikai statisztika. Miskolci Egyetem Földtudományi Kar.
 - [41] József, R., & Julianna, T. Valószínűség-számítás és matematikai statisztika, Nemzeti Tankönyvkiadó, 2008.
 - [42] Suthaharan, S., & Suthaharan, S. (2016). Support vector machine. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, 207-235.
 - [43] Numerikus példák a tartógépekre. <https://gyires.inf.unideb.hu/GyBITT/19/ch05s04.html> megtekintés: 2023.10.27
 - [44] Benesty, J., Chen, J., Huang, Y., Cohen, I. (2009). Noise reduction in speech processing (Vol. 2). *Springer Science & Business Media*.

- [45] Claesen, M., De Smet, F., Suykens, J., & De Moor, B. (2014). *EnsembleSVM: A library for ensemble learning using support vector machines*. arXiv preprint arXiv:1403.0745.
- [46] Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. *In Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 856-863).
- [47] Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808.
- [48] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, No. 2, pp. 1137-1145).
- [49] Tájékoztató, N. A. I. H. (2016). A Nemzeti Adatvédelmi és Információszabadság Hatóság tájékoztatója a munkahelyi adatkezelések alapvető követelményeiről.
- [50] Johansson, C. (2016). The'58 Sound: What Makes a Live Engineer Choose the Shure Beta58 or the Shure SM58?
- [51] SM58 adatlap, <https://pubs.shure.com/guide/SM58/en-US> megtekintés: 2023.10.22
- [52] Tibor, F. Érzelmek automatikus felismerése a beszéd akusztikus jellemzői alapján. *Beszédkutatás* 2007, 151.
- [53] Kiss, G., Sztahó, D., & Vicsi, K. (2019). Depression State Assessment: Application for Detection of Depression by Speech. *In INTERSPEECH* (pp. 966-967).
- [54] Kiss, G., & Vicsi, K. (2017). Mono-and multi-lingual depression prediction based on speech processing. *International Journal of Speech Technology*, 20, 919-935.
- [55] Hofmann, M., Klinkenberg, R. (Eds.). (2016). *RapidMiner: Data mining use cases and business analytics applications*. *CRC Press*.