Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar

A DEPRESSZIÓS BESZÉD AKUSZTIKAI-FONETIKAI JELLEMZŐINEK VIZSGÁLATA

Ph.D. értekezés
BME-VIK Informatikai Tudományok Doktori Iskola

Kiss Gábor
okl. mérnök-informatikus

Témavezető:
Vicsi Klára, DSc

Budapest, 2019

Rövid tartalmi összefoglaló

A depresszió korunk egyik legelterjedtebb, gyógyítható betegsége, azonban jelenleg nem ismert olyan objektív biomarker, ami alapján megfelelő pontossággal felismerhető lenne a depressziós állapot, ami megnehezíti a depresszió időben való felismerését. Kutatásom célja az volt, hogy bemutassam miképp lehetséges a beszédjelfeldolgozás alapján a depressziós állapot gépi felismerését, súlyossága szerinti becslését megvalósítani.

Az értekezésem felépítése a következő: a bevezetésben röviden bemutatom a depressziós állapot beszédjelfeldolgozás alapú fontosabb megválaszolandó kérdéseit. Bemutatom, mi a depresszió, és hogyan diagnosztizálják, miért okozhat a depressziós állapot elváltozást a beszédproduktumban. Ismertetem az orvosi gyakorlatban használt két legelterjedtebb depresszió súlyosságát leíró skálát. Majd összefoglalom, hogy kutatásom kezdetekor milyen eredmények voltak ismertek a tématerületben angol és német nyelv esetében.

Az ezután következő fejezetekben írom le a saját munkámat. Bemutatom, hogy milyen beszédadatbázisokat és milyen módszereket hoztam létre illetve használtam a kutatásom során. Magyar nyelvű depressziós beszédadatbázis segítségével, statisztikai vizsgálatok alapján, magyar nyelve először igazolom, hogy a depressziós állapot hatására megváltoznak a beszéd egyes akusztikai-fonetikai jellemzőinek értékei. Megvizsgálom, hogy a statisztikai úton kiválasztott jellemzők alapján, szupport vektor gépek segítségével milyen pontossággal valósítható meg a depressziós állapot detektálása, illetve annak négy kategória szerinti elkülönítése. Bemutatom, hogy a nemek szerint eltérő modellek alkalmazása és optimalizálása javítja a depressziós állapot osztályozását. Regressziós modellek segítségével megvizsgálom, hogy mekkora hibával lehetséges a depressziós állapot súlyosság szerinti becslését megvalósítani. Bemutatom, hogy a beszédhangszintű szegmentálás használata, javítja a becslés pontosságát. Bemutatom továbbá, hogy az alkalmazott akusztikai-fonetikai jellemzőkészletből kiválasztható olyan részhalmaz, amely magyar és német nyelv esetén is szignifikáns korrelációt mutat a depresszió súlyosságával, és megvizsgálom, hogy a kiválasztott jellemzők segítségével, mekkora pontossággal becsülhető a depressziós állapot súlyossága többnyelvű modell segítségével. Vizsgálataim idején, ebben a témában, ilyen mély elemzés a szakirodalomban még nem volt olvasható. Bemutatom, a European Space Agency (ESA) Coala projekt keretei közt kidolgozott és megvalósított, a beszélő depressziós állapotát beszédminta alapján többnyelvű modell segítségével becslő rendszeremet. Az értekezésem végén összesítem az elért eredményeimet, téziseimet.

Abstract

Depression is one of the most prevalent and curable illnesses of our time, but currently there is no objective biomarker that can be used to detect depression with convenient accuracy, which makes it difficult to recognize depression in time. The purpose of my research was to show how it is possible to recognize depression and estimate its severity based on speech signal processing.

The structure of my thesis is the following: in the introduction I briefly introduce the most important questions of speech processing regarding a depressed state. I define depression and describe the method of diagnosis, and how a depressive disorder can cause change in the in the speech product, furthermore I describe the two most common scales used in medical practice to express the severity of depression. Then, I will summarize the results known in this research filed at the beginning of my research, in case of English and German speaking subjects.

In the chapters that follow I write about my own work. I describe the speech databases and methods I used during my research.

Based on the Hungarian depressive speech database and statistical analyses, I confirm that the values of certain acoustic-phonetic parameters of speech change as a result of depression. I investigate the accuracy of the detection of depression and its four categories separation based on statistically selected paramaters using support vector machines. I show that the use models created by gender, and their optimalization improves classification of depression. With the help of regression models, I investigate the potential error while determining the severity of depression. I show that using speech-sound-level segmentation when calculating acoustic-phonetic parameters improves the accuracy of the estimation. I show that from the acoustic-phonetic parameter set used, a subset can be selected that shows a significant correlation with the severity of depression in case of Hungarian and German, moreover I examine what accuracy can be achieved while estimating the severity of depression applying the selected parameters, using a multilingual model. I present my system for estimating the depressive state of the speaker using a multilingual model, developed and implemented within the framework of the European Space Agency (ESA) Coala project.

At the end of my dissertation I summarize my achievements and my theses.

Köszönetnyílvánítás

Szeretném megköszönni témavezetőmnek, Vicsi Klárának sok éves útmutatását, türelmét, bíztatását. Beszédakusztikai laboratórium csoport összes dolgozójának – elsősorban Sztahó Dávid, Tulics Miklós Gábriel, Nagy Ildikó, Tündik Máté Ákos, Szaszák György - és jelenlegi illetve korábbi hallgatójának észrevételeit, kritikáit és segítségét munkám során.

Szeretném megköszönni Beszédkommunikáció és Intelligens Interakciók Laboratórium több dolgozójának a kritikáit és ötleteit téziseim megvitatása során, névszerint: Németh Géza, Mihajlik Péter, Zainkó Csaba, Csapó Tamás Gábor, Tarján Balázs.

Szeretném megköszönni Simon Lajos és Dér Bálint pszichiátereknek a segítségét a felvételek gyűjtése során.

Szeretném megköszönni édesanyámnak, hogy lektorálta az értekezésemet. Szeretném megköszönni feleségemnek és fiaimnak, hogy mindenben támogattak kutatásaim elvégzése során.

Tartalom

1. Bevezetés ... 8

2. Depresszió ... 12

2.1. Depresszió meghatározása ... 13

2.2. Depresszió diagnosztizálása ... 14

2.3. A beszéd mint a depresszió objektív biomarkere ... 15

2.4. Depressziós súlyosságot értékelő skálák ... 16

3. A depresszió és a beszéd kapcsolatának külföldi vizsgálatai ... 17

3.1. Jelentősebb depressziós beszédadatbázisok ... 17

3.2. Szakirodalmi eredmények a depresszióra jellemző akusztikai-fonetikai jellemzőkkel kapcsolatban ... 18

3.3. Depressziós állapot detektálásának és súlyosság szerinti becslésének eredményei a szakirodalomban ... 21

4. Kutatásom során létrehozott és felhasznált beszédadatbázisok és módszertan ... 23

4.1. Felhasznált depressziós beszédadatbázisok ... 23

4.1.1. Magyar Depressziós Beszédadatbázis ... 24

4.1.2. Német Depressziós Beszédadatbázis ... 25

4.1.3. Olasz Depressziós Beszédadatbázis ... 25

4.1.4. Depressziós beszédadatbázisok összehasonlítása ... 25

4.2. Statisztikai módszerek ... 27

4.2.1. Kolmogorov-Szmirnov-próba ... 28

4.2.2. Levene-próba ... 28

4.2.3. Független kétmintás t-próba ... 29

4.2.4. Pearson korreláció ... 30

4.2.5. Statisztikai vizsgálataim ... 30

4.3. Osztályozási és regressziós módszerek ... 31

4.3.1. SVM matematikai hátterének rövid leírása ... 32

4.3.2. SVR matematikai hátterének rövid leírása ... 34

4.3.3. Modellépítés és tesztelési eljárások 35
4.3.4. Kiértékelési módszerek 35
4.3.5. Statisztikai alapú jellemzővektor kiválasztása 36
4.4. Kutatás során vizsgált akusztikai-fonetikai jellemzők 37
4.4.1. Felhasznált alapjellemzők leírása 39
4.4.2. A felhasznált akusztikai-fonetikai alapjellemzők mérési helye 45
4.4.3. Származtatott akusztikai-fonetikai jellemzők 46
4.5. Felhasznált programok 47
5. A depressziós állapot hatásának statisztikai vizsgálata a beszéd akusztikai-fonetikai jellemzőire 48
5.1. Depressziós és egészséges beszélők akusztikai-fonetikai jellemzőinek összehasonlítása 48
5.2. A depresszió és az akusztikai-fonetikai jellemzők közötti korreláció vizsgálata 53
5.3. Statisztikai vizsgálataim eredményeinek összegzése 56
6. A depressziós állapot detektálásának beszédjel feldolgozásán alapuló vizsgálataim 57
6.1. Kiindulási eredményeim a depressziós állapot gépi detektálásában 58
6.2. Depressziós állapot detektálásának javítása jellemzővektor kiválasztással 60
6.3. Depressziós állapot detektálásának javítása nemek szerinti eltérő modellek alkalmazásával 62
6.4. Depressziós állapot osztályozási eredményeim összegzése 64
7. A depressziós állapot súlyosság szerinti becslésének beszédjel feldolgozásán alapuló vizsgálataim 66
7.1. Depressziós állapot súlyosságának becslése beszédhangszintű szegmentálás felhasználása nélkül 67
7.2. Depressziós állapot súlyosságának becslése beszédhangszintű szegmentálás felhasználásával 71
7.3. Depressziós állapot detektálása regressziós eljárás segítségével 75
7.4. Depressziós állapot súlyosság szerinti becslésére kapott eredményeim összegzése 78

8. A depressziós állapotot tükröző akusztikai-fonetikai jellemzők nyelvfüggőségének vizsgálatakor kapott eredményeim 80

8.1. A depresszió és az akusztikai-fonetikai jellemzők közötti korreláció nyelvfüggőségének vizsgálata 81

8.2. Depressziós állapot súlyosságának többnyelvre érvényes becslése 83

8.3. Depressziós állapot súlyosságának több nyelvre érvényes becslésére kapott eredményeim összefoglalása 87

9. Az eredmények gyakorlati alkalmazhatósága 89

10. Összegzés és téziseim összefoglalása 91

11. Rövidítések, jelölések 96

12. Irodalomjegyzék 97

13. A szerző tudományos közleményei 106

13.1. Értekezéshez szorosan kapcsolódó publikációk 106

13.2. A tématerületen megjelent egyéb publikációk 107

1. Bevezetés

Az emberi beszéd bonyolult folyamat, ami sokrétű információt hordoz, és a kommunikáció egyik legalapvetőbb eszköze. Az emberi beszédnek mint kommunikációs eszköznek, elsődleges célja a gondolatok szimbolikus úton történő közvetítése a nyelvi tartalom segítségével [1], azonban ezenfelül további információkat is képes közvetíteni. Ilyen további információk például a beszélő érzelmi állapota, attitűdje, és képes utalni a beszélő fiziológiai állapotára is, amennyiben az állapot hatással bír a létrejött beszédproduktumra, vagyis hatással bír a beszédkeltés folyamatára [2][3][4][5][6][7].

A beszédkeltés alapvető fiziológiai szervei: a tüdő, a légcső, a gége, a garat, a száj- és orrüreg. Működésüket az agy irányítja [8]. Ezek alapján azok a fiziológiai elváltozások detektálhatók a beszédjel feldolgozása során, amelyek vagy elváltozást okoznak a beszédkeltésért felelős szervekben (például: megfázás, tumor), vagy az agyi működésben okoznak elváltozást (például: Parkinson-kór vagy depresszió), esetleg egyszerre mindkettőben. Kutatásomban a depressziós állapot beszédre gyakorolt hatását és a megváltozott akusztikai-fonetikai jellemzők által lehetséges depressziós állapot detektálását és a depresszió súlyosságának becslését vizsgáltam.

Már 1921-ben megállapította Emil Kraeplin, a modern pszichológia egyik megalapozója, hogy a depressziós állapot hatására megváltozik az emberi beszéd [9]. A depresszió és a beszéd kapcsolata az 1980-as évekre már fontos kutatási területnek számít, és több akusztikai, illetve fonetikai paramétert kapcsolatba hoztak a depresszióval, mint például az átlagos alapfrekvencia érték, az alapfrekvencia tartománya, beszédtempó [10].

Kutatásom kezdetekor (2013) a depressziós állapot és a beszéd kapcsolatának vizsgálata, illetve a beszédjel feldolgozásán alapuló depressziós állapot gépi detektálásának lehetősége előtérbe került kutatási területnek számított. Alapvetően kétféle gépi detektálási módszert alkalmaztak a kutatók: a beszélő depressziós állapotának osztályozását, illetve a depresszió súlyosságának a becslését regressziós eljárással. Kétosztályos gépi tanuló eljárás alkalmazása esetén (nem depressziós/depressziós) elégséges lehet az orvosi diagnózis megléte a minták címkézéséhez. Azonban többosztályos osztályozás esetén, ami képes a depressziós állapot különböző kategóriáit is elkülöníteni súlyosság alapján, illetve regressziós eljárás megvalósítása esetén, elengedhetetlen valamilyen depressziós állapotot leíró skála. Többféle depresszió súlyosságát leíró skála létezik, ami megnehezíti az egyes kutatók eredményeinek az összehasonlítását a depressziós állapot és az emberi beszéd kapcsolatában, mivel ezek a skálák nem feleltethetők

meg egymásnak egyértelműen, így a különböző kutatók eredményei csak nehézkesen hasonlíthatók össze, mivel az egyes depressziós és nem depressziós bemondók beszédmintáit tartalmazó beszédadatbázisokban (továbbiakban depressziós beszédadatbázis) eltérő módon van jelölve a bemondók depresszió szerinti súlyossága [10].

A nemzetközi irodalom több beszédparamétert is megemlít, ami a depresszió hatására megváltozik (alapfrekvencia, szünetek hossza, intenzitás dinamikája) [11][12]. Azonban abban még nincs általános egyezség, hogy melyek az igazán jellemző paraméterek, illetve az adott beszédparamétereket hogyan érdemes mérni. Természetesen a mérési lehetőség nagyban függ a beszédadatbázis feldolgozottságától, ami lehet csupán egy egyszerű beszéd/nembeszéd szerinti [11][12], de akár egy pontos beszédhangszintű szegmentálása is az adatbázisnak. Az utóbbit értelemszerűen lényegesen időigényesebb és költségesebb megvalósítani, emiatt a legtöbb eddigi kutatásban nem alkalmaztak beszédhang szintű szegmentálást, viszont ebből kifolyólag egyes beszédparamétereket csak pontatlanul, nagyobb szórással tudtak megmérni, illetve egyes beszédparaméterek meg sem mérhetők a beszéd szegmentálása nélkül. Talán pont ezen okokból kifolyólag egyes beszédparaméterek értékei esetén eltérő tendenciákat mértek a kutatók a depressziós állapot hatására [10][11][13].

Másik alapvető probléma a különböző kutatásokban, hogy a női és a férfi beszédmintákat egyben kezelik [11][12]. A jelenleg létező depressziós beszédadatbázisok viszonylag kisméretűek: 30-160 beszélőtől tartalmaznak beszédmintákat [10], így ha külön vizsgálják a nőket és a férfiakat, akkor a vizsgált halmaz mérete tovább csökken. Ugyanakkor a saját kutatásaimból kiderült, a beszélő neme nagyban befolyásolhatja a beszéd egyes akusztikai-fonetikai jellemző értékeinek a megváltozását a depressziós állapot hatására. Ez az újszerű megfigyelésem egyezik a depresszióval foglalkozó pszichiáterek megfigyelésével, ugyanis a depresszió hatása különbözőképpen mutatkozik meg a férfiak és a nők esetében [14][15], illetve Hönig és társai is hasonló eredményre jutottak a 2014-ben publikált kutatásukban [16], ahol nemek szerint külön vizsgáltak a depresszió és beszéd akusztikai-fonetikai jellemzőinek összefüggését beszédhang szintű szegmentálás segítségével. Így ha közös modellt használunk a nők és a férfiak esetében, az problémát okozhat a depressziós állapot felismerésében, ami mindenképpen megoldandó feladat.

Továbbá a megfelelő beszédminta típusa is nyitott kérdés még, miszerint olvasott szöveg vagy spontán beszédminta alkalmasabb-e a beszélők depressziós állapotának detektálására, a depressziós állapotuk szerinti súlyosságuk becslésére.

Az eddigi nemzetközi vizsgálatok elsősorban angol és német nyelvű depressziós beszédadtabázisok beszédmintáinak elemzésével készültek. A depressziós állapot és a beszéd kapcsolatának mélyebb vizsgálata magyar nyelvű beszédminták elemzésével eddig még nem történt meg.

Kutatásom kezdetekor a depressziós állapot beszédre gyakorolt hatásának egynyelvű vizsgálati eredményei voltak csak ismertek. A kutatásomban bemutatott depressziós állapot beszédre gyakorolt hatásának nyelvek közötti összehasonlítása, a depressziós állapot súlyosságának nyelvfüggetlen becslése egészen új kutatási területnek számít, az ezirányú kutatási eredményem publikálásakor (2016, 2017), annak mélyebb vizsgálata nem található meg a nemzetközi szakirodalomban. Ezzel kapcsolatos komolyabb vizsgálatot egyedül Alghowinem és társai publikáltak az én publikálásommal egyidőben 2016-ban [17], ahol angol, amerikai angol és német nyelvű depressziós beszédadatbázisok segítségével vizsgálták a nem depressziós és súlyosan depressziós személyek elkülönítésének lehetőségeit a beszélő nyelvétől függetlenül.

A kutatási célkitűzéseimet ennek megfelelően alakítottam ki, és kutatási eredményeim nemcsak a magyar nyelvű vizsgálataim esetében számítanak újszerűnek, hanem több kutatási eredményem nemzetközi szinten is újszerűnek számítottak a publikálásukkor.

Dolgozatomban bemutatott kutatási eredményeimmel a depresszió beszédre gyakorolt hatásának mélyebb megértéshez, a depressziós állapot beszédalapú detektáláshoz és a depressziós állapot súlyosságának becsléséhez kívánok hozzájárulni. Megvizsgáltam a depressziós állapot hatását az emberi beszédre. Továbbiakban megvizsgáltam még a megváltozott akusztikai-fonetikai jellemzők által lehetséges beszédjelfeldolgozás-alapú depressziós állapot detektálását és a depressziós állapot súlyosság szerinti becslését magyar nyelv esetében. Emellett két- (magyar, német) és többnyelvű (magyar, német, olasz) vizsgálatokat is végeztem. Úgymint a depressziós állapot beszédre gyakorolt hatásának nyelvfüggetlen megnyilvánulásait és az eredményeim alapján több nyelvre alkalmazható depressziós állapot súlyosság szerinti becslésének lehetőségeit.

Továbbiakban megvizsgáltam még, hogy az akusztikai-fonetikai paramétereket milyen módon érdemes mérni a pontosabb detektálás és becslés érdekében. Megvizsgáltam a depressziós állapot beszédre gyakorolt hatásának nemek szerinti függését is.

Konkrét fő célkitűzéseim a következők:

- Olyan akusztikai-fonetikai paraméterek kiválasztása, amelyek értéke igazolhatóan megváltozik a depressziós állapot hatására magyar nyelv esetében. (I. téziscsoport).
- Megvizsgálni gépi tanuló eljárás segítségével, hogy automatikusan, kizárólag a beszédjel feldolgozása alapján, milyen pontossággal lehet osztályozni a bemondókat két (nem depressziós/depressziós), illetve négyfokozatú (nem depressziós/enyhe depresszió/közepes depresszió/súlyos depresszió) depressziós skálán magyar nyelv esetében (II. téziscsoport).
- Megvizsgálni gépi tanuló eljárás segítségével, hogy automatikusan, kizárólag beszédjel feldolgozás alapján, milyen pontossággal lehet becsülni az egyes bemondók depressziós állapotának súlyosságát magyar nyelv esetében (III. téziscsoport).
- Megvizsgálni, hogy a depressziós állapot hatásainak a beszédre gyakorolt realizációi mennyiben függetleníthetők a beszélő által használt nyelvtől. Milyen pontossággal lehetséges automatikusan detektálni, illetve becsülni a depressziós állapotot nyelvfüggetlen modell segítségével (IV. téziscsoport).
- Olyan rendszer kidolgozása és implementálása, amely képes eltérő anyanyelvű vizsgált személyek depressziós állapotának megfigyelésére beszédjel feldolgozás alapján.

Kutatásom fő célja egy olyan rendszer alapjainak lefektetése és kidolgozása, amely képes a beszélő depressziós állapotának súlyosságát becsülni kizárólag beszédjel feldolgozás alapján nem invazív módon, egy rövid 8-10 mondatos szöveg felolvasásának segítségével.

2. Depresszió

A depresszió korunk egyik legelterjedtebb, gyógyítható betegsége [18], ám diagnosztizálása szaktudást igényel, és így a betegség felismerése a társadalom egy szűk rétegére hárul. A depresszió a World Health Organization (WHO) 2002-es felmérése alapján a negyedik, 2012-es felmérése alapján már a harmadik leggyakoribb betegség világszerte, és a szervezet előrejelzései alapján 2030-ra a második leggyakoribb betegség lehet [19][20].

A depresszió súlyossága nagyban befolyásolja az ebben szenvedő egyének életminőségét. A depressziós állapot hatására, illetve annak súlyosbodásával az ettől szenvedő betegek elveszíthetik teljes motiváltságukat, így képtelenek lehetnek ellátni a napi teendőiket, elvégezni a rájuk bízott munkát [21]. Tehát a depresszió hatása nemcsak az ettől szenvedő betegek életében jelenik meg mint probléma, hanem komoly gazdasági károkat is okoz. Egy 2010-es tanulmány kimutatta, Európa 30 országát vizsgálva, hogy a depresszió 92 milliárd eurós veszteséget okozott a gazdaságban, melyből 54 milliárd euró a termelés kiesés következtében keletkezett [22]. A gazdasági károk mellett kimutatható, hogy a depresszió súlyosbodásával megnő az öngyilkosság kockázata, illetve az öngyilkosságok fele vezethető vissza depressziós okokra [23][24].

Depresszió hatására megváltoznak az emberi beszédproduktum egyes jellemzői, amelyek számszerűsíthetők és mérhetők. Emiatt lehetőség nyílik a depresszió beszédjel alapú gépi detektálásra, ami megkönnyítheti, illetve szélesebb körben lehetővé teheti a betegség diagnosztizálását. Ezen okok miatt fontos kutatási terület a depressziós állapot beszédjel alapú felismerése és súlyosságának becslése.

2.1. Depresszió meghatározása

A depresszió, mint betegség definícióját a The Diagnostic and Statistical Manual of Mental Disorders (DSM) definiálja. Abban az esetben beszélhetünk általánosságban depresszióról, ha vizsgált személynél legalább két héten át vagy hosszabb ideig a következő tünetek közül négy vagy annál több fennáll:

- pszichomotoros gátlás vagy remegés
- csökkent kognitív állapot vagy megnövekedett határozatlanság
- ájulás vagy energiaveszteség
- insomnia vagy hypersomnia
- jelentős súlyveszteség vagy súlygyarapodás
- jelentős negatív önértékelés vagy alaptalan túlzott bűntudat
- gyakori gondolatok a halálról vagy gyakori öngyilkos gondolatok.

A klinikai pszichiátria megkülönbözteti a depresszió különböző fajtáit. A depressziós állapotot két fő kategória szerint sorolják be, úgymint unipoláris és bipoláris depresszió. Az unipoláris depresszió, azaz egyfázisú depresszió esetén csak a rossz hangulat irányába van eltérés. A bipoláris depresszió esetén mind a rossz, mind pedig a jó hangulat túlzott felfokozottsága felváltva követi egymást. A fő kategóriákon belül további típusokat definiálnak, ezek közül az unipoláris depresszió esetén a főbb altípusok: a major depresszió (klinikai depresszió), illetve a disztímia. A bipoláris depresszió esetén a főbb altípusok: az I. típusú bipoláris, más néven mániás depresszió, II. típusú bipoláris más néven hipománia, illetve a ciklotímia [25].

A depresszió pontos élettani okaiban még nincs általános egyezség. Azonban a depresszióra többnyire úgy tekintenek, mint a kortikális limbikus rendszer egyfajta diszfunkciója, vagyis a limbikus kéreg csökkent aktivitása és összekapcsolhatósága [26][27][28]. Bár a pontos élettani okok nem ismertek, azonban megállapították, hogy a depresszió kialakulásában genetikai hajlam, illetve környezeti tényezők játszanak elsősorban szerepet, úgymint a stressz, traumatikus életesemény, bezártság vagy a napfény hiánya [29].

A kutatásaim során az unipoláris depresszión belül a klinikai depresszió hatását vizsgáltam az emberi beszédre. Továbbiakban a klinikai depresszió helyett, az egyszerűség kedvéért, mindig a depresszió kifejezést használom.

2.2. Depresszió diagnosztizálása

A depresszió diagnózisának felállítása jelenleg egy szűk szakorvosi csoportra hárul. Ez a tény, figyelembe véve, hogy világszerte hány ember szenved ebben a betegségben [20], nagyban megnehezíti a depresszió időben való felismerését és a kezelés korai szakaszban való megkezdését, aminek következtében feleslegesen meghosszabbodhat és súlyossá válhat a beteg depressziós állapota [30].

A diagnózis felállításának nehézsége abból adódik, hogy jelenleg nem ismert a depressziónak pontos, objektív biomarkere. A diagnózis felállítása gyorsítható, egyszerűsíthető lehetne megfelelő objektív diagnosztikai eszközökkel, a vizsgált személy biológiai, fiziológiai és viselkedési jellemzőinek megfelelő vizsgálatával. Bár biológiai markerek széles választéka, mint az alacsony szerotonin szint, neurotranszmitter diszfunkció és genetikai rendellenességek összefüggésben állnak a depresszióval és az öngyilkossági hajlandósággal, azonban nem találtak még olyan a biomarkereket, amelyek megfelelő specifikussággal és szenzitivitással rendelkeznének a depresszió diagnóziáshoz a klinikai gyakorlatban. Ráadásul ezen biomarkerek mérése általában csak költséges úton és invazív beavatkozást igénylő módon végezhetők [31][32][33].

Ezen okok miatt a klasszikus (invazív beavatkozást igénylő) biomarkerek vizsgálatán kívül az elmúlt évtizedben egyre nagyobb hangsúlyt kapnak a nem invazív biomarkerek vizsgálata is, mint például az arcon megfigyelhető jellemző elváltozások a depresszió hatására [34].

A nem invazív biomarkerek közül az emberi beszéd ígéretes jelöltnek tűnik a depresszió diagnosztizálásnak segítségében, mivel egyrészt a beszédkeltés természetes, nem jelent különösebb megterhelést a vizsgált személy számára, a beszédjel elemzése olcsón elvégezhető, nem igényel különösebb beavatkozást és speciális tudás meglétét a vizsgált személy részéről. Másrészt a pszichiáterek a diagnózis során egyébként is felhasználják (szubjektíven) a vizsgált személy beszédkeltési sajátosságait, mivel a depressziótól szenvedő betegek beszédére jellemző a csökkent beszédtevékenység, a beszéd monotonitása, élettelen hangszínezete, lassulása [35][36].

A depresszió diagnosztizálását segítő objektív, olcsó, gyors és nem invazív eljárás elsősorban az elsődleges egészségügyi ellátásban lenne hasznos, mivel a depressziós tüneteket tapasztaló egyének leggyakrabban csak a háziorvosukhoz fordulnak a panaszaikkal, ugyanakkor felmérések igazolják, hogy a háziorvosok csupán az esetek felében képesek a depressziót

felismerni, így a depressziótól szenvedő egyént a megfelelő szakorvosi ellátáshoz irányítani [30][37][38].

2.3. A beszéd mint a depresszió objektív biomarkere

A beszédkeltés bonyolult, komplex folyamat, így a beszédproduktum létrejöttét több tényező is jelentősen befolyásolja, és a fiziológiai és a kognitív elváltozás jelentős, észlelhető akusztikai-fonetikai elváltozásokat okoz benne [39][40].

Az emberek a beszédkeltés során több motoros idegrostot használnak, mint bármilyen más mechanikai tevékenység során. Másodpercenként akár 20-25 beszédhangot is képesek vagyunk kiejteni, így a beszédkeltés a leggyorsabb, különálló motoros tevékenységnek számít. Beszédkeltésben több fő izomcsoport is szerepet játszik, úgymint a légző-, gége- és artikulációs izmok. A légzőizmok közé tartoznak a rekeszizom és a bordaközi izomcsoportok, illetve a gége kilenc porcát kontrolláló öt izomcsoport. Az artikulátorok közé tartoznak az állkapocs, az ajak, a nyelv, a lágy szájpadlás és garatösszehúzó izmok, melyek együttes működése mind szerepet játszik a végleges beszédproduktum kialakulásában [41][42].

A depresszió hatására megváltozik az izmok feszítettsége és kontrolláltsága, ezáltal megváltoznak a beszédproduktum prozódiai jellemzői. Az esetleges izomzavarok a gégében a hangszalagok viselkedését változtathatják meg, így a depresszió hatással bírhat az alapfrekvenciára, a beszéd érdességére, levegősségére, glotalizáltságára [43][44][45][46].

A depresszió hatására megváltoznak (csökkennek) az egyén kognitív képességei, mivel a depresszió hatással bír a depressziótól szenvedő személy munkamemóriájára. Beszédkeltés során a munkamemória felelős a fonológiai hurok megfelelő működéséért, mely segít kontrollálni az artikulációs rendszert. Mivel a depresszió hatással van a fonológiai hurokra, így a depressziótól szenvedő egyén beszédkeltése során különböző artikulációs és fonációs hibákat vét [47][48][49][50].

Emellett a kognitív képességek csökkenése befolyással bír a beszéd megtervezésre. Szignifikáns korreláció mutatható ki a depresszió súlyossága, illetve a beszédben található szünetek hossza között, amiből arra lehet következtetni, hogy a depresszió hatására nehezebbé válik a depressziótól szenvedő egyénnek a beszéd megtervezése vagy létrehozása [51][52][53].

2.4. Depressziós súlyosságot értékelő skálák

A depressziós állapot súlyosságának mérésére elsősorban két, széles körben elterjedt skála létezik. A Hammilton Rating Scale for Depression (HAM-D) és a Beck Depression Inventory, illetve annak módosított, javított változata a Beck Depression Inventory II. Mivel a Beck Depression Inventory II skála használata egyértelműen kiszorította a Beck Depression Inventory skála használatát, és a dolgozatomban csak a Beck Depression Inventory II skálát tárgyalom és használom, ezért az egyszerűség kedvéért a Beck Depression Inventory II skálára, mint BDI-skálára hivatkozom a továbbiakban. Mindkét skála (HAM-D és BDI) széles körben elterjedt és elfogadott a depresszió súlyosságának leírására, ugyanakkor lényeges különbségek találhatók a két skála tulajdonságaiban [54][55][56].

A leglényegesebb különbség a két skála között, hogy míg a HAM-D-értéket kizárólag a pszichiáter állapíthatja meg a diagnózis felállítása során, addig BDI értéket bárki képes meghatározni egy önkitöltős kérdőív segítségével. Emellett a HAM-D-érték megállapításához körülbelül 20-30 percre van szüksége, addig a BDI értéket 5-10 perc alatt is meg lehet határozni. A HAM-D elsősorban a vizsgált személy neuro-vegetatív tüneteit vizsgálja, addig a BDI főleg a negatív önértékelési tünetekre fókuszál.

A HAM-D-skála alapesetben 17 tünetet vizsgál, ahol az egyes tünetekre a pszichiáter 0-2, 0-3 vagy 0-4 tartományban adhat értéket a tünet fontosságának függvényében, így végül 0-48 tartományban ad értéket a vizsgált személyről, az egyes tünetekre adott pontszámok egyszerű összegzésével. A kapott pontszám alapján 5 kategóriát különböztet meg a depresszió súlyossága alapján. A HAM-D-értékek és a hozzárendelt kategóriák a 1. táblázatban láthatók.

A BDI-skála 21 kérdésből áll, és mindegyik kérdésre 0-3 tartományban kapható pontszám, így végül 0-63 tartományban ad értéket a vizsgált személyről, az egyes kérdésekre kapott pontszámok egyszerű összegzésével. A kapott pontszám alapján a BDI-skála 4 kategóriát különböztet meg a depresszió súlyossága alapján. A BDI értékek és a hozzárendelt depresszió súlyossága szerinti kategóriák a 1. táblázatban láthatók.

1. táblázat. A HAM-D- és a BDI-skálák által definiált depresszió súlyossága szerinti kategóriák leírása

Kategória neve	HAM-D-érték	BDI érték
nem depressziós	0-7	0-13
enyhe depresszió	8-13	14-19
közepes depresszió	14-18	20-28
súlyos depresszió	19-23	29-63
nagyon súlyos depresszió	23-48	nem értelmezett

3. A depresszió és a beszéd kapcsolatának külföldi vizsgálatai

Ebben a fejezetben röviden összefoglalom, hogy kutatásom kezdetekor (2013) milyen depressziós beszédadatbázisok léteztek. Milyen eredmények voltak ismertek a depresszió és az akusztikai-fonetikai jellemzők között, és milyen eredményeket értek el a depressziós állapot detektálása, illetve a depressziós állapot súlyosságának becslése terén gépi tanuló eljárásokat alkalmazva a beszédjel feldolgozása alapján.

Kutatásom kezdetekor az európai nyelvek közül elsősorban angol nyelvű vizsgálatok léteztek a beszéd és a depressziós állapot kapcsolatáról. 2013-ban publikálták az első jelentős, nem angol nyelvű depressziós beszédadatbázist, ahol a beszédminták német anyanyelvű bemondóktól származtak.

Ahogy arra már a bevezetőben is utaltam az egyes eredmények nehezen összehasonlíthatók, mivel az egyes depressziós beszédadatbázisok alapjaikban eltérőek, mind az adatbázisban szereplő bemondók számában és depressziós állapotuk szerinti súlyosságuk eloszlásában, mind a depressziós állapotot leíró skála használatában, illetve a rögzített beszédhang típusában (szabad szöveg, interjú, olvasott, kitartott hangok vagy hanggyakorlatok) [10].

3.1. Jelentősebb depressziós beszédadatbázisok

Az első jelentősebb európai depressziós beszédadatbázist 2000-ben publikálták France és társai. Az adatbázisban 59 (38 nő; 21 férfi) depressziós személytől és 34 egészséges személytől (10 nő és 24 férfi) találhatók beszédminták. A depressziós csoportba csak közepes vagy súlyosabb depressziós személyektől gyűjtöttek mintákat (BDI > 19), míg az egészséges csoportba csak egészséges személyektől gyűjtöttek mintákat (BDI < 14), vagyis enyhén depressziós

személyektől származó beszédmintát nem tartalmaz a beszédadatbázis. A bemondóktól szabad szöveget és interjú típusú felvételeket rögzítettek [57].

2004 és 2013 között még 7 másik jelentősebb beszédadatbázist publikáltak, ami egészében vagy részben depressziós és egészséges bemondók beszédmintáit tartalmazza. Mindegyik depressziós beszédadatbázisban angol nyelvű bemondók beszédmintáit rögzítették. Az egyes beszédadatbázisokban a bemondók száma 10 és 165 között változik. A depressziós állapot leírására három beszédadatbázis esetében a bemondók depressziós állapotuk szerinti súlyosságát nem jelölték, csak a depressziós állapot meglétét [58][59][60]. Egy depressziós beszédadatbázis esetében a BDI-skálát [61], míg három másik depressziós beszédadatbázis esetében a HAM-D-skálát használták a depresszió súlyosságának leírására [62][63][64], így ezekben a depressziós beszédadatbázisokban megadták a bemondók depressziós állapotuk szerinti súlyosságát is. Eltérő ezekben a depressziós beszédadatbázisokban még a depressziós és egészséges bemondók aránya, a bemondók depressziós állapotuk szerinti súlyosságuknak eloszlása, illetve három beszédadatbázis esetében csupán depressziós bemondóktól tartalmaz beszédmintákat a beszédadatbázis [62][63][64]. Megjelenik a beszédadatbázisokban az olvasott beszédminta használata is [58][60][61][62][64], és két adatbázis esetében különböző beszédhang gyakorlatokat is rögzítettek [59][64]. 2013-ban publikálták az első jelentős nem angol nyelvű depressziós beszédadatbázist. Az adatbázisban német anyanyelvű beszélők beszédét rögzítették. Az eddigi beszédadatbázisokhoz képest a német nyelvű depressziós beszédadatbázis a nagyobb méretű beszédadatbázisok közé sorolható a maga 150 beszélőjével, és a depresszió súlyosságának széles skálájából tartalmaz bemondókat az egészséges állapottól a súlyos depressziós állapotig. A depressziós állapot súlyosságának leírására a BDI-skálát használták, és olvasott, valamint interjú típusú bemondásokat is tartalmaz ~~a beszédadatbázis~~ [65][66].

3.2. Szakirodalmi eredmények a depresszióra jellemző akusztikai-fonetikai jellemzőkkel kapcsolatban

Már a XX. század második felében több érzeti változást figyeltek meg a beszédben. A pszichiáterek és a különböző kutatók a depressziós emberek vizsgálatakor azt figyelték meg, hogy a depressziós emberek beszéde monotonabb, fémesebb az egészséges emberek beszédéhez képest [44][67][68][69]. Ezek az érzeti megfigyelések számszerűsíthetők, mérhetők a beszéd egyes akusztikai-fonetikai jellemzőinek segítségével, így a mai depressziós beszédadatbázisok megjelenésével lehetőség nyílt a megváltozott akusztikai-fonetikai

jellemzőket megmérni, azokat összehasonlítani egészséges kontrollcsoport beszédmintáiból nyert akusztikai-fonetikai jellemzők értékeivel. A depressziós állapot és az akusztikai-fonetikai jellemzők vizsgálatához statisztikai vizsgálatokat végeztek, hogy megállapítsák a depressziós beszélők beszédmintái alapján, mely akusztikai-fonetikai jellemzők értékei mutatnak szignifikáns eltérést az egészséges beszélők beszédmintáiból nyert akusztikai-fonetikai jellemzőihez képest. A vizsgálatokat angol és német egynyelvű depressziós beszédadatbázisok segítségével végezték.

Bár több akusztikai-fonetikai jellemző esetében (intenzitás dinamikája, alapfrekvencia dinamikája, szünetek hossza, jitter, formánsfrekvenciák) egyes kutatások szignifikáns eltéréseket mutattak ki a depressziós beszélők beszédét vizsgálva, azonban ezek közül az akusztikai-fonetikai jellemzők közül a legtöbb esetében nincsen nemzetközi egyetértés a depressziós állapot és a vizsgált akusztikai-fonetikai jellemzők értékének kapcsolatában.

Ennek a jelenségnek több oka is lehetséges. Maga az emberi beszéd rendkívül változatos, nagy a természetes varianciája a beszéd akusztikai-fonetikai jellemzőinek mind egy adott beszélőn belül, mind egyes beszélők között még egészséges beszélők esetében is. Emiatt szignifikáns eltérés kimutatása csak viszonylag nagyméretű depressziós beszédadatbázisok segítségével lehetséges. Ugyanakkor a vizsgálatokban használt egyes depressziós beszédadatbázisok lényeges jellemzőikben eltérnek egymástól, úgymint a beszélők száma, beszélők neme és depressziós állapotuk súlyossága szerinti eloszlásuk, a vizsgált beszédminták típusa és a beszédmintában szereplő beszédhangok sorozata. Emellett az egyes depressziós beszédadatbázisok leírásánál nem egyértelmű, hogy a depressziós személyek nem szenvednek-e más beszédhangot befolyásoló betegségben, hiszen a depresszió gyakori kísérő tünete más olyan betegségeknek (például: Parkinson-kór), amelyek szintén befolyásolják a létrejött beszédproduktumukat, így az akusztikai-fonetikai jellemzők értékeit.

A másik alapvető probléma, hogy nincs megegyezés abban sem, hogy az egyes akusztikai-fonetikai jellemzőket pontosan hogyan és hol kell mérni, ami nagyban befolyásolja a mért akusztikai-fonetikai jellemző értékét. Ez a jelenség megnehezíti az egyes eredmények összehasonlíthatóságát.

Ezek az okok magyarázatot adhatnak arra, hogy az egyes vizsgálatok eredményei néha egymásnak ellentmondók, vagyis az egyes akusztikai-fonetikai jellemzők vizsgálata során képesek kimutatni szignifikáns kapcsolatot a depressziós állapottal, míg más vizsgálatok

esetében nem találnak szignifikáns kapcsolatot a depressziós állapottal ugyanezen akusztikai-fonetikai jellemzőket vizsgálva.

Például egyes kutatásokban kimutatták az alapfrekvencia átlagos értékének és dinamikájának csökkenését depressziós beszélők esetében [62][70], míg más kutatásokban ugyanezt nem tudták kimutatni [53][64]. Ugyanez a jelenség megfigyelhető még számos más akusztikai-fonetikai jellemző esetében, úgymint az intenzitás átlagos értéke [71][53] (ráadásul ezekben az eredményekben változás iránya a depresszió hatására is egymásnak ellentmondó), jitter [72][46], spektrális meredekség [72][46], formánsfrekvenciák (első és második) átlagos értéke [57][62][64][73].

A különböző kutatások eredményei lényegében egyedül a beszédsebesség csökkenésében értenek egyet a depressziós állapot hatására, bár az okok vizsgálata között az egyes kutatók eltérő eredményre jutottak (hosszabb beszédhangok, megnövekedett szünethosszok, gyakoribb szünetek) [13][43][47][60][62][64]. Az itt tárgyalt akusztikai-fonetikai jellemzőket a 4.1. fejezetben részletesen leírom.

A 2. táblázatban összefoglalom, hogy a fontosabb akusztikai-fonetikai jellemzők esetében milyen tendenciák voltak ismertek a depressziós állapot hatására a kutatásom kezdetekor angol és német nyelvű depressziós beszédadatbázisok vizsgálata alapján. A táblázatban csak olyan fontosabb akusztikai-fonetikai jellemzőt tüntetek fel, ahol a kutatások alapján a tendencia egyértelmű, annak ellenére, hogy nem minden esetben tudtak szignifikáns eltérést kimutatni.

2. táblázat. Kutatásom kezdetekor ismert tendenciák a depresszió hatására a beszéd akusztikai-fonetikai jellemzőin

Jellemző	Változás iránya
alapfrekvencia dinamikája	↓
alapfrekvencia átlaga	↓
intenzitás dinamikája	↓
szünetek átlagos hossza	↑
szünethosszok varianciája	↑
beszédsebesség	↓
artikulációs sebesség	↓
spektrális meredekség	↑
első formánsfrekvencia átlagos értéke	↑
második formánsfrekvencia átlagos értéke	↓

3.3. Depressziós állapot detektálásának és súlyosság szerinti becslésének eredményei a szakirodalomban

Kutatásom kezdetekor alapvetően csak a depressziós állapot meglétének gépi detektálási lehetőségeit vizsgálták angol nyelv esetében, vagyis a kétosztályú (nem depressziós/depressziós) osztályozási lehetőségeket. Ez elsősorban abból fakadt, hogy a kezdeti depressziós beszédadatbázisok esetében csupán a depressziós állapot meglétét jelölték a beszédadatbázisokban, a bemondók depressziós állapotának súlyosságát nem, illetve a beszédadatbázisok kisméretűek voltak, kevés lett volna a beszédminták száma többkategóriájú depresszió súlyosság szerinti elkülönítésére, vagy a depresszió súlyossága szerinti becslésére.

Depressziós állapot detektálása esetén (kétosztályú osztályozás) 50-79%-os pontosságot értek el a különböző depressziós beszédadatbázisok esetében olvasott vagy spontán beszédmintákkal a jellemzők széles skáláját felhasználva a különböző kutatásokban [12][59][74][75][76][79][80]. Az eredmények itt is nehezen összehasonlíthatók az előző fejezetben tárgyaltak miatt. Az osztályozások megvalósításához elsősorban szupport vektor gépeket és gauss mixture modell gépi tanuló eljárásokat használtak. Az osztályozási eljárások összehasonlításánál az SVM használatával érték el a legjobb eredményeket [80]. Azonban feltehetőleg a depressziós beszédadatbázisok relatív alacsony mintaszáma miatt nem kaptak jelentős különbségeket az egyes gépi tanuló eljárások használatában [80].

Az újabb, több mintát tartalmazó depressziós beszédadatbázisok megjelenésesével, valamint az adott depresszió súlyosságát leíró skála segítségével nemcsak a depressziós állapot meglétét rögzítették, hanem a beszélő depressziós állapota szerinti súlyosságnak fokát is, megjelent a depresszió súlyossága szerinti több kategóriájú depressziós állapotot megvalósító osztályozások vizsgálata is. HAM-D-skála alapján 5 kategóriájú depressziós állapot osztályozása esetében 36-50%-os osztályozási pontosságot értek el a különböző depressziós beszédadatbázisokon [47][77][78].

A 2013-as „Audio/Visual Emotion Challenges” (AVEC-2013) versenyben a bemondók depressziós állapotuk szerinti súlyosságukat kellett regressziós eljárás segítségével becsülni a BDI-skála alapján. A versenyhez német nyelvű depressziós beszédadatbázis állt rendelkezésre [65]. Az adatbázisban 150 beszélőtől voltak beszédminták, amit 50/50/50 arányban felosztottak „train”/”development”/”test” halmazokra, és a „train” és „development” halmazokhoz

megadták a beszélők depressziós állapotuk szerinti súlyosságát a BDI-skála szerint. Kiindulási („baseline”) eredményként 14,12 RMSE (root mean square error) hibaérték volt megadva, ami egy relatív gyenge eredménynek mondható, mivel az adatbázisban található személyek depressziós állapotuk szerinti súlyosságának a szórása 11,7 értékű volt. A verseny győztesének sikerült a teszthalmazt 8,68 RMSE hibaértékkel becsülnie [81].

Az eredmények alapján kijelenthető, hogy depressziós állapot meglétének automatikus detektálása (kétosztályú osztályozása), a depressziós állapot többkategóriájú osztályozása, illetve a depressziós állapot súlyosságának becslése beszédjel feldolgozás alapján még nem számított megoldott feladatnak a kutatásom kezdetekor. Fontos hangsúlyoznom, hogy az összes bemutatott szakirodalmi eredmény egynyelvű (nyelvspecifikus) vizsgálat eredménye, és nemek szerinti közös modellt alkalmaztak a depressziós állapot detektálásában, illetve súlyosságának becslésében.

4. Kutatásom során létrehozott és felhasznált beszédadatbázisok és módszertan

Ebben a fejezetben bemutatom a kutatásaim során felhasznált beszédadatbázisokat, alkalmazott statisztikai, osztályozási és regressziós módszereket, a felhasznált és vizsgált akusztikai-fonetikai jellemzőket.

A 4.1. fejezetben bemutatom a felhasznált különböző nyelvű (magyar, német és olasz) depressziós beszédadatbázisokat. Azoknak megadom a részletes leírását, hogy mire használtam őket, és összehasonlítom a bemutatott depressziós beszédadatbázisokat főbb jellemzőikben.

A 4.2. fejezetben bemutatom az általam használt statisztikai módszereket, amelyeknek segítségével átfogó és mély elemzést valósítottam meg a depressziós állapot és az akusztikai-fonetikai jellemzők kapcsolatában. A részletes statisztikai elemzést elsősorban a magyar nyelvű depressziós adatbázison végeztem, azonban statisztikai eszközök használatával megvizsgáltam a depressziós állapot hatásának nyelvfüggetlen megnyilvánulásait az akusztikai-fonetikai jellemzőkön a magyar és német nyelvű depressziós beszédadatbázisokat felhasználva.

A 4.3. fejezetben tárgyalom a kutatásaim során alkalmazott osztályozási és regressziós eljárásokat, továbbá ebben a fejezetben tárgyalom a modellek építése és tesztelése során alkalmazott módszereimet, illetve hogy a modellek pontosságának leírásához milyen kiértékelési leírójellemzőket alkalmaztam. Ebben a fejezetben mutatom be az általam alkalmazott statisztikai alapú jellemzővektor kiválasztási módszereket.

A 4.4. fejezetben bemutatom a kutatásaim során felhasznált akusztikai-fonetikai jellemzőket, azokat részletesen ismertetem. Megadom a számításuk módját és összefoglalom, hogy végül milyen származtatott akusztikai-fonetikai jellemzők segítségével végeztem a statisztikai vizsgálataimat és osztályozási, illetve regressziós modelljeim tanítását.

A 4.5. fejezetben összefoglalom, hogy az egyes módszereket milyen külső programokkal, eszközökkel valósítottam meg.

4.1. Felhasznált depressziós beszédadatbázisok

Kutatási célkitűzéseim között szerepelt a depressziós állapot hatásának vizsgálata az emberi beszédproduktumra és az ezáltal lehetővé váló depressziós állapot becslése magyar nyelv esetében, illetve ennek nyelvfüggetlen vizsgálata. Emiatt magyar, német és olasz nyelvű depressziós beszédadatbázisokat használtam a kutatásaim során.

4.1.1. Magyar Depressziós Beszédadatbázis

Kutatásom kezdetekor még nem létezett magyar nyelvű depressziós beszédadatbázis, így lehetőségem volt részt venni annak megtervezésében és a beszédminták gyűjtésében.

A magyar nyelvű beszédminták gyűjtését a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikájával együtt végeztük.

Fontos szempont volt az a Magyar Depressziós Beszédadatbázis megtervezésénél, hogy a depressziós állapot meglétének jelölése mellett, jelölve legyen az egyes bemondók depressziós állapotuk szerinti súlyossága is.

A depresszió súlyosságának leírásához a Beck Depression Inventory II (BDI) skálát választottuk [56], amely az adott személy depressziós súlyosságát írja le. A BDI-skála 0-tól 63-ig terjed, ahol a 0 a teljesen egészséges állapotot jelöli, míg a 63 a legsúlyosabb depressziós állapotot. A BDI-skála a depresszió súlyossága szerint négyféle kategóriát különbözet meg (3. táblázat).

3. táblázat. Beck Depression Inventory II által definiált kategóriák és a hozzájuk tartozó BDI értékek

BDI érték	Kategória neve
0-13	nem depressziós
14-19	enyhe depresszió
20-28	közepes depresszió
29-63	súlyos depresszió

Azért választottuk a BDI-skálát, mivel ez a skála megbízható, széles körben elterjedt és elfogadott az orvosi gyakorlatban, ugyanakkor az érték megállapítása gyorsan és egyszerűen elvégezhető, ami abból a szempontból volt fontos, hogy minél több beszédmintát lehessen gyűjteni a későbbi statisztikai vizsgálatokhoz és osztályozási és regressziós feladatok megvalósításához.

A beszédminták gyűjtésénél törekedtünk arra, hogy az adatbázisban szereplő bemondók lefedjék a depresszió súlyosságának különböző fokozatait, az egészséges állapottól az egészen a súlyos depresszióig.

Tervezési szempont volt még, hogy az adatbázisba csak olyan beszédminták kerüljenek kiválasztásra, ahol a beszélő nem szenved más olyan betegségben, ami hatással bír az emberi

beszédproduktumra. ~~Továbbá~~ A szakorvosok feladata volt validálni a BDI értéket, illetve ha felmerült, hogy a bemondó nem a valóságnak megfelelően töltötte ki a BDI kérdőívet, akkor joguk volt kizárni a beszédmintát az adatbázisból (erre egy esetben sem került végül sor).

A felvételek csendes helyiségben kerültek rögzítésre 44,1 kHz mintavételi frekvenciával, 16 bites kvantálással, csiptetős mikrofonnal. A magyar nyelvű depressziós beszédadatbázist a magyar nyelvű és többnyelvű vizsgálatok során is felhasználtam.

Erre az adatbázisra a későbbiek során, mint a „Magyar Depressziós Beszédadatbázis" hivatkozom.

4.1.2. Német Depressziós Beszédadatbázis

A német nyelvű depressziós beszédadatbázis része az „Audio-visual depressive language corpus"-nak (AViD-Corpus) [65].

Azért ezt az adatbázist választottam a kutatásom elején rendelkezésre álló külföldi depressziós beszédadatbázisok közül, mivel ez volt az egyetlen, kutatási célra szabadon felhasználható, olyan megfelelően nagyméretű depressziós beszédadatbázis, ahol a bemondók depressziós állapota is jelölve volt, és a jelöléshez a BDI-skálát használták.

Az adatbázist a többnyelvű vizsgálataim során használtam.

Erre az adatbázisra a későbbiek során, mint „Német Depressziós Beszédadatbázis" hivatkozom.

4.1.3. Olasz Depressziós Beszédadatbázis

Az olasznyelvű depressziós beszédadatbázist a nemzetközi Coala projekt keretében gyűjtötték az olasz kollégák a Magyar Depressziós Beszédadatbázis mintájára. Mivel a depressziós beszédadatbázis kevés mintát tartalmaz, így azt csak a többnyelvű modellek teszteléséhez, illetve a kutatási eredményeim hasznosításaként megvalósított többnyelvű modell tanításához használtam.

Erre az adatbázisra a későbbiek során, mint az „Olasz Depressziós Beszédadatbázis" hivatkozom.

4.1.4. Depressziós beszédadatbázisok összehasonlítása

Mindhárom beszédadatbázisban a bemondóknak (többek között) egy rövid mesét ("Az északi szél és a Nap") kellett felolvasniuk az anyanyelvükön, ami széles körben elterjedt a különböző beszéd vizsgálatokban.

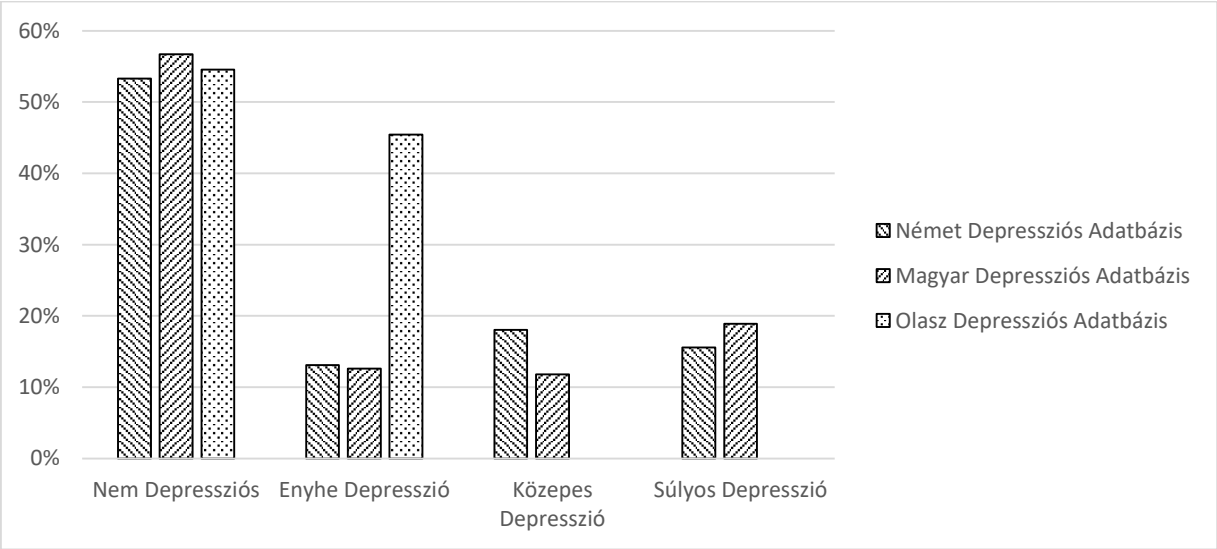
A beszédadatbázisba gyűjtött felvételekhez elkészítettem az egyes felvételekhez tartozó beszédhangszintű szegmentálást, az általam fejlesztett automatikus nyelvfüggetlen szegmentáló program segítségével [82].

Minden bemondó esetében (akár egészséges volt, akár depressziós), rögzítésre került a bemondó depressziós állapota szerinti súlyossága a BDI-skála alapján. A BDI pontszámok 0-tól 50-ig fordultak elő a 3 vizsgált beszédadatbázisban.

A 4. táblázatban láthatók a három beszédadatbázisban található bemondók főbb leíró jellemzői, míg az 1. ábra mutatja külön-külön a három beszédadatbázisban szereplő bemondók depressziós állapotuk szerinti súlyosságának az eloszlását a BDI-skála alapján.

4. táblázat. A három depressziós beszédadatbázis fő jellemzői

Nyelv	Bemondók száma (nők/férfiak)	Bemondók BDI szerinti átlaga és szórása	Bemondók életkorának átlaga és szórása
magyar	127 (79/48)	14.2 (+-13.5)	42.2 (+-14.4)
német	150 (99/51)	14.9 (+-11.7)	31.5 (+-12.3)
olasz	11 (8/3)	12 (+-5.2)	nem állt rendelkezésre adat



1. ábra. BDI értékek eloszlása a három depressziós adatbázisban.

4.2. Statisztikai módszerek

Kutatásom során statisztikai összehasonlító elemzést végeztem az egészséges és depressziós emberek beszédéből kinyert akusztikai-fonetikai jellemzők értékein. A statisztikai elemzés segít a depresszió beszédparaméterekre gyakorolt hatásának megértésében, másrészt a gépi tanuló eljárások esetében, mint egy gyors jellemzővektor kiválasztási algoritmust is fel lehet használni.

Az összehasonlító elemzéseket különböző hipotézisvizsgálatokkal végeztem. A hipotézisvizsgálat lényege, hogy adott egy nullhipotézis (H0) és egy alternatív hipotézis (H1) a H0 tagadása. Így a hipotézisvizsgálat eredményeként vagy megtartjuk a H0-t, vagy pedig elvetjük azt, ami magában foglalja a H1 elfogadását. Ha a hipotézisvizsgálat során elvetjük a H1-t, de valójában a H0 volt igaz, akkor elsőfajú hibát követünk el, míg ellenkező esetben másodfajú hibát vétünk. Az elsőfajú hiba nagysága mindig felülről becsülhető. Adott mintaszám esetén (adott próbát használva), ha csökkenteni akarjuk az egyik hiba valószínűségét, az megnöveli a másik hiba valószínűségét. A két hiba valószínűségének együttes csökkentésére csak a minták számának növelésével van lehetőség. A H0 és a H1 között egy adott próba segítségével és szignifikancia szint mellett lehet döntést hozni a minták alapján. Ahol a szignifikancia szint egyértelműen megadja az elsőfajú hiba felső becslését (100% - szignifikancia szint), vagyis ha a H0-t elutasítjuk, akkor maximum ekkora eséllyel tévedünk [83].

A gyakorlatban a modern statisztikai elemző programok a hipotézisvizsgálat során azt a legkisebb elsőfajú hibát (p) adják vissza, amikor még az adott próba alapján el kell vetni a nullhipotézist. A hipotézisvizsgálataim során én is ezt a gyakorlatot alkalmaztam, és a hipotézisvizsgálatok eredményeinek jobb átláthatósága érdekében a p értékeket az általános gyakorlat szerint négy kategóriába soroltam. A 5. táblázatban látható, hogy adott p érték esetén milyen jelölést alkalmaztam.

5. táblázat. p értékének 4 kategória szerinti jelölése

p értéke	Kategória jelölése	Kategória jelentése
>0.05	-	nem szignifikáns
0.05-0.01	*	szignifikáns az elsőfajú hiba nagysága maximum 5%
0.01-0.001	**	szignifikáns az elsőfajú hiba nagysága maximum 1%
<0.001	***	szignifikáns az elsőfajú hiba nagysága maximum 0.1%

4.2.1. Kolmogorov-Szmirnov-próba

Kolmogorov-Szmirnov-próba (KS-próba) a nemparaméteres próbák családjába tartozik. A próba használatával többek között (egymintás KS-próba) az vizsgálható, hogy egy adott minta alapján a becsült eloszlás egyezőnek tekinthető-e egy referencia eloszlással.

A KS-próba a minta empirikus eloszlásának és a referencia eloszlásnak a távolsága alapján dönti el, hogy a minta eloszlása egyezőnek tekinthető-e a referencia eloszlásával (H0), vagy pedig szignifikáns eltérés mutatható ki a két eloszlás között (H1). A H0 elfogadása vagy elutasítása a próba alapján a következő képlettel határozható meg:

$$D_n = \sup_x |F_n(x) - F(x)| \tag{4.1}$$

ahol a D_n, (n a minták száma) a próbastatisztika által számított távolsága az empirikus eloszlásfüggvénynek és a referencia eloszlásnak, $F(x)$ a referencia eloszlás, és $F_n(x)$ az empirikus eloszlásfüggvény, ami a következő képlettel számítható:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x)}(X_i) \tag{4.2}$$

ahol a $I_{(-\infty,x)}(X_i)$ az indikátor függvény, ami 1, ha $X_i \leq x$, egyébként 0. Ha D_n nagyobb, mint az adott mintaszámhoz és adott szignifikancia szinthez tartozó kritikus érték, akkor elveti a H0-t, egyébként megtartja [83].

4.2.2. Levene-próba

Levene-próba a paraméteres próbák családjába tartozik, a próba alapján az dönthető el, hogy több mintacsoport alapján becsült valószínűségi változók szórása egyező-e (H0), vagy sem (H1). Elsősorban más hipotézisvizsgálatok előtt használják. Kiválasztják a szórások egyezése vagy különbözősége alapján a megfelelő próbát a hipotézis vizsgálathoz. A Levene-próba lényegében felfogható úgy mint az F-próba általánosítása kettőnél több mintacsoport szórás egyezésének vizsgálatához. A H0 elfogadása vagy elutasítása a próba alapján a következő képlettel határozható meg:

$$W = \frac{(N-k)}{(k-1)} * \frac{\sum_{i=1}^{k} N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k}\sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2} \tag{4.3}$$

ahol k a különböző csoportok száma, N_i az i-dik csoportban lévő minták száma, N az összes minta száma, és Z_{ij} a következő képlettel számítható:

$$Z_{ij} = \left|Y_{ij} - \bar{Y}_i\right| \tag{4.4}$$

ahol Y_{ij} az a *i*-dik csoport *j*-dik mintájának értéke, és $\bar{Y}_i$, az átlaga az *i*-dik csoportnak, és a

$$Z_{i.} = \frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ij} \tag{4.5}$$

$$Z_{..} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{j} Z_{ij} \tag{4.6}$$

A W közelítőleg F eloszlást követ *k-1* és *N-k* szabadsági fokokkal. Ha W nagyobb, mint az adott szabadsági fokokhoz és adott szignifikancia szinthez tartozó kritikus érték, akkor elveti a H0-t, egyébként megtartja [84].

4.2.3. Független kétmintás t-próba

A független kétmintás t-próba a paraméteres próbák családjába tartozik, a próba alapján az dönthető el, hogy két mintacsoport alapján becsült valószínűségi változók várható értéke egyező-e (H0), vagy sem (H1). A H0 elfogadása vagy elutasítása a próba alapján a következő képlettekkel határozható meg, annak függvényében, hogy a két becsült valószínűségi változó szórása egyező-e, vagy sem:

-egyező szórás esetén

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(n-1)s_x^{*2} + (m-1)s_y^{*2}}} * \sqrt{\frac{n * m(n+m-2)}{n+m}}, \qquad df = n + m - 2 \tag{4.7}$$

-különböző szórás esetén

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^{*2}}{n-1} + \frac{s_y^{*2}}{m-1}}}, \qquad df = \frac{\frac{s_x^{*2}}{n-1} + \frac{s_y^{*2}}{m-1}}{\left(\frac{s_x^{*2}}{n-1}\right)^2 * \frac{1}{n+1} + \left(\frac{s_y^{*2}}{m-1}\right)^2 * \frac{1}{m+1}} - 2 \tag{4.8}$$

ahol az $\bar{x}$ és $\bar{y}$ a két mintacsoport átlaga, az s_x^* és s_y^* a két mintacsoport korrigált szórása, az *n* az egyik, míg az *m* a másik mintacsoport elemeinek száma, és *df* pedig a próba szabadsági foka. Ha *t* nagyobb, mint az adott szabadsági fokhoz és adott szignifikancia szinthez tartozó kritikus érték, akkor elveti a H0-t, egyébként megtartja.

A független kétmintás t-próba alkalmazásának feltétele, hogy a két becsült valószínűségi változó független és normális eloszlású (illetve jól közelíthetők normális eloszlással) [83][85].

4.2.4. Pearson korreláció

A statisztikában a korreláció két valószínűségi változó lineáris kapcsolatát fejezi ki. Ha két valószínűségi változó független, akkor korrelálatlanok is, a megfordítása általánosságban nem igaz, bár normális eloszlású valószínűségi változók esetében a korrelálatlanság függetlenséget is jelez. Ha a korrelációs együttható 0 (H0), akkor a két változó korrelálatlan, egyébként van köztük lineáris kapcsolat (H1). A H0 elfogadása vagy elutasítása a korrelációs együttható értéke alapján a következő képlettel határozható meg:

$$t = r * \sqrt{\frac{n-2}{1-r^2}}\ , \quad \text{df = n-2} \tag{4.9}$$

ahol a *df* a szabadsági fok, az *n* a két valószínűségi változóból vett mintapáros elemszáma, az *r* pedig a korrelációs együttható, ami a következő képlettel számítható:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y})}{(n-1) * s_x^* * s_y^*} \tag{4.10}$$

ahol az x_i és y_i a i-dik mintapár a mintapárosban, és a $\bar{x}$ és $\bar{y}$ pedig a mintapár egyik illetve másik tagjának az átlagát jelöli. Ha t nagyobb, mint az adott szabadsági fokhoz és adott szignifikancia szinthez tartozó kritikus érték, akkor elveti a H0-t, egyébként megtartja [83].

4.2.5. Statisztikai vizsgálataim

Kutatásaim során (független) kétmintás t-próbát alkalmaztam annak vizsgálatára, hogy a depressziós állapot a beszéd mely paramétereinek várható értékére van hatással, vagyis a depressziós személyek mely akusztikai-fonetikai jellemzőinek átlagértéke mutat szignifikáns eltérést az egészséges személyekhez viszonyítva a vizsgált beszédminták alapján. A t-próba esetén a nullhipotézis azt állítja, hogy a két vizsgált valószínűségi változó várható értéke megegyezik, míg az alternatív hipotézis azt állítja, hogy azok nem egyeznek meg. A t-próba feltétele, hogy a két valószínűségi változó normális eloszlást követ, vagy legalábbis jól közelíthetők normális eloszlással. Emiatt t-próba alkalmazása előtt minden esetben megvizsgáltam, hogy az adott akusztikai jellemzők jól közelíthetők-e normális eloszlással, melyhez ~~Ehhez~~ Kolmogorov-Szmirnov-próbát használtam 99%-os szignifikancia szint mellett. A t-próba használata előtt minden esetben ellenőriztem, hogy a két minta szórása egyezőnek tekinthető-e 95%-os szignifikancia szint mellett Levene-próbát alkalmazva, és ennek a próbának az eredménye alapján vagy az egyenlő szórást, vagy a különböző szórást feltételező kétmintás t-próbát alkalmaztam.

Kutatásaim során korreláció vizsgálatot alkalmaztam még, hogy mely akusztikai-fonetikai jellemzők értékei mutatnak szignifikáns korrelációt a depresszió súlyosságát leíró BDI értékkel. A korreláció vizsgálat alkalmazása esetében Pearson (lineáris) korrelációt vizsgáltam. A korreláció vizsgálat nullhipotézise, hogy a korrelációs együttható értéke 0, az alternatív hipotézis pedig, hogy a korrelációs együttható értéke nem nulla.

A statisztikai vizsgálatokat az SPSS program [86] segítségével végeztem el.

4.3. Osztályozási és regressziós módszerek

Gépi tanuló eljárások használatával osztályozási és regressziós feladatokat valósítottam meg annak érdekében, hogy megvizsgáljam a depressziós állapot detektálása, illetve a depresszió súlyossága mekkora pontossággal állapítható meg automatikusan a beszédből kinyert akusztikai-fonetikai jellemzők segítségével.

Kutatásaim során többféle gépi tanuló eljárást is kipróbáltam az osztályozások és regressziós feladatok megvalósítására, ezek a szupport vektor gépek (support vector machines - SVMs), mesterséges neurális hálók (artifical neural networks – aNN), k-legközelebbi szomszéd (k-nearest neighbours – k-NN) gépi tanuló eljárások voltak [87]. Az eljárások alkalmazását a rapidminer [88] keretrendszer segítségével valósítottam meg.

Legjobb eredményeket szuppport vektor gépek, illetve neurális hálók alkalmazásával értem el, általában valamivel jobb eredményeket kapva szupport vektor gépek használatával. Szupport vektor gépeken belül osztályozás esetében szupport vektor gép (support vector machine – SVM) [89][90] gépi tanuló eljárást használtam, azon belül pedig C-SVM-et, míg regressziós feladatok esetében szupport vektor regresszió (support vector regression – SVR) [91] gépi tanuló eljárást használtam, azon belül pedig epsilon-SVR-t. Neurális hálók esetében 1 illetve 2 rejtett rétegű feed-forward neurális hálókat használtam, (szigmoid rejtett neuronokkal és softmax kimeneti neuronokkal). A rejtett rétegekben 5 és 40 közötti neuronok alkalmazását próbáltam ki, a rejtett rétegekben lévő neuronokat ötösével változtatva.

Mivel a legjobb eredményeket SVMs típusú gépi tanuló eljárások alkalmazásával kaptam, és kutatásom kezdetekor ez volt az egyik legelterjedtebb gépi tanuló család a depressziós állapot detektálására, illetve súlyosságának becslésére, így az ezzel kapott eredményeimet mutatom be a disszertációmban. Annak az oka, hogy nem kaptam jelentős különbséget a neurális hálók és a szupport vektor gépek használata között, feltehetőleg abból fakadhat, hogy a kutatásomban használt depressziós beszédadatbázisok relatív kisméretűek. Azonban az utóbbi években más

kutatási területen elért eredmények alapján egyértelműen megmutatkozik a neurális hálókban rejlő hatalmas potenciál nagyméretű adatbázisok esetében, így ha megjelennek nagyobb méretű depressziós beszédadatbázisok, mindenképpen érdemes újra megvizsgálni a neurális hálók használatát.

SVM gépi tanuló eljárás segítségével két (nem depressziós/depressziós) illetve négyosztályos (nem depressziós/enyhén depressziós/közepesen depressziós/súlyosan depressziós) osztályozási feladatokhoz alkalmas modelleket hoztam létre. A depresszió súlyosságának becsléséhez pedig SVR eljárás segítségével regressziós feladat megvalósításához alkalmas modelleket hoztam létre. A LibSVM [92] által implementált SVM és SVR eljárásokat használtam.

Az SVM és az SVR eljárások kernel függvényének és hiperparamétereinek megválasztását kimerítő kereséssel végeztem. Kernel függvények közül a lineáris, polinomiális és rbf kerneleket teszteltem. Legjobb eredményt minden esetben az rbf kernellel értem el, ahol az rbf kernel két hiperparaméterének (cost és gamma) értékeinek a 2 hatványait teszteltem a kitevő -10 és +10 közötti tartományban.

4.3.1. SVM matematikai hátterének rövid leírása

Az SVM célja, hogy a jellemzővektorral reprezentált minták osztályozására optimális modellt alkosson, vagyis a jellemzővektor által meghatározott térben annak a hipersíknak a meghatározása, ami a tanítómintákat megfelelően szeparálja. Alapvetően kétosztályú osztályozásra alkalmas, ami kibővíthető többosztályú osztályozás alkalmazására is. A kétosztályú osztályozás esetében a két osztályt -1 és +1-gyel jelölik.

Lineárisan szeparálható eset

A tanító adathalmaz mintái lineárisan szeparálhatók, ha a minták által definiált osztályok közé hipersíkot lehet úgy megadni, hogy az hiba nélkül kettéválassza a mintákat az osztályok alapján. Azonban elképzelhető, hogy több ilyen hipersík is megadható, ebben az esetben az SVM a maximális margóval rendelkező hipersíkot adja vissza, ami az általánosító képesség maximalizálását biztosítja. A hipersíkot a $w \cdot x + b = 0$ egyenlettel lehet megadni, ahol w egy a hipersíkra merőleges vektor, $\frac{b}{||w||}$ pedig 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 hipersíkhoz, ezeket szupport vektoroknak nevezik. Az SVM

fő feladata a w és a b értékének meghatározása úgy, hogy a tanító halmazra teljesüljön a következő egyenletrendszer:

$$y\,(x_i * w\ + b) - 1\ \geq 0, \forall i \in I, \tag{4.11}$$

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

Lineárisan nem szeparálható eset

Az SVM képes lineárisan nem szeparálható mintákat is osztályozni, az úgynevezett kernel függvény segítségével úgy, hogy az eredeti problémát magasabb dimenziójú térbe transzformálja, ahol már lehetséges a hipersík megadása. A legelterjedtebb kernel függvények a következők:

$$\text{lineáris: } K\big(\theta(x_i), \theta(x_j)\big) = x_i^T x_j \tag{4.12}$$

$$\text{polinomiális: } K\big(\theta(x_i), \theta(x_j)\big) = \big(\gamma x_i^T x_j + r\big)^d, \gamma > 0 \tag{4.13}$$

$$\text{radiális bázisfüggvény (RBF): } K\big(\theta(x_i), \theta(x_j)\big) = exp\left(-\gamma\left\|x_i - x_j\right\|^2\right), \gamma > 0 \tag{4.14}$$

$$\text{szigmoid: } K\big(\theta(x_i), \theta(x_j)\big) = tanh\big(\gamma x_i^T x_j + r\big) \tag{4.15}$$

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

Azonban a túl magas dimenziójú térbe transzformálás negatív következménye lehet a modell általánosító képességének romlása, emiatt érdemes egy ξ_i gyengítő változó bevezetése, ami megadja minden vektor esetében, hogy a döntés szempontjából hol helyezkedik el. ξ_i=0 azt jelzi, hogy az adott vektor jól lett osztályozva, és hipersík margóján kívül helyezkedik el. ξ_i=1 azt jelzi, hogy az adott vektor rosszul lett osztályozva, és a $0<\xi_i<1$ azt jelzi, hogy bár az adott vektor jól lett osztályozva, de a hipersík margóján belül helyezkedik el. Ennek alapján az eredeti egyenletrendszer a következő képpen módosul:

$$y\,(\theta(x_i) * w\ + b) - 1\ +\ \xi_i \geq 0, \forall i \in I \tag{4.16}$$

továbbra is az $\frac{1}{||w||}$, érték maximalizálása mellett, ami a következő kifejezés minimalizálását jelenti, az előbbi egyenletrendszer teljesülése mellett:

$$min_{w,b,\delta}\left\{\frac{1}{2}||w||^2 + C\sum_{i=1}^{k}\xi_i\right\} \tag{4.17}$$

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.

4.3.2. SVR matematikai hátterének rövid leírása

Az SVR esetében a tanítómintákhoz nem egy osztálycímke van rendelve, 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.

Az SVM hipersíkjához hasonlóan felírható az SVR döntési függvénye:

$$f(x) = w \cdot x + b \tag{4.18}$$

Ahhoz, hogy a függvény változását minimalizáljuk, *w* értékét kell minimalizálni. Ez az SVM-hez hasonlóan, itt is egyenértékű $\frac{1}{2}||w||$ minimalizálásával, miközben teljesül a következő egyenletrendszer:

$$\begin{cases} y_i - w \cdot x_i - b \leq \varepsilon \\ w \cdot x_i + b - y_i \leq \varepsilon \end{cases} . \forall\, i \in I \tag{4.19}$$

Az SVM működéséhez hasonlóan itt is lehetséges, hogy nem oldható meg az egyenletrendszer, vagyis hogy a maximum ε hibával rendelkező *f(x)* függvény a tanítóminták eredeti terében nem létezik. Ekkor az SVM esetében bemutatott módszert itt is alkalmazható, tehát engedélyezni kell a hibákat is, a gyengítő változók hozzáadásával (ξ_i és ξ_i^*) és kernel függvények segítségével transzformálni kell az eredeti problémateret. Ennek alapján az eredeti egyenletrendszer a következőképpen módosul:

$$\begin{cases} y_i - w \cdot \theta(x_i) - b \leq \varepsilon + \xi_i \\ w \cdot \theta(x_i) + b - y_i \leq \varepsilon + \xi_i^* \end{cases} . \forall\, i \in I \tag{4.20}$$

továbbra is az $\frac{1}{2}||w||$ minimalizálása mellett, ami a következő kifejezés minimalizálását jelenti az előbbi egyenletrendszer teljesülése mellett:

$$min_{w,b,\delta} \left\{ \frac{1}{2}||w||^2 + C \sum_{i=1}^{k} (\xi_i + \xi_i^* \right\} \tag{4.21}$$

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

4.3.3. Modellépítés és tesztelési eljárások

A modellépítés során (tanítás) a jellemzővektorban szereplő akusztikai-fonetikai jellemzők értékeit egyenként -1 és 1 közé normáltam. Majd a tesztelés során a jellemzővektorban szereplő értékeket a tanítás során eltárolt normalizáló értékek alapján transzformáltam.

Az adatbázisokban szereplő viszonylag alacsony mintaszám miatt az ilyenkor szokásos, teljes keresztvalidációs eljárást (LOOCV – leave-one-out cross validation) használtam a tesztelések és a hiperparaméterek optimalizálása során is. Az eljárás lényege, hogy a rendszer pontosságának a leírására mindegyik mintát pontosan egyszer használja fel mint teszthalmaz, míg a maradék mintákat a modell tanításához használja, és így a tesztelő mintákon kapott eredmények írják le a rendszer teljes mintahalmazon számított pontosságát. Természetesen az eljárás során a teszthalmaz és a tanítóhalmaz metszete mindig üres.

A többnyelvű vizsgálataim esetében használtam a szokásos tanító/tesztelő halmaz felosztást is, hogy az általam elért eredményeket összehasonlíthassam más kutatók által közölt eredményekkel.

4.3.4. Kiértékelési módszerek

Osztályozási eljárásaim pontosságának leírására alkalmazott jellemzők

Osztályozási eljárásaim pontosságának jellemzésére minden esetben megadtam a tévesztési mátrixot, ahol az egyes cellákban lévő esetek számát az eredeti kategória szerint százalékban kifejezve adtam meg, vagyis a mátrix $C_{i,j}$-dik cellája a következő képlettel számítható:

$$C_{i,j} = \frac{\sum_{k=1}^{n} I_{i,j}(x_k)}{N_i} * 100 \tag{4.22}$$

ahol a $I_{i,j}(x_k)$ értéke 1 ha a k-dik minta az i-dik osztályból származik, és az osztályozás a j-dik osztályba sorolta míg, az N_i pedig az (eredeti) i-dik osztály mintáinak száma és n az összes minta száma.

Ennek a megoldásnak a fő előnye, hogy így a főátlóba a fedés (recall) értékek kerültek, ami a kétosztályos osztályozás esetén specificitás és szenzitivitás értékeknek felelnek meg, amelyek az orvosi gyakorlatban az egyik leggyakrabban használt leíró jellemzői egy diagnózis pontosságának, és a következő képletekkel adhatók meg:

$$specificitás\ (negatív\ osztály\ fedése) = \frac{TN}{N} * 100 \tag{4.23}$$

$$szenzitivitás\ (pozitív\ osztály\ fedése) = \frac{TP}{P} * 100 \tag{4.24}$$

ahol az *N* a negatív osztály (esetemben a nem depressziós), a *P* a pozitív osztály (esetemben a depressziós) mintáinak a számát jelöli. A *TN* a jól osztályozott negatív minták, míg a *TP* a jól osztályozott pozitív minták számát jelöli.

A tévesztési mátrix mellett megadtam még a pontosságot (accuracy) az osztályozások jellemzésére, ami a következő képlettel számítható:

$$pontosság = \frac{\sum_{k=1}^{n} I(x_k)}{n} * 100 \tag{4.25}$$

ahol az $I(x_k)$ értéke 1, ha a mintát helyesen osztályozta az osztályozó, egyébként 0. Az *n* pedig az összes minta számát jelöli.

Regressziós eljárásaim pontosságának leírására alkalmazott jellemzők

A regressziós eljárásaim pontosságának jellemzésére három leíró jellemzőt használtam, amelyek széles körben elterjedtek a regressziós eljárások pontosságának jellemzésére: az átlagos hibaértékét (MAE – mean absolute error), az átlagos négyzetes hiba értékének a gyökét (RMSE – root mean square error), illetve az eredeti BDI pontszámoknak és az eljárás által becsült BDI pontszámoknak a Pearson korrelációs együttható (r) értékét, ezek a következő képletekkel számíthatók:

$$MAE = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{4.26}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (e_i^2)}{n}} \tag{4.27}$$

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) * (y_i - \bar{y})}{(n-1) * s_x^* * s_y^*} \tag{4.28}$$

ahol az e_i az a *i*-dik minta becslésének a hibáját jelöli, az *n* a minták számát. Az x_i és y_i az *i*-dik minta eredeti és becsült BDI értékét jelöli, míg az $\bar{x}$ és $\bar{y}$ ezek átlagát és a s_x^* és s_y^* ezek korrigált szórását jelöli.

4.3.5. Statisztikai alapú jellemzővektor kiválasztása

Mind az SVM, mind az SVR pontosságát és általánosító képességét befolyásolja a megfelelő bemeneti jellemzővektor kiválasztása [92], vagyis a zajszerű nem releváns jellemzők elhagyása, illetve a tanítóminták számának megfelelő nagyságú jellemzővektor kialakítása. Azonban a lehetséges jellemzővektorok száma az 2^n, ahol n a lehetséges jellemzők száma,

emiatt az összes kipróbálása nagyobb n-ek esetében a gyakorlatban fizikai képtelenség. Sokféle jellemző kiválasztási módszer létezik [93][94], a kutatásaim során én ezek közül az alábbiakat használtam:

Független kétmintás t-próba alapú jellemző kiválasztást, amikor a depressziós állapottal szignifikáns eltérést mutató paramétereket választottam be a jellemzővektorba. *Pearson korreláció alapú* jellemző kiválasztást, amikor a depressziós állapot súlyosságával való szignifikánsan korreláló paramétereket választottam be a jellemzővektorba. Ezen eljárások előnye, hogy könnyen megvalósíthatók, és függetlenek a később használt gépi tanuló eljárástól. A hátránya, hogy nem vizsgálja az egyes paraméterek egymásra gyakorolt hatását, illetve az így kialakított jellemzővektor dimenziója esetlegesen túlságosan nagy a tanítóminták számához viszonyítva.

A harmadik jellemzővektor kiválasztó algoritmusnak *a Fast Forward Selectio*n-t (FFS) használtam, ami $\frac{n(n+1)}{2}$ darab különböző hosszúságú jellemzővektort tesztel. Az eljárásnak a lényege, hogy viszonylag gyorsan kiválaszt egy, az algoritmus által optimálisnak ítélt *n* elemű jellemzővektort, ami az algoritmus által előállított eltérő számú szuboptimális jellemzővektorok közül a legjobban teljesít az osztályozás, illetve a regressziós eljárás esetében. Az algoritmus működését pontokba szedve ismertetem:

- Az eljárás az üres jellemzővektorból indul ki.
- Ezután *n* lépés alatt mindig 1-gyel növeli a jellemzővektor méretét úgy, hogy azt a jellemzőt adja hozzá a jellemzővektorhoz, amivel a legnagyobb pontosság érhető el, és elmenti az adott *n* hosszúságú jellemzővektort és azt, hogy mekkora pontosság érhető el azzal.
- Végül az *n* lépés alatt előállított *n* darab eltárolt jellemzővektor közül azt adja vissza, amelyikkel a legnagyobb pontosság elérése volt lehetséges.

A jellemző kiválasztási módszereket együttesen is használtam például az FFS eljárás meggyorsítása érdekében, a bemenő alapjellemző halmazt a *Független kétmintás t-próbával* vagy a *Pearson korreláció vizsgálattal* előválogattam.

4.4. Kutatás során vizsgált akusztikai-fonetikai jellemzők

Az emberi beszéd rendkívül változatos, pont emiatt az emberi beszédet számos akusztikai-fonetikai jellemzővel le lehet írni, amelyek mind hordoznak valamilyen speciális információt a

realizálódott beszédproduktumról. A beszédjelfeldolgozás-alapú, a bemondó állapotának (például depresszió vagy álmosság) gépi felismerést megcélzó versenyek esetében bevett gyakorlat, hogy több ezer különböző származtatott akusztikai-fonetikai jellemző értéket megadnak az egyes beszédmintákhoz mint kiindulási akusztikai-fonetikai jellemzőhalmaz. Természetesen a megadott akusztikai-fonetikai jellemzők között több hasonló, egymással nagymértékben korreláló jellemző is van [65][95].

Ezzel a gyakorlattal szemben a kutatásom alatt fontosnak tartottam, hogy egy kisméretű, előválogatott származtatott akusztikai-fonetikai jellemzőhalmazzal dolgozzam, mivel kutatási célkitűzéseim között volt a depressziós állapot okozta beszédproduktum érzeti jellemzőinek (monoton, lassabb, stb...) és a beszéd akusztikai-fonetikai jellemzőinek megváltozása közti kapcsolat feltárása. Így a korábbi tapasztalataim alapján olyan akusztikai-fonetikai jellemzőket, továbbiakban ***alapjellemzőket***, választottam ki és vizsgáltam meg a kutatásaim során, amelyek értéke feltehetőleg megváltozik a depressziós állapot hatására, ~~és~~ egymással kevéssé korrelálnak (tehát a beszédproduktum viszonylag jól elkülöníthető tulajdonságait jellemzik), és jelentésük könnyebben összekapcsolható a beszéd érzeti leíró jellemzőivel. A 6. táblázatban látható az általam vizsgált alapjellemzők halmaza, illetve hogy a beszéd mely részein értelmezett a számításuk.

6. táblázat. Kutatásom során felhasznált akusztikai-fonetikai jellemzők

Jellemző neve	Értelmezett számítási tartomány
alapfrekvencia (f0)	a beszéd zöngés szakaszai
jitter	a beszéd zöngés szakaszai
shimmer	a beszéd zöngés szakaszai
első formánsfrekvencia (F1)	a beszéd zöngés szakaszai
első formánsfrekvencia sávszélessége (F2)	a beszéd zöngés szakaszai
első formánsfrekvencia sávszélessége (B1)	a beszéd zöngés szakaszai
második formánsfrekvencia sávszélessége (B2)	a beszéd zöngés szakaszai
intenzitás	teljes beszédminta
mel-sávos energiaértékek	teljes beszédminta
artikulációs sebesség (AS)	teljes beszédminta kivéve a csend szakaszok
relatív szünethossz (RSZH)	csend szakaszok
tranziens arány (RoT)	teljes beszédminta kivéve a csend szakaszok

4.4.1. Felhasznált alapjellemzők leírása

Alapfrekvencia (f0) [Hz]:

Az emberi beszédhangok két csoportba oszthatók: zöngés és zöngétlen hangokra. A zöngés hangok esetében, amikor a hangszalagok nyomásviszonyok változásának a függvényében kvázi periodikusan szétnyílnak, majd összezáródnak, a kialakult amplitúdó-idő függvény periodikusnak, illetve kvázi periodikusnak tekinthető. A zönge frekvenciája adja a beszéd alapfrekvenciáját (alaphangját), az amplitúdója pedig a hangerősséget. Az alapfrekvencia a vizsgált zöngés szakasz periódus idejenek a reciprokát jelenti. Az alapfrekvencia értéke a beszéd során az időben folyamatosan változik, és ez a változás feleltethető meg a beszéd dallamosságának érzetével. Általában a beszélő széles tartományban képes az alapfrekvencia értékét változtatni, azonban az alapfrekvencia átlagos értéke különböző és meghatározó a gyermekek (200-300 Hz), a nők (180 Hz) és a férfiak (100 Hz) esetében [96]. Korábbi kutatások kimutatták, hogy szomorú érzelem hatása például többek között a lecsökkent alapfrekvencia értékében fejeződik ki [97].

Az alapfrekvencia értéke a következő képlettel határozható meg:

$$alapfrekvencia\ [Hz]: \frac{1}{T} \tag{4.29}$$

ahol a T a vizsgált szakasz periódusidejét jelenti másodpercben kifejezve.

Az alapfrekvencia értékének számítását 50 ms-os elemzési ablakkal (60 Hz-es minimum alapfrekvencia kiszámítását teszi lehetővé) autokorrelációs függvény segítségével határoztam meg. Az autokorrelációs eljárás függvénye:

$$r_t(\tau): \frac{\sum_{i=t-\frac{w}{2}}^{t+\frac{w}{2}} s(i) * s(i-\tau)}{\sum_{i=t-\frac{w}{2}}^{t+\frac{w}{2}} s(i)^2} \tag{4.30}$$

ahol *s(t)* a beszédjel, *w* az elemzési ablak mérete.

Jitter [%]:

Az alapfrekvencia hosszabb szakaszon végbemenő változása adja a beszéd dallamosságát, dallammenetét. Ezzel szemben az alapfrekvencia értékének kis elemzési ablakban (néhány periódus) számított változása a beszéd rekedtes, érdes érzetével van összefüggésben. A jitter az alapfrekvencia periódusidejének kis elemzési ablakban vett százalékos ingadozását jelenti a vizsgált elemzési ablakban vett átlagos periódushosszhoz képest [99]. A jitter (jitter ddp) a következő képlettel határozható meg:

$$jitter\ (ddp)\ [\%]: \frac{\frac{\sum_{i=2}^{N} |(T_{i+1} - T_i) - (T_i - T_{i-1})|}{N-2}}{\frac{\sum_{i=1}^{N} T_i}{N}} * 100 \tag{4.31}$$

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

Shimmer [%]:

A shimmer a jitter jellemzőhöz hasonlóan a zöngés szakaszok esetében megjelenő periódusok kis elemzési ablakban értelmezett változását méri, azonban a jitterrel szemben, ami az egymást követő periódushosszok százalékos ingadozását, addig a shimmer az egymást követő periódusok amplitúdójának a százalékos ingadozását méri.

Az alapfrekvencia amplitúdójának a hosszabb szakaszon végbemenő változása adja a beszédhangosság dinamikájának érzetét. Ezzel szemben az alapfrekvencia amplitúdójának kis

elemzési ablakban (néhány periódus) számított változása szintén a beszéd rekedtes, érdes érzetével van összefüggésben. A shimmer az amplitúdó változásnak a vizsgált periódusok átlagos amplitúdójához képest százalékban kifejezett értékét jelenti [99]. A shimmer (shimmer ddp) a következő képlettel határozható meg:

$$shimmer\ (ddp)\ [\%]: \frac{\frac{\sum_{i=2}^{N}|(A_{i+1}-A_i)-(A_i-A_{i-1})|}{N-2}}{\frac{\sum_{i=1}^{N}A_i}{N}} * 100 \qquad (4.32)$$

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

Mind a jitter, mind a shimmer számításánál először meghatároztam a periódusok határait 50 ms-os elemzési ablakkal, majd $N = 5$ értékkel számíttattam a jitter és shimmer értékét.

Első és második formánsfrekvenciák (F1, F2): [Hz]:

Zöngés hangok esetében a hangszalagrezgéssel előállított zönge rezgésformája gerjeszti az artikulációs csatornát. A zönge felbontható alapfrekvenciára és annak felharmonikusaira (alapfrekvencia egész számú többszörösei), egészséges esetben mintegy 5000 Hz-ig. A felharmonikus szerkezet amplitúdó értékeiben általában 12 dB/oktáv csökkenés figyelhető meg. A legnagyobb amplitúdóval az alapfrekvencia rendelkezik. Ez a felharmonikus szerkezet jut az artikulációs csatornába, ahol a felharmonikusok bizonyos csoportjában és környékén erősítés figyelhető meg, mégpedig az artikulációs csatorna által kialakított rezonanciafrekvenciákon és környékükön. A spektrumban mérhető ilyen felerősödött felhangcsoportokat a fonetikai szakirodalomban formánsnak nevezik. Az adott formáns frekvenciáján pedig a rezonanciafrekvencia tartomány burkológörbéjének maximum amplitúdó helyét értjük [100].

Minden zöngés hang esetében jellemző azok tipikus formánsai, illetve formánfrekvencia értékei, és a magánhangzók esetében az első és második formánsfrekvencia már egyértelműen meghatározza az adott magánhangzót [100]. Ha egy adott hang esetében a formánsfrekvenciák értékei jelentősen eltérnek a nyelv által meghatározott adott hangra jellemző formánsfrekvencia értékektől, az artikulációs hibának számít.

Első és második formánsfrekvenciák sávszélessége (B1, B2) [Hz]:

A beszédtechnológia a formánsfrekvencia sávszélessége alatt a jelfeldolgozásban általánosan használt sávszélesség definícióját érti, vagyis az alsó és felső határfrekvencia tartományát, ahol

a határfrekvenciák azt az alsó, illetve felső frekvencia értéket jelölik, ahol az erősítés értéke 3 dB-lel kisebb a formánsfrekvenciánál elért erősítésnél. Az egyes formánsfrekvenciák sávszélességeit a formánsfrekvenciákhoz hasonlóan B1, B2, B3, B4, B5-tel jelöljük. Minél magasabb a formáns sorszáma, általában annál szélesebb a sávszélessége [100]. A normális beszédhez képest megnövekedett sávszélesség értékek azt jelzik, hogy a formánsfrekvenciák erősítése nem megfelelő, ami artikulációs hibának számít.

A formánsfrekvenciákat és a formánsfrekvenciák sávszélességét 30 ms-os elemzési ablakkal a Burg-algoritmus segítségével határoztam meg [101].

Intenzitás [dB]:

A hangintenzitás(I) definíció szerint az egységnyi felületen merőlegesen adott időegység alatt átáramlott hangenergia [W/m^2] mennyiséget jelenti. Az intenzitás alatt ezzel szemben az aktuális hangintenzitás és egy adott referencia hangintenzitás hányadosának 10-es alapú logaritmus értékének tízszeresét értem a disszertációm során, és a következő képlettel számítható:

$$intenzitás\ [dB]: 10 * \log_{10}\left(\frac{I}{I_{ref}}\right) \tag{4.33}$$

ahol az I az aktuális hangintenzitást jelöli, míg az I_{ref} egy adott referencia hangintenzitást jelöl, referenciaértéknek a hallásküszöböt szokás megadni, ami 10^{-12} [W/m^2]. Az intenzitás értéke a hangerősség érzettel áll szoros kapcsolatban, és elsősorban a zöngés szakaszok intenzitás értéke a meghatározó az emberi beszédben [100].

Az intenzitás számításánál 30 ms-os elemzési ablakkal végeztem az amplitúdó értékek négyzetre emelését, azután Gauss ablakoló függvénnyel való konvolúció által kaptam meg az aktuális elemzési ablakra számított intenzitás értéket a következő képlet alapján:

$$E(n) = \sum_{k=0}^{N} x^2\,(k) * G(n-k) \tag{4.34}$$

ahol az N a Gauss-ablak mérete, G az ablokoláshoz használt Gauss-függvény, az $x(k)$ a beszédhang amplitúdója.

Mel-sávos energiaértékek [dB]

Az emberi fül frekvencia felbontóképessége az úgynevezett kritikus sávokkal írható le. A kritikus sávok jól illeszkednek az emberi fülben található csiga frekvenciafelbontó

képességéhez, ami fontos szerepet játszik a hallás folyamatában, ugyanis, ha fülünket egyszerre több hang éri, és ezek egy kritikus sávon belül vannak, akkor ezek intenzitása összegződik, és hallás szempontjából nem számítanak különálló hangnak. A kritikus sávok sávszélessége 500 Hz alatti tartományban körülbelül 100 Hz értékűek, míg 500 Hz fölött a sávszélesség a sávközép-frekvencia növekedésével nő, megközelítőleg a sávközép-frekvencia 1/5.

A kritikus sávok és az érzeti hangmagasságskála (mel-skála) között szoros kapcsolat áll fenn. Az érzeti hangmagasság mértékegysége a mel, amely értékeinek duplázódása a hangmagasságérzetet is duplázza. A 131 Hz megfelel 131 melnek [102]. A frekvencia értékek mel-skálába történő átszámolása a következő képlettel számítható:

$$x\,[mel] = 2595 * \log_{10}\left(1 + \frac{f}{700}\right) \tag{4.35}$$

ahol *x* a mel értéket, *f* pedig a frekvencia értéket jelöli.

A mel-sávos energiaértékeket 30 ms-os elemzési ablakkal számítottam. 27 mel-sávot számoltattam ki 100. meltől kezdve sávonként 100 melt használva. Így a beszédspektrumot körülbelül 60Hz-től 8000 Hz-ig vizsgáltam, ami elégséges a beszédhangok spektrális tulajdonságainak elemzéséhez.

Artikulációs sebesség (AS) [hang/s]:

Az artikulációs sebesség a kiejtett beszédhangok számát jelenti adott időegység alatt. Az artikulációs sebesség nagyban függ a beszéd típusától (olvasott vagy spontán beszéd), beszédtervezés sebességétől, és mértéke viszonylag nagy határok között mozoghat a beszélő szándékának függvényében, ami még normálisnak tekinthető. Az artikulációs sebesség túl nagy értéke hadarás érzetét kelti, míg a túl lassú értéke esetén a folytonos beszéd érzet sérül [103]. A kutatásom során az artikulációs sebesség (AS) alatt az átlagos artikulációs sebességet értem a teljes beszédmintán számítva, az alábbi képlettel:

$$Artikulációs\ sebesség\ [hang/s]: \frac{\sum beszédhangok\ száma}{\sum beszédhangok\ hossza} \tag{4.36}$$

Az artikulációs sebesség meghatározásához szükséges a beszédminta beszédhang szintű szegmentálása vagy a beszédhangok számának ismerete és a beszédmintában található szünetek jelölése.

Relatív szünethossz (RSZH) [%]:

A beszédalapú kommunikáció természetes részét képezik a beszédben található szünetek. A szünetek egyrészt lehetőséget biztosítanak a levegővételre, másrészt a beszédtervezés fontos részét képezik. A megfelelő szünethosszok alkalmazása segíthet a beszéd megértésében, és nyomatékosító hatása is lehet. A szünetek hossza nagyban függ a beszéd típusától (olvasott vagy spontán beszéd). A gyakori és hosszú szünetek a beszédet töredezetté tehetik, míg a ritka és rövid szünetek gyors beszédsebesség érzetét keltik [103]. Kutatásom során a relatív szünethossz (RSZH) alatt a szünetek hosszának és a teljes beszédminta hosszának a százalékos arányát értem a beszédmintákban található kezdő és záró szünetszakaszt leszámítva. Az RSZH értéke az alábbi képlettel számítható:

$$Relatív\ szünethossz\ (RSZH)[\%]: \frac{\sum szünetek\ hossza}{beszédminta\ hossza} * 100 \qquad (4.37)$$

Tranziens arány (ratio of transients – RoT) [%]:

Az emberi beszédproduktum felosztható közel állandó (stacioner) és változó (tranziens) szakaszok egymás utáni sorozatára. A stacioner szakaszok általában a beszédhangok közepén az állandósult szakaszoknál jelennek meg, míg a tranziens szakaszok a beszédhangok átmeneténél, illetve az összetett beszédhangokon belül (például zárfelpattanás) figyelhetők meg. Gyors, hadaró vagy nagy dinamikájú beszéd esetében magas lesz a beszéden belül a tranziensek aránya, míg lassú, monoton beszéd esetében alacsonyabb lesz az értéke. A tranziens arány jellemzőt (ratio of transients) a depressziós állapot monotonitásának jellemzésére én dolgoztam ki. A kutatásom során a RoT jellemző alatt a beszédmintában található tranziens szakaszok hosszának (szüneteket elhagyva) százalékos arányát értem a beszédmintában található beszédszakaszok hosszához képest (szüneteket elhagyva). A RoT értéke az alábbi képlettel számítható (8):

$$Tranziens\ arány(RoT)[\%]: \frac{\sum tranziens\ szakaszok\ hossza}{\sum beszéd\ szakaszok\ hossza} * 100 \qquad (4.38)$$

A beszédminta automatikus felosztását stacioner és tranziens szakaszokra a következő, általam kidolgozott algoritmussal végeztem. Az egymást követő 10 ms hosszú szakaszok automatikusan tranziens vagy stacioner szakaszoknak lettek jelölve. A kategorizálás egy speciális távolságszámítási eljáráson alapul a két egymást követő mérési pont között (9), ahol a TD_i a távolságát jelöli az i-dik és az $i+1$-dik mérési pontnak.

$$TD_i = \frac{W_1 * DI_i + W_2 * DVM_i + W_3 * DMM_i}{W_1 + W_2 + W_3} * 100 \tag{4.39}$$

ahol W_1, W_2, W_3 súly paraméterek a tranziensek detektálásánál és,

$$DI_i = \sqrt{(Intenzitás_i - Intenzitás_{i+1})^2} \tag{4.40}$$

$$DVM_i = \sqrt{(VarMel_i - VarMel_{i+1})^2} \tag{4.41}$$

$$DMM_i = \frac{\sum_{k=1}^{3} \sqrt{\left(MelB_i\left(Max_i(k)\right) - MelB_{i+1}\left(Max_i(k)\right)\right)^2}}{6} + \frac{\sum_{k=1}^{3} \sqrt{\left(MelB_i\left(Max_{i+1}(k)\right) - MelB_{i+1}\left(Max_{i+1}(k)\right)\right)^2}}{6} \tag{4.42}$$

A 10. egyenlet a két egymást követő mérési pont közötti intenzitás változását méri, ahol az $Intenzitás_i$ az i-dik mérési pont intenzitás értéke.

A 11. egyenlet a spektrum szerkezetének változását méri két egymást követő mérési pont között, ahol $VarMel_i$ az i-dik mérési pont mel-sávos energiaértékek varianciája.

A 12. egyenlet a mel-sávos energiaértékek első három maximális értékeinek változását méri két egymást követő mérési pont között, ahol a $Maxi(k)$, a mel-sávos energiaértékeinek k-dik legnagyobb értéke az i-edik mérési pont esetében, és a $MelB_i(j)$ a j-dik mel-sávos energiaértéke az i. mérési ponton.

Egy adott 10 ms-os szakasz esetében, ha a TD_i nagyobb, mint az adott küszöbérték *(TH)*, akkor azt a részt tranziens kategóriába sorolja az algoritmus, egyébként pedig stacioner kategóriába.

A tranziens szakaszok automatikus detektálása 4 szabad paramétert tartalmaz, amelyek befolyásolják a végső RoT értékét. A paraméterek megválasztásánál korábbi eredményeimet használtam fel, ahol úgy határoztam meg az értékeiket, hogy azok minél pontosabban jelöljék a hangátmeneteket. Így az algoritmust a következő értékekkel használtam: W1 = 2, W2 = 4, W3 = 1 és TH = 10.

Az utolsó három alapjellemzőn kívül az összes alapjellemző értékének kiszámítását a praat [98] programmal végeztettem.

4.4.2. A felhasznált akusztikai-fonetikai alapjellemzők mérési helye

A beszéd egyes akusztikai-fonetikai jellemzőinek az értéke nagyban függhet attól, hogy a beszéd mely részén mérjük őket. Emiatt két típusba soroltam a jellemzőket. Az „**I. típus**”-ba azokat a jellemzőket soroltam, amik nagyban függhetnek a mérési helytől (jitter, shimmer, F1,

F2, B1, B2, mel-sávos energiaértékek), és ezeket két különböző módszerrel is megmértem, amelyeket lejjebb részletezek. A „**II. típus”**-ba azokat jellemzőket soroltam, ahol nem érdemes a kétféle helyfüggő mérést elvégezni (f0, intenzitás), illetve azokat a jellemzőket, amik fonémánként nem értelmezettek (AS, RSZH, RoT).

Az **„I. típusú”** jellemzők mérése esetén a következő két módszert használtam:

„a módszer”: beszédhangszintű szegmentálást igénylő módszer, ahol a jellemző értékének a számítását ugyanazon beszédhang közepén elhelyezett megfelelő időablakban mértem, például „E” magánhangzó (SAMPA jelöléssel) közepén. Majd ezekből az értékekből származtattam a statisztikai jellemzőket. Ezekben az esetekben azért az „E” hangot választottam, mivel ez volt a leggyakoribb magánhangzó a bemondott mesében.

„b módszer”: egyszerű előfeldolgozást igénylő módszer, ahol a jellemző értékének a számítását az összes lehetséges helyen kimértem 10 ms-os lépésközzel.

Mindkét számítási módszernek megvan a maga előnye és hátránya. A **„b módszer”** esetén nem szükséges bonyolultabb előfeldolgozása a beszédmintának, elég csupán egy beszéd/nem beszéd, illetve egy zöngés/nem zöngés meghatározás, ami automatikusan is könnyen és elég pontosan elvégezhető, azonban a beszéd természetéből adódóan a mért értékeknek így nagyobb lesz a szórásuk mind az egészséges, mind a depressziós beszélők esetén. Az **„a módszer”** esetén kisebb lehet a mért paraméter értékének a természetes szórása, így jobban tükrözheti a depressziós állapot okozta változást, azonban az automatikus pontos beszédhangszintű szegmentálás megnehezíti a paraméterértékek mérését.

A **„II. típusú”** jellemzők mérése esetén az f0-t a zöngés szakaszokon 10 ms-os lépésközzel, az intenzitást a teljes beszédszakaszon 10 ms-os lépésközzel, az AS-t, a RoT-ot a teljes beszédszakaszon (szüneteket elhagyva), míg az RSZH-t a teljes felvételen (elhagyva a felvétel elején és lévő szüneteket) mértem.

4.4.3. Származtatott akusztikai-fonetikai jellemzők

A mért alapjellemzők értékeiből (AS, RSZH és RoT-ot leszámítva, hiszen ezek már eleve leíró jellemzők) három statisztikai függvény segítségével: átlag, szórás, százalékos tartomány (alsó és felső 1 százalékának elhagyása utána képzett tartomány) három származtatott leíró jellemzőt készítettem: átlag, szórás és százalékos tartomány.

A leíró jellemzőhalmaz kialakítása során törekedtem annak minimalizálásra, így mivel a RoT jellemző számítása során közvetve felhasználja a mel-sávos energiaértékek dinamikáját, a mel-

sávos energiaértékeknél csak az átlagot vizsgáltam. Mivel a beszélő és a mikrofon távolsága, illetve a felvétel rögzítésének erőssége nagyban befolyásolja az átlagos intenzitást, a beszédmintákat amplitúdó szerint csúcsra normáltam és az intenzitás érték esetén csak a szórást és százalékos tartományt számíttattam ki.

Egy beszédmintából összesen 98 származtatott jellemzőt számíttattam ki (7. táblázat).

7. táblázat. Származtatott akusztikai-fonetikai jellemzők száma beszédmintánként

	Jellemző neve	Kétféle számolási módszer	Átlag	Szórás	Százalékos tartomány	Származtatott jellemzők száma
I. típusú jellemzők	jitter	igen	igen	igen	igen	6
	shimmer	igen	igen	igen	igen	6
	F1	igen	igen	igen	igen	6
	F2	igen	igen	igen	igen	6
	B1	igen	igen	igen	igen	6
	B2	igen	igen	igen	igen	6
	mel-sávos energiaértékek[1]	igen	nem	igen	igen	54
II. típusú jellemzők	f0	nem	igen	igen	igen	3
	intenzitás	nem	nem	igen	igen	2
	RSZH	nem	önmagában leíró jellemző			1
	AS	nem	önmagában leíró jellemző			1
	RoT	nem	önmagában leíró jellemző			1
	Összesen:					**98**

[1]: 27 mel-sáv került kiszámításra a beszédmintákból

4.5. Felhasznált programok

BME-TMIT-LSA nyelvfüggetlen, kényszerített beszédhangszegmentáló: beszédhangok és beszédhang határok automatikus jelölésére [82].

Praat: hangelemző és annotáló program, akusztikai-fonetikai alapjellemzők mérésére [98].

LibSVM: SVM/SVR alapú osztályozási és regressziós feladatok megvalósítására [92].

SPSS_20: statisztikai elemző szoftver, statisztikai vizsgálatok megvalósítására [86].

Rapidminer: adatelemző program, aminek segítségével osztályozási és regressziós feladatokat valósítottam meg [88].

5. A depressziós állapot hatásának statisztikai vizsgálata a beszéd akusztikai-fonetikai jellemzőire

Kutatásom kezdetekor egyértelmű volt számomra, hogy a depresszió hatására megváltozhat az emberi beszéd, és ezek a változások mérhetők, számszerűsíthetők. Angol és német nyelvű depressziós beszédadatbázisok segítségével a különböző kutatások kimutatták, hogy a vizsgált nyelvek esetében valóban kimutatható tendenciaszerű elváltozás a depressziós emberek beszédének különböző akusztikai-fonetikai jellemzőiben. Így kutatásom kezdetekor az volt a hipotézisem, hogy a depressziós állapot hatásra az akusztikai-fonetikai jellemzők értékének megváltozása magyar nyelv esetében is kimutatható, és matematikai módszerek segítségével a szignifikáns eltérés igazolható.

Annak ellenére, hogy korábbi eredmények alapján kimutatták angol és német nyelvű beszédminták esetében, hogy a depressziós állapot hatására megváltozik az emberi beszédproduktum, annak az alapos vizsgálata, hogy ez az elváltozás milyen módon függ a depressziós állapot súlyosságától, kevésbé kutatott terület. Emiatt kutatási célkitűzéseim közt szerepelt az akusztikai-fonetikai jellemzők értékeinek a depressziós állapot súlyosságával való korrelációjának vizsgálata is. Így kutatásom kezdetekor azt a hipotézist fogalmaztam meg, hogy az akusztikai-fonetikai jellemzők értékei és a depressziós állapot súlyossága között korreláció mutatható ki, ami matematikai módszerek segítségével igazolható.

Vizsgálataimat a Magyar Depressziós Beszédadatbázis beszédmintái segítségével végeztem el.

5.1. Depressziós és egészséges beszélők akusztikai-fonetikai jellemzőinek összehasonlítása

Megvizsgáltam, hogy mely akusztikai-fonetikai alapjellemzők származtatott értékei mutatnak szignifikáns eltérést a depressziós és nem depressziós emberek beszédéből kinyert akusztikai-fonetikai jellemzők értékeit összehasonlítva. A vizsgálatokat nemek szerint külön és együttesen is elvégeztem. Mivel a beszéd legtöbb alapjellemzőjének értékét jelentősen befolyásolja a beszélő neme, így nemek szerint az egészséges minták alapján z-normalizáltam az akusztikai-fonetikai jellemzők értékeit, ami a következő képlettel számítható:

$$\widehat{x_{i,k}} = \frac{\left(x_{i,k} - \bar{x}_i\right)}{s_i^*} \; \forall i \in I \text{ és } \forall k \in K \tag{5.1}$$

ahol az $\widehat{x_{i,k}}$ az i-dik jellemző k-dik mintájának a z-normalizált értéke, $x_{i,k}$ az eredeti értéke, $\bar{x}_i$ az i-dik jellemző átlaga az egészséges minták alapján, s_i^* az i-dik jellemző korrigált szórása az

egészséges minták alapján, I a jellemzők halmaza, K pedig a minták halmaza. Így az összes akusztikai-fonetikai jellemző értéke az egészséges csoport esetében 0 átlagúra és 1 szórásúra lett transzformálva.

A 4.4. fejezetben bemutatott 98 származtatott akusztikai-fonetikai jellemzőt vizsgáltam független kétmintás t-próbával.

A 8. táblázatban láthatók az **„I. típusú”** jellemzők közül azok a jellemzők, amelyek legalább az egyik nem szerint, legalább az egyik statisztikai függvény szerint szignifikáns eltérést mutattak a két csoport között.

A 9. táblázatban láthatók a **„II. típusú”** jellemzők esetén a szignifikáns eltérést mutató jellemzők, amelyek legalább az egyik nem szerint, legalább az egyik statisztikai függvény szerint szignifikáns eltérést mutattak a két csoport között.

8. táblázat. Kétmintás t-próba eredményei az I. típusú akusztikai-fonetikai jellemzők esetében

Jellemző	Számítási módszer	Statisztikai függvény	Szignifikancia | Változás iránya					
			Együtt		Nők		Férfiak	
jitter	"a"	Átlag	**	↑	**	↑	-	↑
		Szórás	**	↑	**	↑	-	↑
		Sz. tartomány	*	↑	-	↑	*	↑
	"b"	Átlag	**	↑	*	↑	-	↑
		Szórás	*	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
shimmer	"a"	Átlag	***	↑	**	↑	*	↑
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
	"b"	Átlag	***	↑	***	↑	**	↑
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
F1	"a"	Átlag	***	↓	***	↓	***	↓
		Szórás	***	↑	**	↑	*	↑
		Sz. tartomány	-	↑	*	↑	-	↑
	"b"	Átlag	**	↓	-	↓	**	↓
		Szórás	-	↓	-	↓	-	↓
		Sz. tartomány	-	↓	-	↓	-	↓
F2	"a"	Átlag	***	↓	**	↓	**	↓
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
	"b"	Átlag	*	↓	-	↓	-	↓
		Szórás	*	↑	-	↑	-	↑
		Sz. tartomány	-	↑	*	↓	-	↑
B1	"a"	Átlag	***	↑	**	↑	***	↓
		Szórás	***	↑	***	↑	***	↓
		Sz. tartomány	***	↑	**	↑	**	↓
	"b"	Átlag	-	↑	-	↑	-	↓
		Szórás	-	↑	-	↑	-	↓
		Sz. tartomány	-	↑	-	↑	-	↓
B2	"a"	Átlag	***	↑	**	↑	**	↑
		Szórás	*	↑	-	↑	**	↑
		Sz. tartomány	*	↑	-	↑	***	↑
	"b"	Átlag	***	↑	***	↑	-	↑
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
„alacsony" mel-sávos energiaértékek[1]	"a"	Átlag	** (65-492Hz)	↑	** (65-603 Hz)	↑	** (65-390 Hz)	↑
	"b"	Átlag	** (65-390 Hz)	↑	** (65-390 Hz)	↑	** (65-214 Hz)	↑
„magas" mel-sávos energiaértékek[1]	"a"	Átlag	** (1330-6330 Hz)	↓	** (1330-6330 Hz)	↓	** (1330-5187 Hz)	↓
	"b"	Átlag	** (1330-5734 Hz)	↓	** (1330-5734 Hz)	↓	** (1330-4688 Hz)	↓

[1]: a „Szignifikancia" oszlopban zárójelek között feltüntettem, hogy a szignifikáns eltérést ($p<0.01$) mekkora tartományban kaptam

- : nem szignifikáns; * : $p < 0.05$; ** : $p< 0.01$; *** : $p<0.001$

A vizsgálat eredményeként elmondható, hogy több akusztikai-fonetikai jellemző származtatott értéke is szignifikáns eltérést mutatott a két depressziós és egészséges csoportokat összehasonlítva az **„I típusú”** jellemzőhalmaz esetében. A számítás módszereket (**„a- és b módszer”**) összehasonlítva kijelenthető, hogy jellemzően az **„a módszerrel”** számított jellemzők nagyobb szignifikáns eltérést mutattak, mint a **„b módszerrel”** számítottak. Emiatt a későbbi kutatásaimban az **„I. típusú”** jellemzőknél mindig az **„a módszert”** használtam a jellemző értékeinek kinyeréséhez, az akusztikai-fonetikai jellemzőhalmaz redukálása érdekében.

Továbbiakban a vizsgálatból az is megállapítható, hogy szignifikánsan (95%) nagyobb átlag, szórás és százalékos tartomány értékek jellemezték a depressziós csoportot a jitter, shimmer, B1 és B2 esetében, míg kisebb átlag érték jellemezte a depressziós csoportot az F1 és F2 esetében. A mel-sávos energiaértékek esetében az alsó tartományokban (65-390 Hz között) magasabb értékek jellemezték a depressziós csoportot, míg a felső tartományokban (1330-5190 Hz) között ennek ellenkezője volt megfigyelhető, ami felfogható úgy, mint egy megemelkedett spektrális meredekség.

9. táblázat. Két mintás t-próba eredményei az II. típusú akusztikai-fonetikai jellemzők esetében

Jellemző	**Statisztikai függvény**	**Szignifikancia | Változás iránya**					
		Együtt		**Nők**		**Férfiak**	
f0	Átlag	**	↓	**	↓	*	↓
	Szórás	*	↓	*	↓	*	↓
	Sz. tartomány	*	↓	*	↓	*	↓
intenzitás	Szórás	***	↓	***	↓	***	↓
	Sz. tartomány	***	↓	***	↓	**	↓
AS	nincs	**	↓	*	↓	**	↓
RSZH	nincs	*	↑	*	↑	**	↑
RoT	nincs	*	↓	*	↓	*	↓

- : nem szignifikáns; * : $p < 0.05$; ** : $p< 0.01$; *** : $p<0.001$

A **„II. típusú”** jellemzők esetében az összes vizsgált jellemző, mindkét nem esetében szignifikáns eltérést mutatott a két csoport között. A depressziós csoportra a kisebb artikulációs sebesség és tranziens arány, és a magasabb relatív szünetek hossza volt jellemző. Emellett a depressziós csoportot még az alapfrekvencia kisebb átlagos értéke, szórása és kisebb százalékos tartománya, illetve az intenzitás kisebb szórás értéke és százalékos tartománya jellemezte.

Az eredményeimet összehasonlítva a kutatásom kezdetekor ismert tendenciákkal (3.2. fejezet 2. táblázat), és a következő akusztikai-fonetikai jellemzők esetében igazoltam azok szignifikáns változását a magyar nyelv esetében: alapfrekvencia átlaga és dinamikája, F1 átlaga, intenzitás

dinamikája, artikulációs sebesség, szünetek átlagos hossza, spektrális meredekség növekedése. Ezeken kívül a jitter átlaga esetében is szignifikáns növekedést mutattam ki a nemek szerinti együttes elemzésben, azonban nemek szerinti külön elvégzett elemzés során kiderült, ez csak a női beszélők esetében igazolható az adataim alapján. Az F2 átlagában az ismert tendencia ellenkezőjét igazoltam, feltehetőleg a formánsfrekvenciák depresszió hatására való megváltozásának nyelvfüggő következményeként. Azonban az eredményeim alapján felmerül, hogy kiválasztható több olyan akusztikai-fonetikai jellemző, amely depressziós állapot hatására nyelvfüggetlen módon változik.

A 2. táblázatban bemutatott jellemzők közül a hangosságot nem vizsgáltam, mivel véleményem szerint annak mérése problematikus, és precíz megvalósítása a gyakorlatban, például az elsődleges egészségügyi ellátórendszerben, nem kivitelezhető, amit alátámaszt az a tény is, hogy ennek a jellemzőnek a vizsgálatánál kapták a legkülönfélébb eredményeket az egyes kutatók. Közvetlenül a beszédsebességet sem mértem, azonban ennek az értéke az artikulációs sebességből és a szünetek átlagos hosszából megállapítható, így kijelenthető, hogy a depressziós beszélőket a magyar nyelv esetében is a jellemzi a lecsökkent beszédsebesség.

A B1 és B2 vizsgálata nem számít széles körben elterjedtnek a nemzetközi szakirodalomban a depressziós állapot okozta kutatások során, azonban eredményeimből egyértelműen kiderül, hogy azok megnövekedett átlagos értéke is jól jellemzi a depressziós személyeket (magyar nyelv esetében). A RoT értéke esetében a feltételezésemnek megfelelő eredményt kaptam, miszerint depressziós beszélők esetében lecsökken az értéke, így ennek a jellemzőnek használata más kutatók számára is hasznos lehet, ha a RoT jellemző változása más nyelvek esetében is megfigyelhető.

I.1. tézis: [B1, C4, C6, C7, J1] Statisztikai úton bizonyítottam, hogy a depressziós állapot szignifikáns változást okoz a beszéd egyes akusztikai-fonetikai alapjellemzőiből származtatott jellemzők átlagos értékében, az egészséges átlaghoz képest. 99%-os szignifikancia szinten az alábbi alapjellemzők származtatott értékei mutattak szignifikáns eltérést mindkét nem esetén: shimmer, F1, F2, B1, B2, mel-sávos energiaértékek (65-390 Hz és 1330-5190 Hz közöttiek), intenzitás.

5.2. A depresszió és az akusztikai-fonetikai jellemzők közötti korreláció vizsgálata

Korreláció vizsgálatot végeztem a beszédminták BDI-skálán mért értékeivel és az akusztikai-fonetikai alapjellemzők azon származtatott értékei között, amiknél szignifikáns (95%-os szignifikancia szint mellett) különbséget kaptam a depressziós és nem depressziós csoportok között az 5.1. fejezetben bemutatott vizsgálat alapján. Az akusztikai-fonetikai jellemzők számítását az **„a módszerrel”** végeztem az **I. típusú** jellemzők esetében, így összesen 53 származtatott akusztikai-fonetikai jellemzőt vizsgáltam.

A vizsgálatot Pearson korreláció módszerrel nemek szerint együtt és elkülönítve végeztem. Az együttes vizsgálat esetén nemek szerint z-normalizáltam a jellemzőket az 5.1. fejezetben leírtak szerint. Azok az alapjellemzők, amik legalább 95%-os szignifikancia szint mellett, szignifikáns korrelációt mutattak a depresszió súlyosságát leíró BDI értékkel legalább az egyik nem esetén a 10. táblázatban láthatók.

A vizsgált alapjellemzők közül, 99%-os szignifikancia szinten a B2 átlaga, a shimmer átlaga, az F1 átlaga, az intenzitás szórása és százalékos tartománya, és a mel-sávos energiaértékek mutattak szignifikáns korrelációt a BDI értékekkel mindkét nem esetében, bár maga a korreláció abszolút mértéke nem mondható kifejezetten erősnek (0.32 - 0.60).

A kapott eredményeket összehasonlítva az **I.1. tézisben** kapott eredményekkel, azt a következtetést lehet levonni, hogy bár több alapjellemző származtatott értéke is szignifikánsan megváltozik a depresszió hatására, ugyanakkor a depresszió súlyosságával való lineáris kapcsolatuk már nem olyan erős.

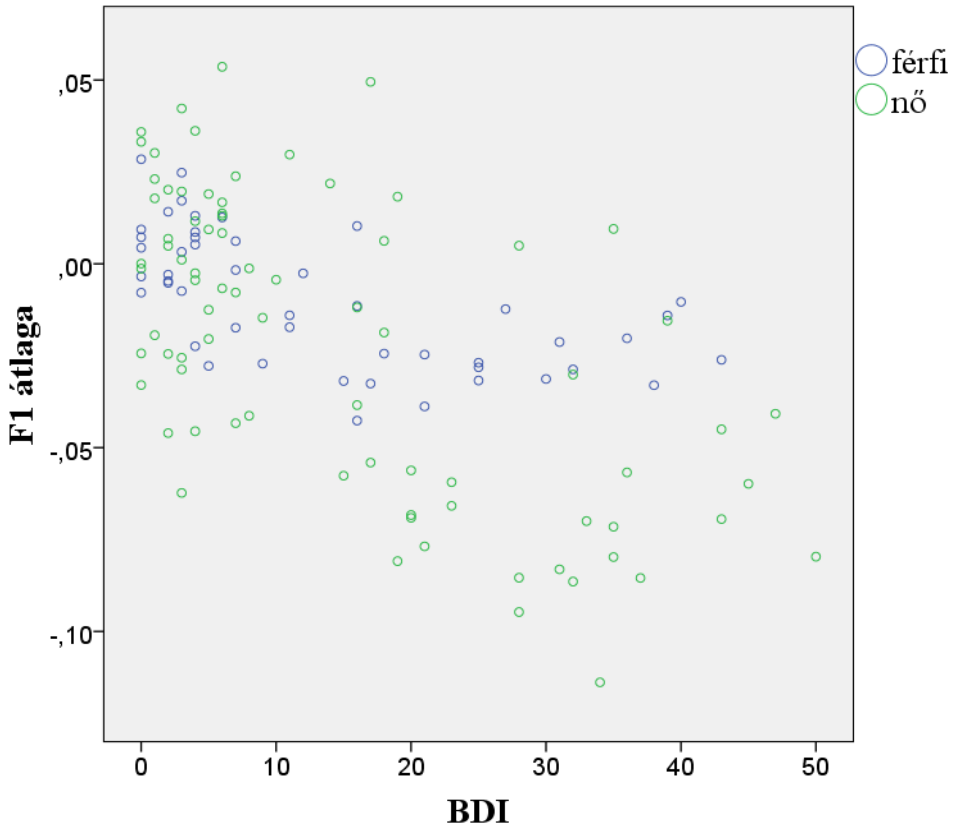
10. táblázat. Akusztikai-fonetikai alapjellemzők és a BDI értékek Pearson féle korrelációja

Jellemző	Korreláció értéke | Szignifikancia szint					
	Együtt		Nők		Férfiak	
jitter átlaga	0.22	*	0.25	*	0.2	*
jitter szórása	0.22	*	0.26	*	0.19	*
shimmer átlaga	0.34	**	0.3	**	0.43	**
shimmer szórása	0.24	**	0.35	**	0.2	*
F1 átlaga	-0.6	**	-0.62	**	-0.62	**
F1 szórása	-0.25	**	-0.29	**	-0.2	-
F2 átlaga	-0.29	**	-0.27	*	-0.36	*
B1 átlaga	-0.31	**	-0.27	*	-0.39	**
B1 szórása	-0.37	**	-0.44	**	-0.36	*
B1 sz. tartománya	-0.33	**	-0.34	**	-0.36	*
B2 átlaga	0.32	**	0.28	**	0.41	**
B2 szórása	0.23	**	0.1	-	0.44	**
B2 sz. tartománya	0.25	**	0.05	-	0.5	**
„alacsony” mel-sávos energiaértékek[1]	0.54 -0.38 (65-390 Hz)	**	0.51 - 0.36 (65-492 Hz)	**	0.58 - 0.41 (65-390 Hz)	**
„magas” mel sávos energiaértékek[1]	-0.47 - -0.37 (1330-5735 Hz)	**	-0.43 - -0.37 (1518-4688 Hz)	**	-0.53 - -0.37 (1518-4688 Hz)	**
f0 átlaga	-0.25	**	-0.32	**	-0.15	-
f0 Sz. tartománya	-0.35	**	-0.55	**	-0.23	*
intenzitás szórása	-0.51	**	-0.54	**	-0.46	**
intenzitás sz. tartománya	-0.53	**	-0.58	**	-0.44	**
AS	-0.29	**	-0.22	*	-0.54	**
RSZH	0.25	**	0.23	*	0.28	*
RoT	-0.28	**	-0.27	*	-0.34	*

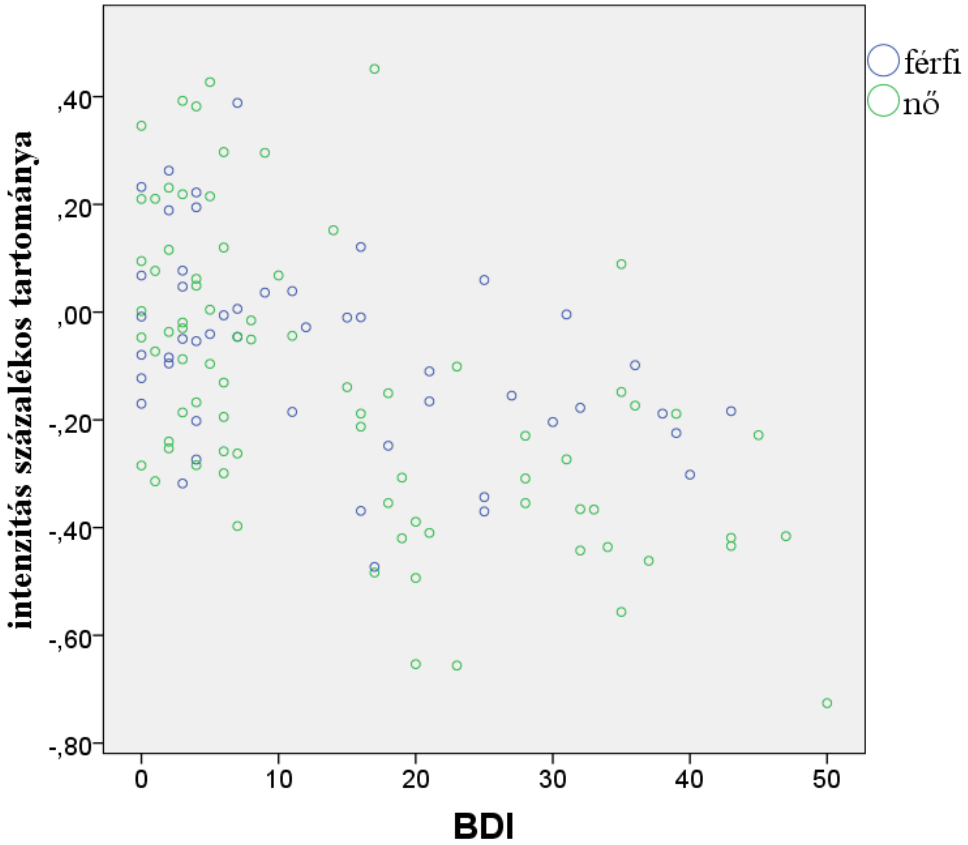
[1]: a „Korreláció értéke” oszlopban zárójelek között feltüntettem, hogy a szignifikáns korrelációt ($p<0.01$) mekkora tartományban kaptam

-: nem szignifikáns; *: $p < 0.05$; **: $p< 0.01$; ***: $p<0.001$

Emiatt további vizsgálatokat végeztem, hogy megállapítsam, hogy a közepesen erős korrelációs értékeknek mi lehet az oka. Ehhez megvizsgáltam az akusztikai-fonetikai jellemzők (z-normalizált) értékeinek, és a BDI értékek szórásgörbéjét (scatter plot), ami alapján arra a következtetésre jutottam, hogy bár a depressziós állapot súlyosabb szintjével valóban nagyobb az akusztikai-fonetikai jellemzők átlagos eltérése az egészséges csoport átlagához viszonyítva, azonban ennek nagy a szórása, ami feltehetőleg a beszéd természetes szórásából adódik. Ezt a dolgozatomban az F1 átlagával (2. ábra) és az intenzitás százalékos tartományával (3. ábra) illusztrálom, azonban a többi akusztikai-fonetikai jellemző esetében is hasonló tendenciákat tapasztaltam.



2. ábra. F1 átlagának és a BDI értékek szórásgörbéje



3. ábra. Intenzitás százalékos tartományának és a BDI értékek szórásgörbéje

I.2. tézis: [J1] Statisztikai úton bizonyítottam, hogy a beszéd egyes akusztikai-fonetikai alapjellemzőinek származtatott értékei szignifikáns korrelációt mutatnak mindkét nem esetében a beszédminta depresszió súlyosságát leíró Beck Depression Inventory (BDI) skálán mért értékével. 99%-os szignifikancia szinten az alábbi alapjellemzők mutattak szignifikáns korrelációt mindkét nem esetén: shimmer, F1, B2, mel-sávos energiaértékek (65-390Hz; 1518-4688Hz), intenzitás.

5.3. Statisztikai vizsgálataim eredményeinek összegzése

Statisztikai vizsgálatok segítségével igazoltam, hogy a depressziós állapot hatására szignifikánsan megváltoznak a beszédnek egyes származtatott akusztikai-fonetikai jellemzői, és ezek a változások több akusztikai-fonetikai jellemző esetében a depresszió súlyosságával szignifikánsan korrelálnak. Azonban az egyes akusztikai-fonetikai jellemzők eltérő mértékben képesek tükrözni a depressziós állapotot mérésük helyétől függően, illetve a beszélő nemétől függően. Továbbiakban az is kiderült a vizsgálataimból, hogy egyik származtatott akusztikai-fonetikai jellemző sem képes önmagában megfelelő pontossággal leírni a depressziós állapot súlyosságát, így egy esetleges gépi osztályozás vagy regresszió megvalósítása esetén több származtatott akusztikai-fonetikai jellemző együttes használatára lesz szükség.

Az eredményeim nemcsak a magyar nyelv esetében számítanak újnak, hanem azok feltehetőleg hasznosíthatók más nyelvek esetében is, és így a nemzetközi szakirodalomban is újszerűek voltak publikálásakor. Ilyenek az akusztikai-fonetikai jellemzők depresszió súlyosságával való szignifikáns korrelációnak alapos vizsgálata, a depressziós állapotnak az akusztikai-fonetikai jellemzőkre gyakorolt hatásának a beszélő neme szerinti függése és egy általam javasolt új akusztikai-fonetikai jellemző (RoT) depressziós állapottal való kapcsolatának kimutatása.

6. A depressziós állapot detektálásának beszédjel feldolgozásán alapuló vizsgálataim

A depressziós állapot diagnosztizálásának támogatása elsősorban az elsődleges egészségügyi ellátórendszerben lenne fontos. Az akusztikai-fonetikai jellemzőkön elvégzett statisztikai vizsgálataim alapján a depressziós állapot gépi detektálása (kétosztályú osztályozása) feltehetőleg megvalósítható, és ezáltal létrehozható a depressziós állapot diagnosztizálásának automatikus támogatása beszédjel feldolgozása alapján. Azonban a létrehozott rendszer pontossága, illetve specificitása és szenzitivitása kritikus.

Angol és német nyelvű beszédminták alapján létrehozott modellek segítségével különböző kutatások 50-80%-os pontossággal voltak képesek a depressziós állapot detektálására (3.2. fejezet), ami alapján kijelenthető, hogy a depressziós állapot gépi osztályozása nem triviális, és annak megfelelő megtervezésével feltehetőleg lényeges javulás érhető el az osztályozás pontosságában. Azonban az 5.2. fejezetben bemutatott eredményeim alapján feltételeztem, hogy nemcsak a depressziós állapot detektálása lehetséges, hanem a depressziós állapot súlyossága szerinti több kategóriájú osztályozása is.

Az osztályozások során a pontosság maximalizálását tűztem ki célul. Mielőtt rátérnék az eredményeim bemutatására, bevezetnék két fogalmat, amelyek mint minimum pontosság értékek elvárhatók a modelljeimmel elért eredményekkel szemben. Az egyik a *„véletlen döntés”*, ebben az esetben az egyes kategóriák között egyenletes eloszlás szerint véletlenül döntünk, ennek az értéke $\frac{1}{N} * 100\%$, ahol N a kategóriák száma. Azonban, ha ismerjük a vizsgált adatbázisban az egyes kategóriák arányait, akkor ennél lehetséges nagyobb pontosságot elérni anélkül, hogy bármilyen leíró jellemzőt használnánk az egyes minták osztályozásához, mégpedig úgy, hogy a kategóriák arányai alapján döntünk véletlenül, ami könnyen belátható, hogy megegyezik azzal az értékkel, mintha minden mintát a legnagyobb mintaszámú kategóriára döntenénk (ha a kategóriák között egyenletes a minták száma, akkor értéke megegyezik a *„véletlen döntéssel”*), ennek az értéke $\frac{k}{m}\%$, ahol k a legnagyobb kategória elemszámát jelenti, az m pedig az összes minta számát, ezt *„eloszlás alapú döntésnek”* hívom a továbbiakban. A Magyar Depressziós Beszédadatbázis esetében a *véletlen döntés* esélye 50% a kétosztályú osztályozás esetében és 25% a négyosztályú osztályozás esetében, míg az *eloszlás alapú döntés* mindkét esetben 57%-os pontosságú (nem depressziós személyek aránya az adatbázisban).

Vizsgálataimat a Magyar Depressziós Beszédadatbázis beszédmintái segítségével végeztem el.

6.1. Kiindulási eredményeim a depressziós állapot gépi detektálásában

Két- (nem depressziós/depressziós), illetve négyosztályos (nem depressziós/enyhén depressziós/közepesen depressziós/ súlyosan depressziós) osztályozási modelleket hoztam létre SVM gépi tanuló eljárást használva.

A jellemzővektor kialakításában az összes származtatott akusztikai-fonetikai alapjellemzőt felhasználtam, és közös modellt alakítottam ki nőkre és férfiakra. Mivel a nők és a férfiak egyes beszédjellemzőinek értékei nagyban különbözhetnek, a jellemzők értékeit nemek szerint normáltam az egészséges beszédminták alapján az 5.1. fejezetben leírtaknak megfelelően.

A modellek tesztelésével előállított tévesztési mátrixok a 11. táblázatban láthatók a kétosztályos modell esetén, és a 12. táblázatban láthatók a négyosztályos osztályozási modell esetén.

11. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, az akusztikai-fonetikai alapjellemzők alapján

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**65%**	35%
	Depressziós	27%	**73%**

68%-os pontosságot értem el (11. táblázat) kétosztályos osztályozás esetén nemek szerinti közös modellel az összes akusztikai-fonetikai alapjellemzőt felhasználva. Az eredményt SVM gépi tanuló eljárással értem el RBF kernellel (cost = 2^5 és gamma = 2^{-5}). A depressziós állapot detektálására 65%-os specificitást és 73%-os szenzitivitást értem el. A modellemmel elért eredmény jobb, mint a *„véletlen döntés”* (50%) és az *„eloszlás alapú döntés”* (57%), az előbbihez viszonyítva 36%-os, míg az utóbbihoz viszonyítva 26%-os relatív csökkenést értem el a hibásan osztályozott személyek számában.

12. táblázat. Négyosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, az akusztikai-fonetikai alapjellemzők alapján

		Osztályozott kategória			
		Nem depressziós	**Enyhén depressziós**	**Közepesen depressziós**	**Súlyosan depressziós**
Eredeti kategória	**Nem depressziós**	**64%**	28%	7%	1%
	Enyhén depressziós	29%	**21%**	21%	29%
	Közepesen depressziós	14%	15%	**43%**	28%
	Súlyosan depressziós	11%	4%	45%	**40%**

52%-os pontosságot értem el (12. táblázat) négyosztályos osztályozás esetén nemek szerinti közös modellel az összes akusztikai-fonetikai alapjellemzőt felhasználva. Az eredményt SVM gépi tanuló eljárással értem el RBF kernellel (cost = 2^5 és gamma = 2^{-5}).

A modellemmel elért pontosság ebben az esetben is jobb, mint a *„véletlen döntés”* (25%), azonban nem éri el, az *„eloszlás alapú döntés”* pontosságát (57%), az előbbihez viszonyítva 36%-os relatív csökkenést értem el a hibásan osztályozott személyek számában, míg az utóbbihoz viszonyítva 12%-os relatív növekedést kaptam a hibásan osztályozott személyek számában.

Megfigyelhető, hogy bár az enyhén depressziós eset kivételével minden kategóriában jobb eredményt értem el, mint a véletlen döntés, azonban az egyes kategóriák elkülönítése relatív nagy hibával valósult meg.

Ezekre az eredményekre (11. táblázat és 12. táblázat) a későbbi osztályozási modelljeim tesztelésekor, mint „kiindulási eredmények” hivatkozom.

II.1. tézis: [C4, J1] Kimutattam, hogy SVM gépi tanuló eljárás alkalmazásával depressziós és egészséges beszélőktől származó beszédminták egymástól elkülöníthetők, továbbá a depresszió négyfokozatú súlyossága szerinti osztályozás is megvalósítható. Kétosztályos osztályozás esetén 68%-os pontosságot, négyosztályos osztályozás esetén 52%-os pontosságot értem el, a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel és nemek szerinti közös modellel.

6.2. Depressziós állapot detektálásának javítása jellemzővektor kiválasztással

A statisztikai vizsgálataim alapján egyértelműen kiderült, hogy bár az általam előválogatott összes akusztikai-fonetikai jellemző hasznos lehet a depressziós állapot detektálásában, azonban egyes származtatott jellemzők esetében nagyobb eltérés mutatható ki, míg más származtatott jellemzők esetében nem tudtam szignifikáns eltérést kimutatni a depressziós állapot hatására, így ezek feltehetőleg, bár nem szükségszerűen, csak zajt visznek a kialakított modellbe. Az is fontos szempont, hogy összesen 127 beszédminta állt rendelkezésemre, és 98 származtatott akusztikai-fonetikai jellemzőt számíttattam ki belőlük. Ezek alapján arra a következtetésre jutottam, hogy feltehetőleg érdemes redukálni a jellemzők számát a modell építése során.

Ennek alapján a 6.1. fejezetben bemutatott modelleket úgy módosítottam, hogy a jellemzővektor kialakítása során, csak azokat az akusztikai-fonetikai jellemzőket használtam fel, amelyek 95%-os szignifikáns eltérést mutattak mindkét nem esetén. Fontos megjegyeznem, hogy a 95%-os szignifikancia szint megválasztása önkényes volt, azonban az eredmények alapján egyértelműen látszik, hogy ez jó kiindulási alap egy esetleges későbbi optimalizáláshoz.

A modellek tesztelésével előállított tévesztési mátrixok a 13. táblázatban láthatók a kétosztályos modell esetén, és az 14. táblázatban láthatók a négyosztályos modell esetén.

13. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**72%**	28%
	Depressziós	29%	**71%**

72%-os pontosságot értem el (13. táblázat) kétosztályos osztályozás esetén nemek szerinti közös modellel, statisztikai jellemző kiválasztási módszerrel. Az eredményt SVM gépi tanuló eljárással értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}). A depressziós állapot detektálására 72%-os specificitást és 71%-os szenzitivitást értem el.

A „kiindulási eredményhez” viszonyítva 12,5%-os relatív csökkenést értem el a hibásan osztályozott személyek számában (11. táblázat és 13. táblázat).

14. táblázat. Négyosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória			
		Nem depressziós	**Enyhén depressziós**	**Közepesen depressziós**	**Súlyosan depressziós**
Eredeti kategória	**Nem depressziós**	**72%**	21%	7%	0%
	Enyhén depressziós	36%	**36%**	21%	7%
	Közepesen depressziós	24%	19%	**36%**	21%
	Súlyosan depressziós	18%	22%	19%	**41%**

57%-os pontosságot értem el (14. táblázat) négyosztályos osztályozás esetén nemek szerinti közös modellel statisztikai jellemző kiválasztási módszerrel. Az eredményt SVM gépi tanuló eljárással értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}).

A „kiindulási eredményhez” viszonyítva 11%-os relatív csökkenést értem el a hibásan osztályozott személyek számában (12. táblázat és 14. táblázat).

Megfigyelhető, hogy az egyes depresszió súlyossága szerinti kategóriák osztályozásának pontossága jobb lett a kiindulás modellemmel elért eredményhez képest, most már minden esetben az adott kategória fedés értéke abszolút értelemben a legnagyobb a többi kategóriára döntéshez viszonyítva, és jobb, mint a *véletlen döntés*. Azonban az adott kategóriába tartozó személyek egy jelentős részét még mindig nagy hibával osztályozza, és a hiba mértéke egy jelentős részben még mindig akár 2 kategória nagyságúnak is adódott.

A kétosztályú és négyosztályú modellek tesztelése alapján kijelenthető, hogy a statisztikai jellemző kiválasztási módszerrel jelentős, akár több mint 10%-os relatív javulás is elérhető a hibásan osztályozott személyek számában.

II.2. tézis: [J1] Kimutattam, hogy statisztikai alapú jellemzővektor kiválasztással akár 12% relatív hibacsökkenés is elérhető a kétosztályú osztályozás esetében, és akár 11% relatív hibacsökkenés is elérhető a négyosztályú osztályozás esetében a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel kapott eredményekhez képest.

6.3. Depressziós állapot detektálásának javítása nemek szerinti eltérő modellek alkalmazásával

Statisztikai vizsgálataim során (5. fejezet) azt figyeltem meg, hogy a legtöbb akusztikai-fonetikai jellemző értéke lényegesen más mértékben tér el a depressziós beszélők esetében a női és férfi beszélőket összehasonlítva. Ezt a megfigyelésemet megerősítette dr. Simon Lajos, aki hasonló következtésre jutott hosszú pszichiátriai gyakorlata során [14]. Ennek alapján azt a hipotézist fogalmaztam meg, hogy a nemek szerinti külön modell kialakítása és alkalmazása javíthatja a depressziós állapot gépi detektálásának pontosságát.

A 6.2.-es fejezetben bemutatott modelleket ennek alapján úgy módosítottam, hogy nemek szerint eltérő jellemzővektorokat alakítottam ki az 5.1. fejezetben bemutatott statisztikai vizsgálat eredménye alapján, vagyis az adott nem szerint szignifikáns (95%) különbséget mutató paramétereket választottam ki, és így nemek szerint eltérő modelleket hoztam létre.

Az osztályozási modellek tesztelésével előállított tévesztési mátrixok a 15. táblázatban láthatók a kétosztályos modell esetén, és az 16. táblázatban láthatók a négyosztályos modell esetén.

15. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti eltérő modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**88%** (nők:84%, férfiak: 93%)	12% (nők: 16%, férfiak: 7%)
	Depressziós	16% (nők: 20%, férfiak: 10%)	**84%** (nők: 80%, férfiak: 90%)

86%-os pontosságot értem el (15. táblázat) kétosztályos osztályozás esetén nemek szerinti eltérő modellel statisztikai jellemző kiválasztási módszerrel. Az eredményt SVM gépi tanuló eljárással értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}). A depressziós állapot detektálására 88%-os specificitást és 84%-os szenzitivitást értem el.

92%-os volt a pontosság a férfiak esetében, 82%-os volt a pontosság a nők esetében. A „kiindulási eredményhez" viszonyítva 56%-os relatív javulást értem el a hibásan osztályozott személyek számában (11. táblázat és 15. táblázat).

16. táblázat. Négyosztályos osztályozás tévesztési mátrixa, nemek szerinti eltérő modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória			
		Nem depressziós	**Enyhén depressziós**	**Közepesen depressziós**	**Súlyosan depressziós**
Eredeti kategória	**Nem depressziós**	**87%** (nők: 89%, férfiak: 85%)	10% (nők:10%, férfiak: 10%)	3% (nők:2%, férfiak: 4%)	0%
	Enyhén depressziós	33% (nők: 30%, férfiak: 40%)	**40%** (nők: 40%, férfiak 40%)	27% (nők: 30%, férfiak 20%)	0%
	Közepesen depressziós	7% (nők11%, férfiak 0%)	20% (nők:11%, férfiak 33%)	**53%** (nők:55%, férfiak 50%)	20% (nők:22%, férfiak: 17%)
	Súlyosan depressziós	8% (nők:6%, férfiak 11%)	12% (nők13%, férfiak 11%)	36% (nők:50%, férfiak: 11%)	**44%** (nők:31%, férfiak: 67%)

69%-os pontosságot értem el (16. táblázat) négyosztályos osztályozás esetén nemek szerinti eltérő modellel, statisztikai jellemző kiválasztási módszerrel. Az eredményt SVM gépi tanuló eljárással értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}).

74%-os volt a pontosság a férfiak esetében, 66%-os volt a pontosság a nők esetében. A „kiindulási eredményhez" viszonyítva 35%-os relatív javulást értem el a hibásan osztályozott személyek számában (12. táblázat és 16. táblázat), és ezzel a modellel már jobb pontosságot

értem el, mint az „eloszlás alapú döntés” (57%), amihez viszonyítva 21%-os relatív javulást értem el a hibásan osztályozott személyek számában

Megfigyelhető, hogy az egyes depresszió súlyossága szerinti kategóriák osztályozásának pontossága lényegesen jobb lett a kiindulás eredményhez képest, most már minden esetben az adott kategória fedés értéke abszolút értelemben a legnagyobb a többi kategóriára döntéshez viszonyítva, és a közepesen depressziós kategória esetében relatív magas 53%-os fedés értéket kaptam, vagyis az ebbe tartozó személyek több, mint felét helyesen kategorizálta. Jelentős javulást értem el a hibásan kategorizált személyek arányában és a hibás döntés esetében a hiba nagyságában is, mivel az esetek többségében a hiba mértéke maximum 1 kategória nagyságú, tehát maximum a szomszédos kategóriára téveszt.

Az eredmények alapján kijelenthető, hogy a nemek szerint eltérő modell kialakításával és használatával jelentős javulást értem el.

II.3. tézis:[J1] Kimutattam, hogy nemek szerint külön létrehozott és statisztikai jellemző kiválasztással optimalizált modellek alkalmazása jelentősen javítja a depresszió osztályozási pontosságát. Az optimalizált modellel 56% relatív csökkenést értem el a hibásan osztályozott személyek számában a kétosztályú osztályozás esetében, és 35% relatív csökkenést értem el a hibásan osztályozott személyek számában a négyosztályú osztályozás esetében, a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel kapott eredményekhez képest.

6.4. Depressziós állapot osztályozási eredményeim összegzése

Osztályozási modellek létrehozásával bemutattam, hogy lehetséges a depressziós állapot detektálását, illetve akár a vizsgált személyek depresszió súlyossága szerinti kategorizálását támogató beszédjel feldolgozás alapú rendszer megalkotása magyar nyelv esetében. A kezdeti modelljeimmel elért eredmények is már hasznosíthatók lehetnének az elsődleges egészségügyi ellátórendszerben, azonban természetesen a megalkotott modell pontossága nagyban befolyásolja a támogató rendszer hasznosságát.

Megmutattam, hogy a megfelelő jellemző kiválasztás fontos a modellek megalkotása során, amire elsősorban a beszédadatbázis mérete miatt van szükség. Azonban a megfelelő depressziós beszédadatbázis kialakítása, a beszédminták gyűjtése időigényes és költséges, így feltehetőleg,

a megfelelő jellemzővektor kiválasztása mindig egy szükséges lépés lesz a minél nagyobb pontosságú diagnózist támogató rendszerek létrehozásában.

Megmutattam, hogy a nemek szerint eltérő modellek kialakításával és használatával jelentős javulás érhető el a depressziós állapot osztályozásában. Így ennek használata mindenképpen kívánatos a minél nagyobb pontosságot elérő diagnózist támogató rendszerek létrehozásában.

A 17. táblázatban a kétosztályú és négyosztályú modellek alkalmazása esetében összefoglalom, hogy a modellek pontossága hogyan változott az általam alkalmazott optimalizáló eljárások használatával, és mekkora relatív csökkenést értem el hibásan osztályozott személyek számában a kiindulási modellel elért eredményemhez viszonyítva.

17. táblázat. Az osztályozás pontossága és a relatív hiba csökkenése az általam alkalmazott optimalizált modellekel

Alkalmazott modellek	**Kétosztályú osztályozás**		**Négyosztályú osztályozás**	
	Pontosság	**Hiba relatív csökkenése**	**Pontosság**	**Hiba relatív csökkenés**
Kiindulási modellek	68%	-	52%	-
Jellemzővektor kiválasztással optimalizált modellek	72%	12%	57%	11%
Nemek szerint eltérő, jellemzővektor kiválasztással optimalizált modellek	86%	56%	69%	35%

A nemek szerinti eltérő modellek alkalmazása a depressziós állapot beszédjel feldolgozás alapú kutatásában nemcsak a magyar, hanem a nemzetközi irodalomban is újszerű volt annak publikálásakor. A legtöbb addigi kutatás nemek szerinti közös modellt alkalmazott, bár Hönig és társai 2014-ben publikált kutatásukban [16] kimutatták, hogy nemek szerint eltérő az akusztikai-fonetikai jellemzők korrelációja a depresszió súlyosságával, azonban a depressziós

állapot gépi osztályozását vagy a depressziós állapot súlyosságának gépi becslésének vizsgálatát nem végezték el. A depressziós állapot detektálásában a nemek szerinti eltérő modellek alkalmazásának jelentőségét bemutató publikációm megjelenésének évében (2017), jelent meg Cummins, Vlasenko, Sagha, és Schuller két kutatási eredménye [104][105], ahol szintén a nemek szerinti eltérő modellek alkalmazásának előnyeit mutatták be.

7. A depressziós állapot súlyosság szerinti becslésének beszédjel feldolgozásán alapuló vizsgálataim

A 6. fejezetben bemutattam, hogy lehetséges a depressziós állapot beszédjel alapú gépi detektálása, illetve akár a depresszió súlyossága szerinti több kategória elkülönítése osztályozási modellek segítségével. Azonban az osztályozási modellek hátránya, hogy az adott probléma esetében nem használja ki, hogy a kategóriák eredetileg folytonosak, és az egy kategórián belül lévő személyeket egységesen kezeli függetlenül attól, hogy azok a kategórián belül az eredeti BDI értéküknek megfelelően hol helyezkednek el. Illetve ebből kifolyólag a vizsgált személyről kapott döntés sem tartalmazza már ezt az információt, holott a diagnózis felállítása során hasznos lehetne, hogy az adott kategórián belül a BDI-skála szerint hol helyezkedik el. Emiatt a továbbiakban megvizsgáltam, hogy mekkora pontossággal lehet megvalósítani a depresszió súlyossága szerinti becslését, regressziós gépi tanuló eljárások megvalósításával, a Magyar Depressziós Beszédadatbázis beszédmintáinak segítségével.

A statisztikai vizsgálataimból kiderült (5. fejezet), hogy a beszédhangszintű szegmentálás felhasználásával számított származtatott akusztikai-fonetikai jellemzők jobban tükrözik a beszélő depressziós állapotát, így az osztályozási modelljeim esetében (6. fejezet) az akusztikai-fonetikai jellemzők értékét a beszédhang szintű szegmentálás felhasználásával állapítottam meg (*„a módszer”*) az **I. típusú** jellemzők esetében. Azonban a beszédhangszintű szegmentálás előállítása, a vizsgált beszédadatbázis beszédhang szintű annotálásának megvalósítása időigényes, és megfelelő automatikus eljárás meglétét feltételezi. Emiatt ebben a fejezetben megvizsgálom azt is, hogy beszédhangszintű szegmentálás felhasználása az akusztikai-fonetikai jellemzők kiszámításánál mennyiben befolyásolja a depressziós állapot súlyosság szerinti becslésének pontosságát.

Szakirodalomban elsősorban a német nyelvű depressziós beszédadatbázis [65] segítségével vizsgálták a depressziós állapot BDI-skála szerinti becslésének lehetőségeit regressziós gépi tanuló eljárások alkalmazásával. A RMSE hibaértékre a különböző kutatások 10,5 és 8,7 közötti értékeket tudtak elérni.

A depressziós állapot súlyosság szerinti becslésénél az RMSE hibaérték minimalizálást tűztem ki célul. A 6. fejezetben bevezetett fogalmakat itt is érdemes megadni. A *„véletlen döntés”* alatt itt azt értem, hogy 0-tól 50-ig egyenletes eloszlással döntünk véletlenszerűen (azért 50-ig, mivel ez volt a legnagyobb BDI értékkel rendelkező személy a beszédadatbázisban), míg az *„eloszlás alapú döntés”* esetében a beszédadatbázisban szereplő személyek átlagos BDI értékére döntünk. Így a *„véletlen döntés”* 17,3 RMSE hibaértéket, míg az *„eloszlás alapú döntés”* 13,5 RMSE hibaértéket jelent a Magyar Depressziós Beszédadatbázis esetében.

7.1. Depressziós állapot súlyosságának becslése beszédhangszintű szegmentálás felhasználása nélkül

SVR gépi tanuló eljárás használatával regressziós modelleket hoztam létre külön női és férfi beszélőkre, ami a beszéd akusztikai-fonetikai jellemzői alapján becsli a beszélő depresszió szerinti súlyosságának mértékét a BDI-skála szerint.

A jellemzővektor kialakításában kizárólag az olyan alapjellemzőket használtam fel, amelyek szignifikáns korrelációt (95%-os szignifikancia szint) mutattak a BDI-skálával, külön a női vagy a férfi beszélők esetében, és az értékük beszédhangszintű szegmentálás nélkül kiszámíthatók, vagyis az **I. típusú** jellemzők értékeit a **„b módszerrel”** számíttattam ki. Az így kialakított jellemzőhalmazból FFS jellemző kiválasztási algoritmussal választottam ki a végleges jellemzővektort.

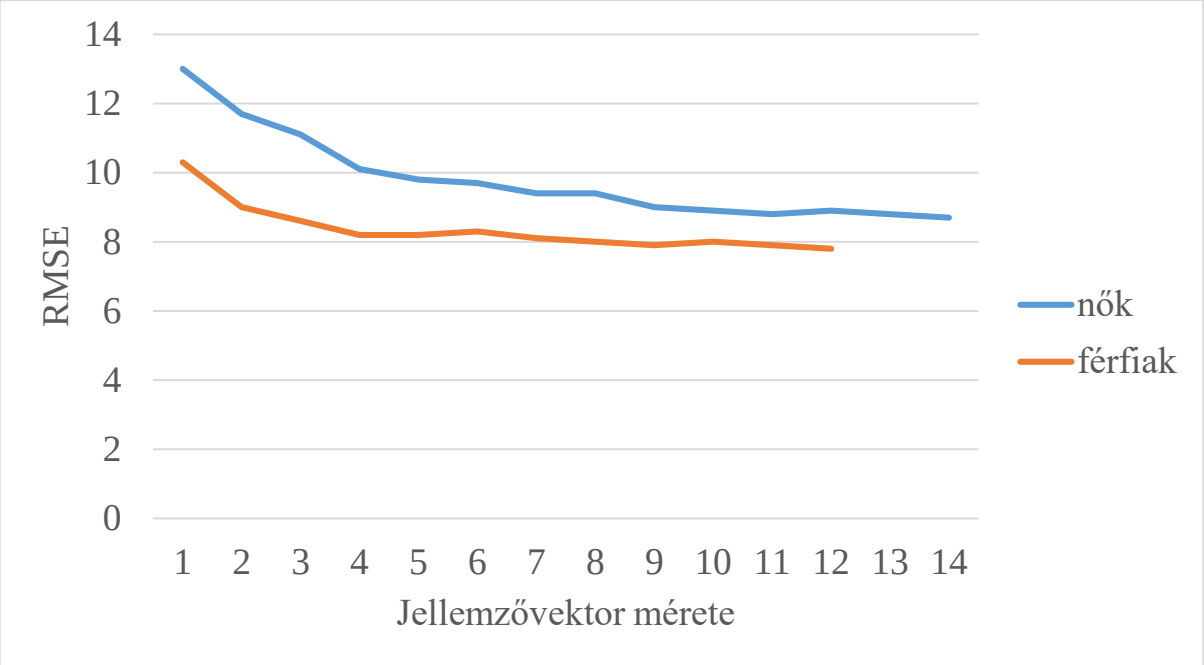
18. táblázat. Az FFS eljárás során a végleges jellemzővektorba válogatott akusztikai-fonetikai jellemzők, az I. típusú jellemzők esetében a „b módszerrel” számítva azokat

	Nők		férfiak	
Jellemzővektor mérete	**Jellemző neve**	**RMSE**	**Jellemző neve**	**RMSE**
1	intenzitás szórása	13	1-es melsáv átlaga	10,3
2	14-es melsáv átlaga	11,7	AS	9
3	RoT	11,1	19-es melsáv átlaga	8,6
4	B2 átlaga	10,1	RoT	8,2
5	intenzitás sz. tartománya	9,8	intenzitás szórása	8,2
6	1-es melsáv átlaga	9,7	F1 átlaga	8,3
7	F2 átlaga	9,4	B1 átlaga	8,1
8	RSZH	9,4	17-es melsáv átlaga	8
9	12-es melsáv átlaga	9	shimmer átlaga	7,9
10	B1 sz. tartománya	8,9	4-es melsáv átlaga	8
11	13-as melsáv átlaga	8,8	B2 szórása	7,9
12	F1 szórása	8,9	14-es melsáv átlaga	7,8
13	shimmer szórása	8,8		
14	f0 százalékos tartománya	8,7		

A 18. táblázatban látható külön a női és a férfi modellek esetében, hogy az FFS eljárás milyen sorrendben, és hány akusztikai-fonetikai jellemzőt választott a végleges jellemzővektorba. A nemenként külön létrehozott végleges jellemzővektor az adott nem szerinti adott oszlop összes jellemzőjét tartalmazza.

A 4. ábrán látható, hogy az FFS eljárás során jellemzővektor méretének növekedésével, hogyan változott az adott nem szerinti modellel elért RMSE hibaérték. A további jellemzők beválogatásával a jellemzővektorba az RMSE hibaérték enyhe növekedése volt megfigyelhető. Megfigyelhető, hogy nemek szerint bár eltérő jellemzővektorokat kaptam, azonban egyes származtatott akusztikai-fonetikai jellemzők, mindkét jellemzővektorban megjelennek

(intenzitás szórása, RoT), amiket viszonylag hamar beválogat az algoritmus (1-5 lépésekben), de a jellemzővektorban megtalálható a legtöbb alapjellemző valamilyen származtatott értéke, ilyenek a formánsfrekvenciák és azok sávszélességei (nők esetében az F1, F2 és B1, B2, míg férfiak esetében csak az F1 és B1), beszédtempóval kapcsolatos jellemzők (nők esetében RSZH, férfiak esetében AS) és a shimmer (nők esetében szórása, férfiak esetében az átlaga). Az FFS eljárást eredményét megvizsgáltam úgy is, hogy az összes alkalmazott akusztikai-fonetikai jellemzőhalmazon futattam, és ugyanazt az eredményt kaptam, mint amikor a statisztikai vizsgálat alapján előválogattam a jellemzőket, így a statisztikai alapú előválogatással esetemben meggyorsítható volt az FFS eljárás alkalmazása.

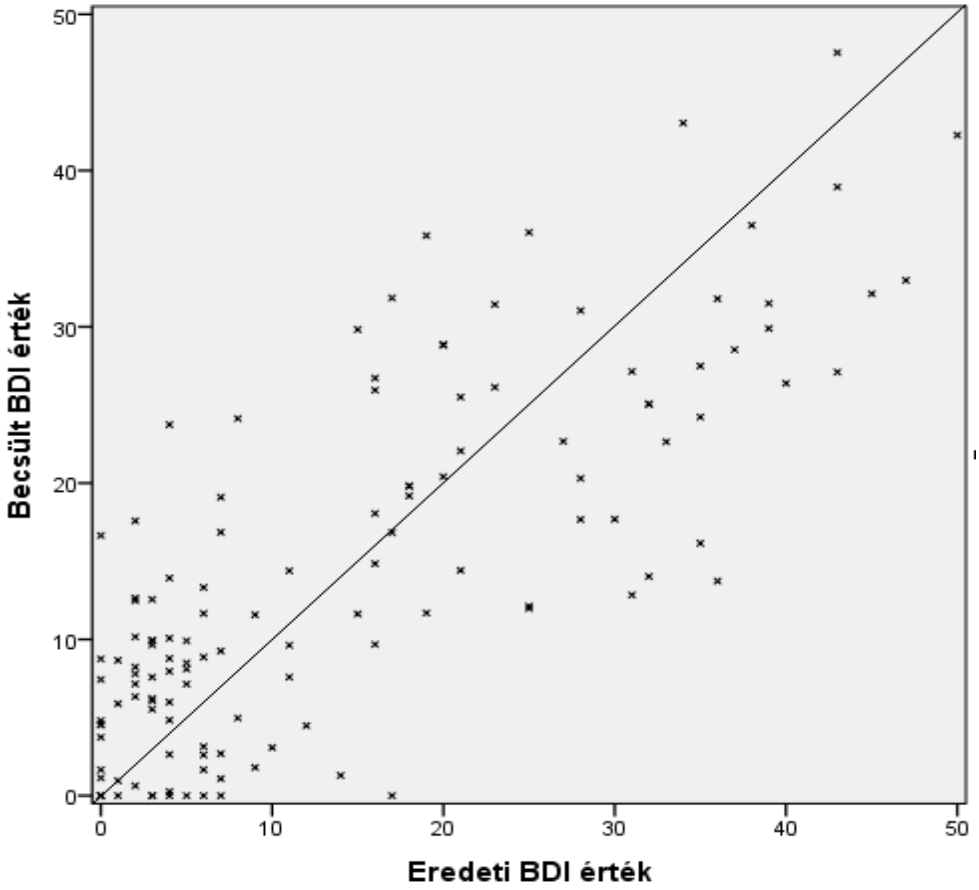


4. ábra. A depressziós állapot súlyosság szerinti becslésének pontossága a jellemzővektor függvényében az FFS eljárás alkalmazása során, az I. típusú jellemzők esetében a „b módszerrel" számítva azokat

A beszédhangszintű szegmentálás nélkül kinyert akusztikai-fonetikai jellemzőkkel épített modell pontosságnak leíró jellemzői a 19. táblázatban láthatók. A 5. ábra mutatja az egyes beszélők eredeti és becsült BDI értékeit. Az eredményeket epsilon SVR típusú gépi tanuló eljárással értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}).

19. táblázat. Depresszió súlyosság becslésének hibája fonéma szintű szegmentálás használata nélkül

	Együtt	Nők	Férfiak
MAE	6,76	7,10	6,20
RMSE	8,37	8,70	7,80
Pearson korreláció	0,79	0,73	0,79



5. ábra. A regressziós modellel becsült, és az eredeti BDI értékek összehasonlítása, abban az esetben, ha a modell építéséhez használt akusztikai-fonetikai jellemzőket a „b módszer" alkalmazásával számítottam ki

A depresszió súlyosságának becslésére kapott eredményem egyértelműen jobb, mint a *„véletlen döntés"* (RMSE 17,3) vagy mint az *„eloszlás alapú döntés"* (RMSE 13,5). Az előbbi esetében 52%-os, míg az utóbbi esetében 38%-os relatív csökkenést értem el az RMSE hibaértékben. Az általam megalkotott modellek alacsony hibával (8,4 RMSE) és magas korrelációval (0,79) képesek becsülni a beszélők depresszió szerinti súlyosságát beszédjel feldolgozás alapján.

Azonban megfigyelhető, hogy egyes estekben a hiba nagysága akár a BDI-skála szerinti több kategória is lehet.

III.1. tézis: [C5] Kimutattam, hogy SVR gépi tanuló eljárás alkalmazásával a beszélő depressziós állapotának súlyossága becsülhető egyszerű (beszédhangszintű szegmentálást nem igénylő) akusztikai-fonetikai jellemzőkkel. A modelleket nemek szerint elkülönítve, statisztikai jellemző kiválasztással optimalizált módon hoztam létre. Az elért átlagos négyzetes hiba gyökére (RMSE) 8,4, míg az átlagos hibára 6,7 (MAE) értékeket kaptam.

7.2. Depressziós állapot súlyosságának becslése beszédhangszintű szegmentálás felhasználásával

A 7.1. fejezetben bemutatott regressziós eljárást úgy módosítottam, hogy az **I. típusú** akusztikai-fonetikai jellemzőket az **„a módszerrel”** számítottam ki. Vagyis a jellemzők kiszámíttatásához felhasználtam az automatikus beszédhangszintű szegmentáló kimenetelét.

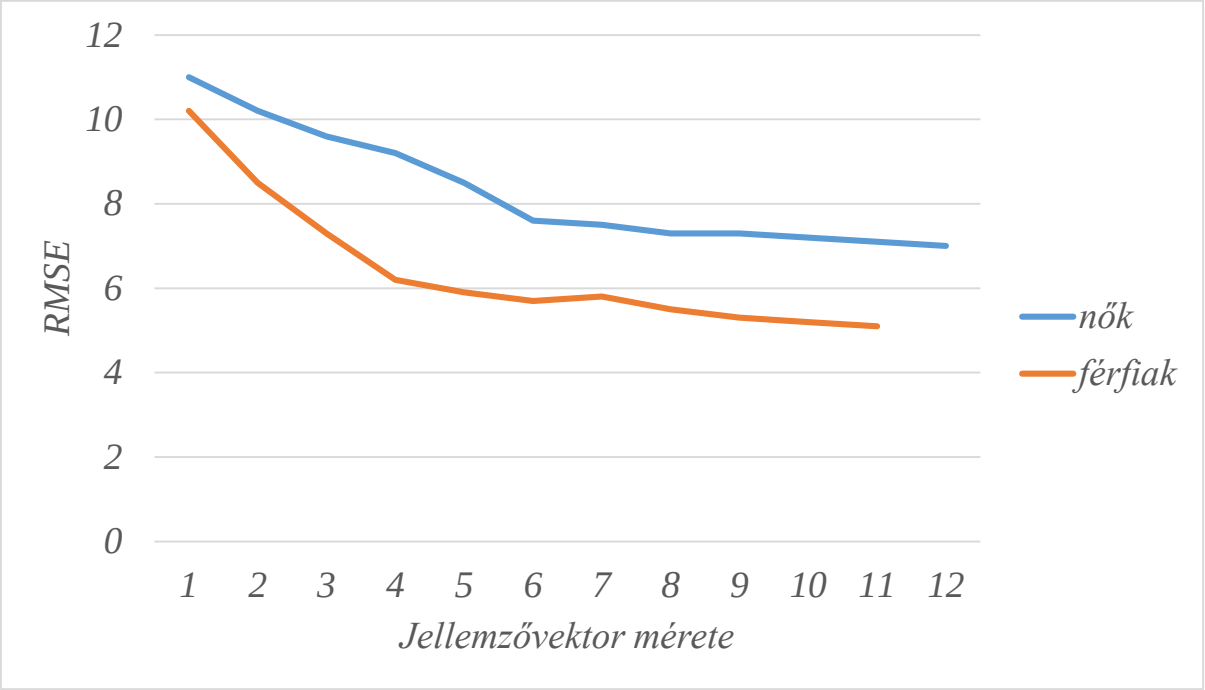
A jellemzővektor kialakításában kizárólag az olyan alapjellemzőket használtam fel, amelyek szignifikáns korrelációt (95%-os szignifikancia szint) mutattak a BDI-skálával, külön a női vagy a férfi beszélők esetében, és az értékük beszédhangszintű szegmentálás felhasználásával számíthatók, vagyis az **I. típusú** jellemzők értékeit a **„a módszerrel”** számíttattam ki. Az így kialakított jellemzőhalmazból FFS jellemző kiválasztási algoritmussal választottam ki a végleges jellemzővektort.

A 20. táblázatban látható külön a női és férfi modellel esetében, hogy az FFS eljárás milyen sorrendben és hány akusztikai-fonetikai jellemzőt választott a végleges jellemzővektorba. A nemenként külön létrehozott végleges jellemzővektor az adott nem szerinti adott oszlop összes jellemzőjét tartalmazza.

20. táblázat. Az FFS eljárás során a végleges jellemzővektorba válogatott akusztikai-fonetikai jellemzők, az I. típusú jellemzők esetében „a módszerrel” számítva azokat

	Nők		**Férfiak**	
Jellemzővektor mérete	**Jellemző neve**	**RMSE**	**Jellemző neve**	**RMSE**
1	F1 átlaga	11	F1 átlaga	10,2
2	F1 szórása	10,2	RoT	8,5
3	RoT	9,6	F2 átlaga	7,3
4	4-es melsáv átlaga	9,2	4-es melsáv	6,2
5	B2 átlaga	8,5	3-as melsáv	5,9
6	22-es melsáv átlaga	7,6	AS	5,7
7	2-es melsáv átlaga	7,5	13-as melsáv	5,8
8	3-as melsáv átlaga	7,3	B1 sz. tartománya	5,5
9	intenzitás szórása	7,3	F1 szórása	5,3
10	F1 sz. tartománya	7,2	20-as melsáv	5,2
11	14-es melsáv átlaga	7,1	B2 szórása	5,1
12	f0 sz. tartománya	7	-	-

A 6. ábrán látható, hogy az FFS eljárás során jellemzővektor méretének növekedésével, hogyan változott az adott nem szerinti modellel elért RMSE hibaérték. A további jellemzők beválogatásával a jellemzővektorba az RMSE hibaérték enyhe növekedése volt megfigyelhető. Ahogy a 7.1. fejezetben a 18. táblázatban, itt is megfigyelhető, hogy nemek szerint eltérő jellemzővektort kaptam, azonban ebben az esetben az 1-4 jellemzők esetében sokkal nagyobb az egyezés (F1 átlaga, RoT, 4-es mel-sáv). Továbbiakban összehasonlítva az eltérő módszerekkel számított akusztikai-fonetikai jellemzőkkel kialakított jellemzővektorokat megfigyelhető, hogy a nemenként kialakított jellemzővektorok hasonló jellemzőket tartalmaznak, azonban ha az **I típusú** jellemzőket az **„a módszerrel”** számítjuk, akkor hamarabb beválogatja őket az FFS algoritmus, és használatuk nagyobb mértékben csökkenti a becslés pontosságának a hibáját, ebből arra lehet következtetni, hogy így pontosabban (kisebb szórással) képesek jellemezni a depressziós állapot szerinti súlyosságot, vagyis hatékonyabban használhatók a regressziós modell tanítására.

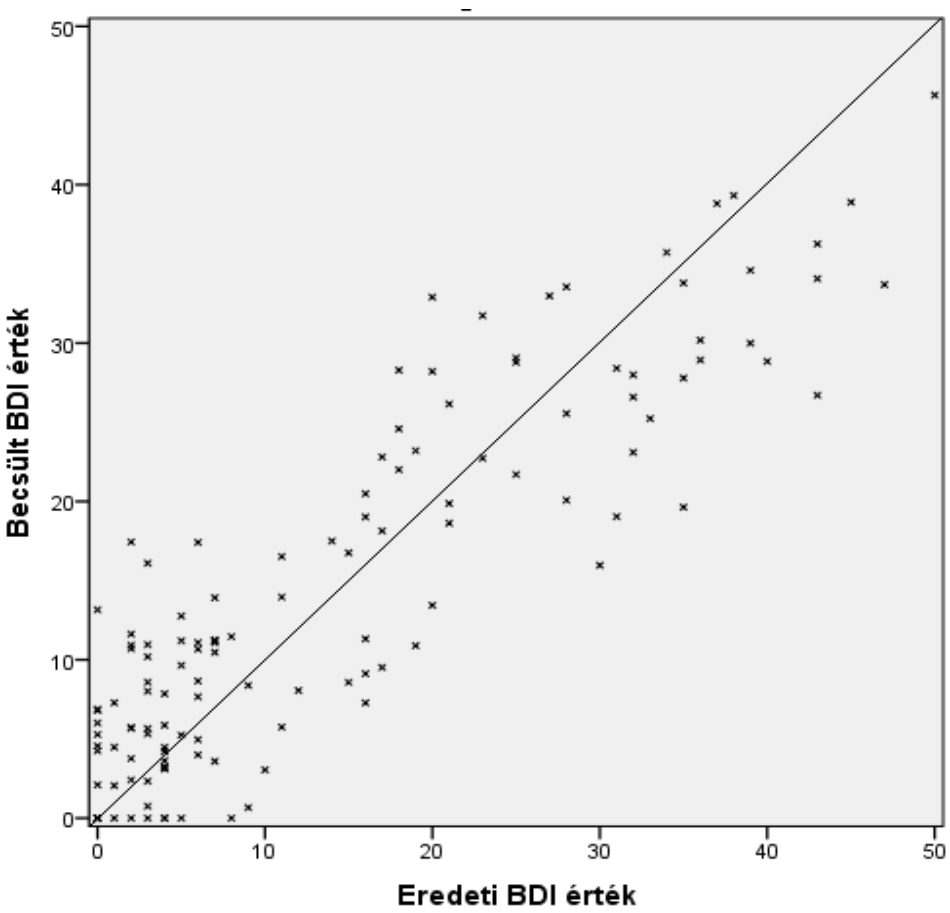


6. ábra. A depressziós állapot súlyosság szerinti becslésének pontossága a jellemzővektor függvényében az FFS eljárás alkalmazása során, az I. típusú jellemzők esetében az „a módszerrel" számítva azokat

A beszédhangszintű szegmentálás feldolgozásával nyert akusztikai-fonetikai jellemzőkkel épített modellek pontosságnak leíró jellemzői a 21. táblázatban láthatók. A 7. ábrán látható az az egyes beszélők eredeti és becsült BDI értékei. Az eredményeket epsilon SVR típusú gépi tanuló eljárással értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}).

21. táblázat. Depresszió súlyosság becslésének hibája beszédhangszintű szegmentálás felhasználásával

	Együtt	**Nők**	**Férfiak**
MAE	5,1	5,7	4,1
RMSE	6,3	7	5
Pearson korreláció	0,89	0,87	0,92



7. ábra. A regressziós modellel becsült, és az eredeti BDI értékek összehasonlítása, ha a modellépítéséhez használt akusztikai-fonetikai jellemzők az „a módszer" alkalmazásával lettek számítva.

A 7.1. fejezetben bemutatott eredményekhez képest a MAE hibaérték esetében 29%-os, míg az RMSE hibaérték estében 25%-os relatív csökkenést értem el, míg a Perason korrelációs együttható esetében 11%-os relatív javulást értem el a pontosabb előfeldolgozó eljárás használatával (beszédhangszintű szegmentálás használatával).

III.2. tézis: [C5] A beszédhangszintű feldolgozással végzett jellemzőkinyerés javít a becslés pontosságán. Az egyszerű előfeldolgozással mérhető (beszédhangszintű szegmentálás nélküli) akusztikai-fonetikai jellemzőkkel kialakított modellel elért eredményhez képest 25%-os relatív csökkenést értem el az RMSE hibaértékben, és 29%-os relatív csökkenést értem el a MAE hibaértékben.

7.3. Depressziós állapot detektálása regressziós eljárás segítségével

A regressziós eljárás segítségével nemcsak a depressziós állapot súlyosság szerinti becslése valósítható meg, hanem a depressziós állapot detektálása, illetve súlyosság szerinti több kategória elkülönítése is. Ennek alapján a 7.2.-es fejezetben bemutatott modellek segítségével elvégeztem a BDI-skála szerinti kétosztályos és négyosztályos tesztelését a regressziós modelljeimnek.

A regressziós modellek tesztelésével előállított tévesztési mátrixok a 22. táblázatban láthatók a kétosztályos tesztelés esetén, és az 23. táblázatban láthatók a négyosztályos tesztelés esetén.

22. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti eltérő modellel, válogatott akusztikai-fonetikai jellemzők alapján, regressziós modellek segítségével

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**97%** (nők:98%, férfiak: 96%)	3% (nők: 2%, férfiak: 4%)
	Depressziós	15% (nők: 17%, férfiak: 10%)	**85%** (nők: 83%, férfiak: 90%)

92%-os pontosságot értem el (22. táblázat) kétosztályos osztályozás esetén, nemek szerinti eltérő modellel, statisztikai jellemző kiválasztási módszerrel regressziós modellek segítségével. Az eredményt SVR gépi tanuló eljárással értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}), a becsült BDI értékek a BDI-skála által definiált kategóriákba sorolásával. A depressziós állapot detektálására 97%-os specificitást és 85%-os szenzitivitást értem el.

94%-os volt a pontosság a férfiak esetében, 91%-os volt a pontosság a nők esetében. A 6.3. fejezetben bemutatott eredményhez viszonyítva 43%-os relatív javulást értem el a hibásan osztályozott személyek számában (15. táblázat és 22. táblázat).

23. táblázat. Négyosztályos osztályozás tévesztési mátrixa, nemek szerinti eltérő modellel, válogatott akusztikai-fonetikai jellemzők alapján, regressziós modellek segítségével

		Osztályozott kategória			
		Nem depressziós	**Enyhén depressziós**	**Közepesen depressziós**	**Súlyosan depressziós**
Eredeti kategória	**Nem depressziós**	**97%** (nők: 98%, férfiak: 96%)	3% (nők:2%, férfiak: 4%)	0%	0%
	Enyhén depressziós	37% (nők: 50%, férfiak: 16%)	**25%** (nők: 30%, férfiak: 17%)	38% (nők: 10%, férfiak 67%)	0%
	Közepesen depressziós	7% (nők11%, férfiak 0%)	13% (nők:11%, férfiak 17%)	**40%** (nők:22%, férfiak 67%)	40% (nők:56%, férfiak: 16%)
	Súlyosan depressziós	4% (nők:0%, férfiak 12%)	4% (nők 6%, férfiak 0%)	17% (nők:25%, férfiak: 0%)	**75%** (nők:69%, férfiak: 88%)

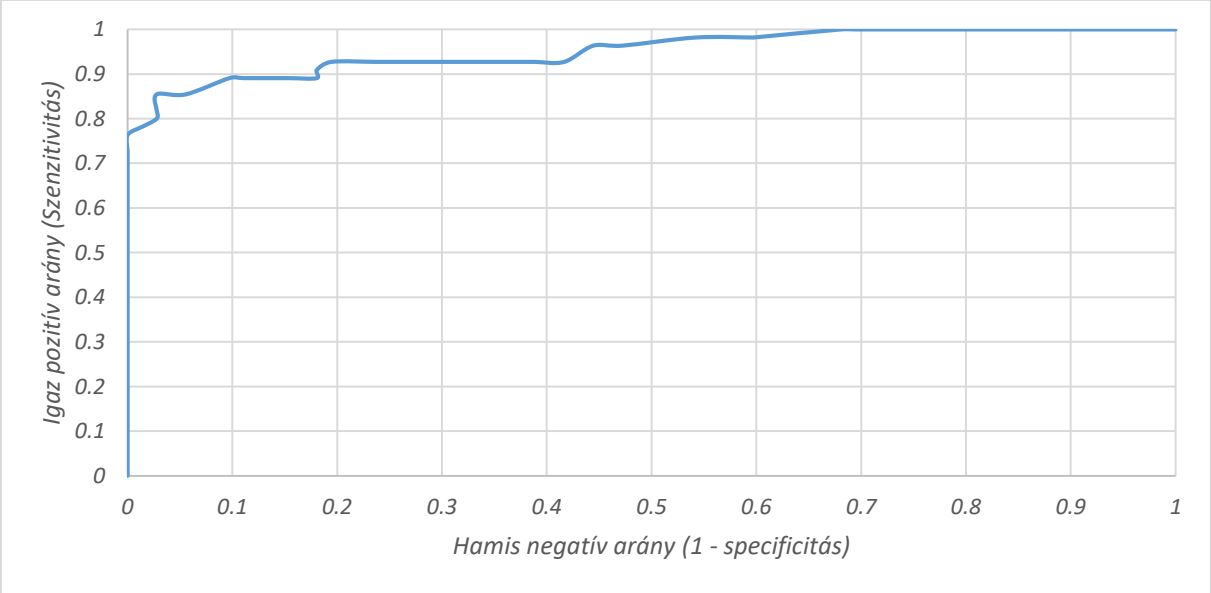
77%-os pontosságot értem el (23. táblázat) négyosztályos osztályozás esetén, nemek szerinti eltérő modellel, statisztikai jellemző kiválasztási módszerrel regressziós modellek segítségével. Az eredményt SVR gépi tanuló eljárásssal értem el RBF kernellel (cost = 2^4 és gamma = 2^{-4}) a becsült BDI értékek a BDI-skála által definiált kategóriákba sorolásával.

81%-os volt a pontosság a férfiak esetében, 75%-os volt a pontosság a nők esetében. A 6.3. fejezetben bemutatott eredményhez viszonyítva 35%-os relatív javulást értem el a hibásan osztályozott személyek számában (16. táblázat és 23. táblázat).

Megfigyelhető, hogy a kétosztályos osztályozás esetében a regressziós modellekkel való detektálása a depressziós állapotnak minden általam vizsgált leíró jellemzőben (pontosság, specificitás, szenzitivitás) jobban teljesített, mint az osztályozási modellekkel elért eredmények. A négy kategória szerinti osztályozás esetében bár jobb pontosságot értem el a regressziós modellek segítségével, azonban az enyhén depressziós és a közepesen depressziós kategóriák fedés értéke alacsonyabb lett. Ennek az oka elsősorban az lehet, hogy ezeknek a kategóriáknak a legkisebb a tartománya a BDI-skála szerint. Így az általam létrehozott

regressziós modellek a gyakorlatban inkább csak három depresszió súlyossága szerinti kategória elkülönítésére alkalmasak igazán (nem depressziós, enyhe+közepes depresszió, súlyos depresszió).

A depressziós állapot detektálása esetében a detektálás pontosságát vizsgáltam elsősorban. Azonban egy gyakorlati orvosi diagnózist támogató rendszer esetében nem feltéttelenül ennek a mérőszámnak az optimalizálása az elsődleges, hiszen nem feltétlenül számít ugyanakkora hibának, ha valakit egészségesként döntünk depressziósra, mint fordítva. Általában a diagnózist támogató eszközöknél a szenzitivitás értékének (vagyis a depressziós személyek felismerése az elsődleges) a maximalizálása az elsődleges szempont, természetesen úgy, hogy a specificitás értéke még értelmes határok között maradjon. A regressziós eljárás osztályozási feladatra adaptált változatának egy további következménye, hogy könnyen módosítható a szenzitivitás értéke (természetesen ezzel a specificitás értéke is) azáltal, hogy a változtatjuk a nem depressziós és depressziós kategóriák BDI szerinti tartományát. Így megvizsgáltam a depressziós állapot detektálása esetében a szenzitivitás és a specificitás összefüggését, amit a ROC (reciver operating characteristic) görbének megadásával szemléltettem (8. ábra). A vizsgálathoz megvizsgáltam, hogy mekkora szenzitivitás és specificitás értékeket kapnék, ha nem depressziós/depressziós kategóriák megállapításához a komperátor BDI értéket 0 és 50 között változtatom 0,5-ös lépésközzel. A ROC-görbe integrálása alapján az AUC (area under curve) értékre 0,96-t kaptam (1 értékű a tökéletes osztályozásé, és 0,5 értékű a véletlen osztályozásé).



8. ábra. A depressziós állapot detektálásának ROC-görbéje a 7.2. fejezetben bemutatott regressziós modellek adaptálásával

A ROC-görbe alapján megfigyelhető, hogy adott igaz pozitív arány esetében (depressziós személy helyes felismerése), mekkora hamis negatív arány (egészséges személy depressziósnak diagnosztizálása) jelenik meg. Ha komparátornak a BDI 13,5-t választjuk, akkor kapjuk a 22. táblázatban bemutatott 85%-os igaz pozitív arányt és a 3%-os negatív hamis arányt. Ha a komparátor értéket 10,5 BDI-re csökkentjük, akkor érhető el 90%-os igaz pozitív arány, de ebben az esetben a negatív hamis arány már 18%-os, és a komparátor érték 7 BDI-re csökkenése esetében érhető el 95%-os igaz pozitív arány, azonban ekkor már 40%-os a negatív hamis arány. Természetesen a relatív kevés mintaszám miatt ezek az értékek inkább tájékoztató jellegűek, azonban legalábbis ekkora tanító beszédadatbázis alapján kijelenthető, hogy bár az általam bemutatott detektálási eljárás önálló diagnosztikai eszközként nem állná meg a helyét, azonban diagnosztikát támogató eszközként nagyon hasznos lehetne elsősorban az elsődleges egészségügyi ellátórendszerben.

7.4. Depressziós állapot súlyosság szerinti becslésére kapott eredményeim összegzése

Regressziós modellek létrehozásával bemutattam, hogy lehetséges a depressziós állapot súlyosság szerinti becslésének megvalósítása. Nemek szerinti eltérő modelleket hoztam létre, ahol nemek szerinti eltérő statisztikai alapú jellemzővektor kiválasztást végeztem. Megvizsgáltam, hogy a becslés mekkora RMSE hibaértékkel valósítható meg, ha a származtatott akusztikai-fonetikai jellemzők kiszámításához rendelkezésre áll beszédhang szintű szegmentálás (6,3 RMSE), és ha nem áll rendelkezésre ilyen (8,37 RMSE). Az eredményeim alapján kijelenthető, hogy a beszédhang szintű szegmentálás felhasználása az akusztikai-fonetikai jellemzők számításában jelentősen csökkenti a jellemzők által létrehozott modellek becslésének hibáját.

Megmutattam, hogy a jellemzővektor növelésével FFS eljárás alkalmazása során, hogyan változik a becslés RMSE hibaértéke, ha az akusztikai-fonetikai jellemzők kiszámításához rendelkezésre áll a beszédhang szintű szegmentálás, vagy ha ilyen nem áll rendelkezésre.

Továbbiakban megvizsgáltam, hogy a regressziós modellek segítségével, mekkora pontosság érhető el a depressziós állapot detektálásában, illetve a depressziós állapot súlyossága szerinti több kategóriájú elkülönítésében. Mindkét esetben nagyobb pontosságot értem el, a 6. fejezetben bemutatót osztályozási modelljeimmel elért eredményeimhez képest, a kétosztályú osztályozás esetében minden általam vizsgált leíró jellemző esetében felülmúlták a regressziós modelljeimmel kapott eredmények az osztályozási modelljeimmel elért eredményeket, azonban

a regressziós modellek 4 osztályos osztályozása esetében a két középső kategória (enyhe depresszió, közepes depresszió) fedésében visszaesést tapasztaltam a 6.3. fejezetben bemutatott osztályozási modelljeimmel elért eredményekhez képest.

8. A depressziós állapotot tükröző akusztikai-fonetikai jellemzők nyelvfüggőségének vizsgálatakor kapott eredményeim

Az emberi beszédproduktumot nagyban meghatározza a beszélő nyelve, így az akusztikai-fonetikai jellemzők értékeit is. Emiatt a beszédtechnológiában egy adott nyelvre megalkotott különböző informatikai megoldások közvetlenül általában nem alkalmazhatók más nyelvek esetében is. Azonban a beszédtechnológia fontos kérdése, hogy az adott problémának mely részei, eljárásai alkalmazhatók nyelvfüggetlen módon. A depressziós állapot gépi felismerésében, a depressziós állapot súlyosságának gépi becslésében is ez a helyzet, vagyis kívánatos lenne, ha egyes megoldások nyelvfüggetlen módon alkalmazhatók lennének, illetve segítségükkel meg lehetne könnyíteni az adott nyelv nyelvspecifikus depressziós állapot gépi detektálásának, depressziós állapot súlyosság szerinti gépi becslésének megvalósítását.

A ma elterjedt gépi tanuló eljárások mind statisztikai alapúak, emiatt a pontos működésükhöz elengedhetetlen a megfelelő számú minta megléte. Azonban a depressziós beszédminták gyűjtése nehézkes és időigényes, emiatt a gyakorlatban nagy mintaszámú depressziós adatbázisok létrehozása nehezen kivitelezhető. Csak az Európai Unióban 24 hivatalos nyelv létezik. Ezek miatt problematikus egy több nyelvre is alkalmazható depressziós állapot diagnosztikáját támogató rendszer megalkotása. A probléma hatványozottan igaz, ha figyelembe vesszük, hogy különböző betegségek beszédjel feldolgozás alapú detektálása [2][5][6][106][107][108][109] aktív kutatási terület, így egy, a gyakorlatban is használható beszédjel feldolgozás alapú, több európai nyelvre alkalmazható, nyelvspecifikus modelleket használó diagnosztikai eszköznek nemcsak minden nyelv esetében kellene megfelelő beszédadatbázist létre hozni, hanem minden betegség típus esetében is.

Ebben a fejezetben azt vizsgálom, hogy lehetséges-e megfelelő közös modell létrehozásával a depressziós állapot súlyosságának becslése többnyelvű (magyar, német és olasz) környezeteben úgy, hogy a beszélő nyelve alapján nem végzek normalizálást. Tehát azt vizsgálom, hogy mekkora pontossággal lehetséges többnyelvű modell alkalmazásával becsülni a depressziós állapot súlyosságát.

Az 5. fejezetben bemutattam, hogy magyar nyelv esetében több származtatott akusztikai-fonetikai jellemző értékében szignifikáns változás mutatható ki a depressziós állapot hatására. A statisztikai vizsgálatom eredményét összehasonlítva más kutatók angol és német nyelvű beszédmintáin kapott eredményeivel, arra a következtetésre jutottam, hogy a legtöbb általam

vizsgált akusztikai-fonetikai jellemző értéke esetében hasonló változás figyelhető meg a depressziós állapot hatására különböző nyelvek esetében (angol, német, magyar). Ennek alapján azt a hipotézist fogalmaztam meg, hogy lehetséges az akusztikai-fonetikai jellemzők egy olyan részhalmazát kiválasztani, ami a beszélő nyelvétől függetlenül tükrözi a depressziós állapot meglétét, és ennek alapján lehetséges a depressziós állapot a beszélő nyelvétől független becslése.

A modelljeim létrehozásához (tanítás) a Magyar és Német Depressziós Beszédadatbázisokat, míg a teszteléshez a Magyar, Német és Olasz Depressziós Beszédadatbázisokat használtam fel.

A 24. táblázatban látható a Magyar Depressziós Beszédadatbázis, a Német Depressziós Beszédadatbázis és a két adatbázis egyesítésén számított *„véletlen döntés”* és *„eloszlás alapú döntés”* RMSE hibaértéke.

24. táblázat. A „véletlen döntés” és az „eloszlás alapú döntés” RMSE hibaértéke a tanítás során használt depressziós beszédadatbázisok esetében

Adatbázis neve	Véletlen döntés RMSE hibaértéke	Eloszlás alapú döntés RMSE hibaértéke
Magyar Depressziós Beszédadatbázis	17,3	13,5
Német Depressziós Beszédadatbázis	15,6	11,7
Magyar + Német Depressziós Beszédadatbázis	16,3	12,5

8.1. A depresszió és az akusztikai-fonetikai jellemzők közötti korreláció nyelvfüggőségének vizsgálata

Korreláció vizsgálatot végeztem a beszélők BDI-skálán mért értékei és a beszédükből kinyert, akusztikai-fonetikai alapjellemzők értékei között. A vizsgálatot Pearson korreláció módszerrel a Magyar Depressziós Adatbázison és a Német Depressziós Adatbázison külön-külön és a két adatbázis egyesítésén is elvégeztem nemek szerint elkülönítve.

Azok az alapjellemzők, amelyek legalább 95%-os szignifikancia szint mellett szignifikáns korrelációt mutattak a depresszió súlyosságát leíró BDI értékkel mindhárom vizsgált adatbázis esetében legalább az egyik nem szerint, a 25. táblázatban láthatók.

25. táblázat. Az akusztikai-fonetikai alapjellemzők és a BDI értékek közötti korreláció elemzés eredményei a magyar és a német beszélők esetében

	Korrelációs együttható mértéke | Szignifikancia szint											
	Német				Magyar				Német + Magyar			
	Nők		Férfiak		Nők		Férfiak		Nők		Férfiak	
B2 átlaga	0.19	*	0.39	*	0.28	**	0.41	**	0.32	**	0.41	**
B2 szórása	0.12	-	0.24	*	0.10	-	0.44	**	0.13	-	0.35	**
jitter átlaga	0.19	*	0.18	*	0.25	*	0.2	*	0.2	*	0.2	*
jitter szórása	0.17	*	0.38	*	0.26	*	0.19	*	0.17	*	0.22	*
shimmer átlaga	0.24	*	0.19	*	0.3	**	0.43	**	0.25	**	0.22	**
shimmer szórása	0.21	*	0.19	*	0.35	**	0.2	*	0.22	**	0.19	*
AS	-0.64	**	-0.23	*	-0.22	*	-0.54	**	-0.28	*	-0.41	**
f0 sz. tartománya	-0.38	*	-0.45	**	-0.55	**	-0.23	*	-0.49	**	-0.32	**
intenzitás szórása	-0.23	*	-0.32	*	-0.54	**	-0.46	**	-0.26	*	-0.27	**
RSZH	0.34	**	0.21	*	0.23	*	0.28	*	0.19	*	0.26	*
RoT	-0.17	*	-0.32	*	-0.27	*	-0.34	*	-0.20	*	-0.21	*

- : nem szignifikáns * : $p < 0.05$; ** : $p< 0.01$; *** : $p<0.001$

A következő akusztikai-fonetikai alapjellemzők mutattak szignifikáns korrelációt (szignifikancia szint 95%) a BDI értékkel mindkét nem esetén (25. táblázat): B2, shimmer, jitter, artikulációs sebesség (AS), alapfrekvencia (f0), intetnzitás, relatív szünethossz (RSZH), RoT.

A vizsgálatom eredményét összehasonlítva a nyelvspecifikus (magyar) vizsgálat eredményével (5.2. fejezet 10. táblázat), azt lehet megfigyelni, hogy csupán néhány származtatott akusztikai-fonetikai jellemző bizonyult mind a két nyelvre érvényesnek a depressziós állapot súlyossága szerinti korrelációval. Elsősorban a spektrális jellemzők (F1, F2, melsávos energiaértékek) hiánya mutatkozik meg, ami persze nem olyan meglepő annak tükrében, hogy ezen jellemzők egészséges, normál értékei nagyban nyelvfüggőek, annak ellenére, hogy számításukat az **„a módszer”** (beszédhangszintű szegmentálás) felhasználásával végeztem.

A vizsgálat alapján kijelenthető, hogy a depressziós állapot hatása hasonló módon jelenik meg a fenti származtatott akusztikai-fonetikai jellemzők értékében (25. táblázat) a vizsgált két nyelv (német és magyar) esetében.

IV. 1. tézis: [B1, C3, J1] Statisztikai úton kimutattam, hogy az alkalmazott akusztikai-fonetikai jellemzőkészletből kiválasztható olyan részhalmaz, amely magyar és német nyelv esetén is szignifikáns korrelációt mutat a beszélők depressziós állapotuk szerinti súlyosságával. Az alábbi jellemzők mutattak 95%-os szignifikancia szinten szignifikáns korrelációt mindkét nem esetén: f0, jitter, shimmer, B2, intenzitás, AS, RSZH, RoT.

8.2. Depressziós állapot súlyosságának többnyelvre érvényes becslése

SVR gépi tanuló eljárás használatával több regressziós modellt is létrehoztam minden esetben külön női és férfi beszélőkre, a depressziós állapot súlyossága szerinti nyelvek közötti becslésének pontosságának vizsgálatára. Vizsgálataim során kétféle tesztelési eljárást alkalmaztam, az egyik esetben az aktuálisan vizsgált beszédadatbázist (beszédadatbázisokat) felosztottam tanító és tesztelő halmazra (normál), a másik esetben a már korábban is használt teljes keresztvalidációs kiértékelést használtam (LOOCV). A megvalósított modellek tesztelésének eredményei a 26. táblázatban láthatók.

Létrehoztam három kiindulási modellt, ahol jellemző kiválasztási eljárás nélkül az összes akusztikai-fonetikai alapjellemzőt felhasználtam a modell építése során (26. táblázat 1-3 sora). A modellek kiértékelését LOOCV eljárással valósítottam meg, és egynyelvű magyar és német, illetve többnyelvű magyar+német modelleket hoztam létre. Ezzel a három modellel elért eredményeket mint kiindulási eredményeket használtam fel a később létrehozott modelljeim eredményeinek kiértékeléséhez.

A kiindulási eredmények esetében a rendszer csak nagy hibával volt képes becsülni a beszélő depressziós súlyosságát, és ahogy várható volt, az egyesített magyar és német nyelvű modell esetében (26. táblázat 3. sor) kaptam a legnagyobb RMSE hibaértékeket, ami nagyobb volt, mint a véletlen döntés RMSE hiba értéke. Az egynyelvű kiindulási modelljeim (20. táblázat 1-2 sorok) RMSE hibaértékei bár jobbak lettek, mint a „véletlen döntés” RMSE hibaértékei, azonban már az „eloszlás alapú döntés” RMSE hibaértékeit csak a magyar nyelvű modellem esetében haladtam meg.

26. táblázat. Regressziós eredmények egy- és többnyelvű modellek esetében különböző tanító és tesztelő adathalmazok esetében

Modell neve	Jellemző kiválasztás	Tanító adathalmaz	Tesztelő adathalmaz	RMSE	MAE
1. Kiindulási_Német	nem	Német Adatbázis	LOOCV	**14.56** nők: 12.6 férfiak: 17.7	**11.58** nők: 10.0 férfiak: 14.6
2. Kiindulási_Magyar	nem	Magyar Adatbázis	LOOCV	**12.43** nők: 12.70 férfiak: 12.00	**9.80** nők: 10.13 férfiak: 9.3
3. Kiindulási_Magyar+Német	nem	Magyar+Német Adatbázis	LOOCV	**19.70** nők: 22.60 férfiak: 13.10	**15.13** nők: 17.9 férfiak: 10.3
4. Német _Normál	igen	Német Adatbázis Tanító Halmaza	Német Adatbázis Tesztelő Halmaza	**8.99** nők: 9 férfiak: 8.9	7.54 nők: 7.9 férfiak: 7
5. Magyar_LOOCV	igen	Magyar Adatbázis	LOOCV	**8.07** nők: 7.8 férfiak: 8.7	**6.31** nők: 6.1 férfiak: 6.8
6. Német_ LOOCV	igen	Német adatbázis	LOOCV	**8.55** nők: 8.9 férfiak: 7.9	**7.02** nők: 7.2 férfiak: 7
7. Magyar+Német_LOOCV	igen	Magyar+Német Adatbázis	LOOCV	**8.59** nők: 8.5 férfiak: 8.7	**7.05** nők: 7.1 férfiak: 7
		Magyar+Német Adatbázis	Magyar Adatbázis	**8.27** nők:7.9 férfiak: 8.9	**6.65** nők: 6.4 férfiak: 7.2
		Magyar+Német Adatbázis	Német Adatbázis	**8.92** nők: 9 férfiak: 8	**7.33** nők: 7.5 férfiak: 7.1
8. Magyar_Német_Normál	igen	Magyar Adatbázis	Német Adatbázis	**9.52** nők: 9.5 férfiak: 9.7	**7.46** nők: 7.6 férfiak: 7.2
9. Német_Magyar_Normál	igen	Német Adatbázis	Magyar Adatbázis	**9.83** nők: 9.8 férfiak: 10	**7.72** nők: 7.4 férfiak: 8.3
10. Magyar+Német_Olasz_Normál	igen	Magyar+Német Adatbázis	Olasz Adatbázis	**5.09** nők: 4.4 férfiak: 6.6	**4.19** nők: 3.4 férfiak: 5.7

A további modellek építése során csak azokat az akusztikai-fonetikai jellemzőket használtam fel, amelyek **IV.1. tézisben** 95%-os szignifikancia szinten a Magyar Depressziós Adatbázis, a Német Depressziós Adatbázis és a két adatbázis egyesítése esetén is szignifikáns korrelációt

mutattak a depresszió súlyosságával, és külön modellt hoztam létre a női és férfi bemondókra. A jellemző kiválasztási algoritmus használata után az eredmények jelentősen javultak.

A Német Depressziós Adatbázist használták az AVEC 2013 versenyben. A versenyben az adatbázist felosztották tanító és tesztelő halmazra, és ez alapján tanítottam be én is a modellemet, és valósítottam meg a tesztelést (26. táblázat 4. sor). Így lehetőségem volt összehasonlítani az akusztikai-fonetikai jellemző számítási és a többnyelvű jellemző kiválasztási módszeremet más kutatók nyelvspecifikus módszerével. A vizsgálat eredményét (RMSE: 8,99; MAE: 7,54) összehasonlítva az AVEC 2013 verseny „baseline" eredményével (RMSE: 14,12; MAE: 10,35), jelentős javulást értem el a depressziós állapot becslésében. Az általam elért eredmények alig maradnak el a verseny győztesének nyelvspecifikus eredményétől (RMSE 8,68; MAE: 7,12) [81].

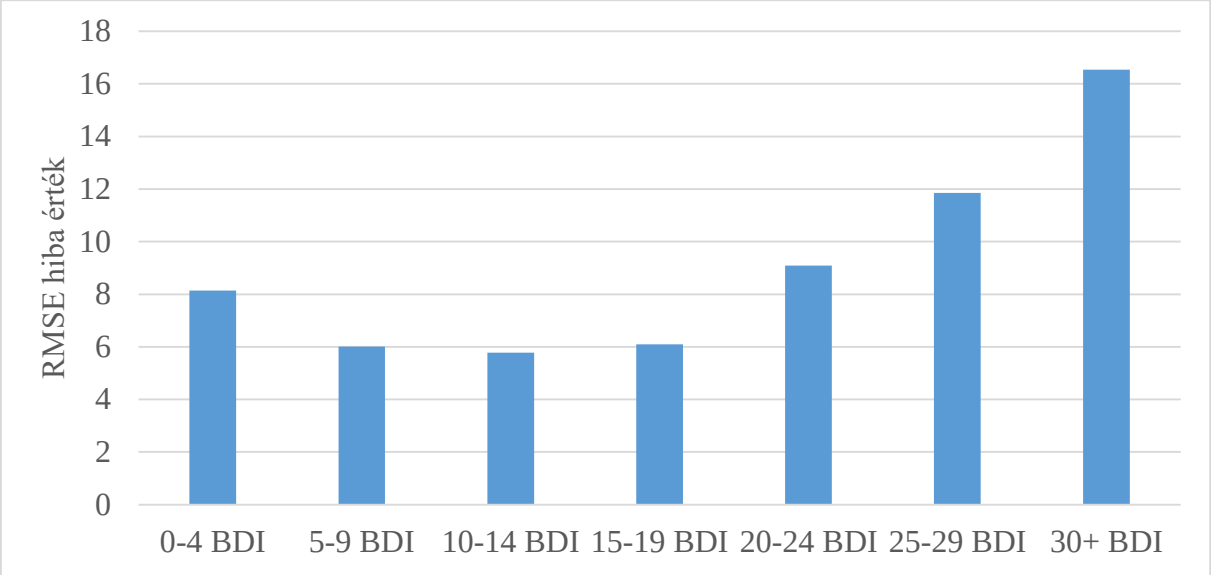
További egy, illetve többnyelvű modelleket hoztam létre, és teszteltem azokat, hogy megvizsgáljam a többnyelvű jellemző kiválasztási módszerem hatását a depressziós állapot becslésére (26. táblázat 5-7 sora). Az eredményeket összehasonlítva a kiindulási eredményekkel, a legjelentősebb javulás a magyar+német többnyelvű modell esetében figyelhető meg, ahol az RMSE értékében 57%-os relatív csökkenést értem el. Tehát kijelenthető, hogy az általam alkalmazott statisztikai jellemző kiválasztási módszerrel jelentős javulás érhető el kétnyelvű modellek esetében.

Megvizsgáltam még azt is, hogy az általam használt módszerekkel milyen pontosságot lehet elérni, ha a tanító és a tesztelő halmazban más anyanyelvű beszélők vannak (26. táblázat 8-10 sora). Az így elért eredmények elmaradnak attól, mint amikor a tanításnál mindkét nyelv mintái szerepelnek (26. táblázat 5-7 sora), de az eredmények összehasonlíthatók, és magyar nyelvű tanítás német nyelven való tesztelési (26. táblázat 8. sora) eredménye még mindig jelentősen jobb, mint az AVEC 2013 „baseline" eredménye.

Amikor az egyesített magyar+német adatbázissal tanítottam a modellt, és az olasz adatbázison teszteltem, kifejezetten jó eredményeket kaptam, azonban fontos megjegyeznem, hogy az olasz adatbázis beszédmintáinak száma alacsony, illetve a BDI eloszlása lényegesen eltérő a másik két adatbázistól (csak nem depressziós és enyhén depressziós személyek beszédmintáit tartalmazza az adatbázis).

Továbbiakban megvizsgáltam az egyesített magyar+német depressziós beszédadatbázissal tanított modell LOOCV teszteléssel (26. táblázat 7. sora) kapott RMSE hibaértékek mértékét a beszélő eredeti BDI értéke alapján csoportosítva (9. ábra). A következő BDIérték szerinti

csoportokat hoztam létre: 0-4 BDI, 5-9 BDI, 10-14 BDI, 15-19 BDI, 20-24 BDI, 25-29 BDI, 30-50 BDI. A csoportok BDI szerinti tartományai egyenletesek, és majdnem fedik a BDI-skála által definiált depresszió súlyossága szerinti kategóriákat (0-13 nem depressziós, 14-19 enyhe depresszió, 20-28 közepes depresszió, 29-63 súlyos depresszió.



9. ábra. RMSE hibaértékek alakulása a beszélők eredeti BDI értéke alapján csoportosítva az egyesített magyar+német depressziós beszédadatbázis mintáival tanított modell esetében LOOCV teszteléssel

A 9. ábrán látható, hogy amíg a beszélők eredeti BDI értéke alacsony, kisebb mint 20, addig a modell által adott becslés pontosságának RMSE hibaértéke 6 és 8 közötti, azonban az eredeti BDI érték növekedésével a modell pontossága folyamatosan romlik. Ennek több oka is lehetséges, elsősorban az, hogy az egyes depressziós adatbázisokban relatív kevés a súlyos depressziós személyek aránya. Azonban megfigyelhető, hogy a legkisebb RMSE hibaérték a 10-14-es BDI csoportban kaptam, ami lényegében a nem depressziós és depressziós állapot határa.

Továbbiakban megvizsgáltam még, hogy az adott regressziós modell alkalmazásával mekkora pontosságot lehet elérni a depressziós állapot detektálásában (kétosztályos osztályozása).

A regressziós modellek tesztelésével előállított tévesztési mátrixok a 22. táblázatban láthatók a kétosztályos tesztelés esetén.

27. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti eltérő modellel, válogatott akusztikai-fonetikai jellemzők alapján, a magyar+német egyesített depressziós beszédadatbázissal tanítva

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**82%**	18%
	Depressziós	26%	**74%**

78%-os pontosságot értem el (27. táblázat) kétosztályos osztályozás esetén nemek szerinti eltérő modellel, statisztikai több nyelvre érvényes jellemző kiválasztási módszerrel, regressziós modellek segítségével. Az eredményt SVR gépi tanuló eljárással értem el RBF kernellel (cost = 2^3 és gamma = 2^{-3}). A depressziós állapot detektálására 82%-os specificitást és 74%-os szenzitivitást értem el.

IV.2. tézis: [B1, C1, C2, C3, J1, J2] Megmutattam, hogy lehetséges olyan SVR alapú regressziós modell létrehozása, amely megfelelő jellemző kiválasztási algoritmus használatával, a tanító beszédmintáktól eltérő anyanyelvű beszélők (vizsgált nyelvek: magyar, német, olasz) depressziós állapotának becslésére is alkalmas.

8.3. Depressziós állapot súlyosságának több nyelvre érvényes becslésére kapott eredményeim összefoglalása

A depressziós állapot gépi becslésének nyelvfüggetlen vizsgálata egy kevéssé kutatott terület, és annak mélyebb vizsgálata mind ez ideig nem történt meg. Ennek elsődleges oka, hogy a kezdeti (európai nyelveken készült) depressziós beszédadatbázisok mind angol nyelvűek voltak, és csak 2013-ban publikálták az első német nyelvű depressziós beszédadatbázist.

Statisztikai vizsgálatok segítségével igazoltam, hogy kiválasztható az alkalmazott akusztikai-fonetikai jellemzők olyan részhalmaza, ami az általam vizsgált magyar és német nyelv esetében, illetve a két nyelv egyesítésével létrehozott beszédadatbázis esetében is szignifikáns korrelációt mutat a depressziós állapot súlyosságával.

Regressziós modellek létrehozásával bemutattam, hogy lehetséges a depressziós állapot súlyosság szerinti több nyelvre érvényes becslésének megvalósítása. Nemek szerinti eltérő modelleket hoztam létre, ahol nemek szerinti eltérő statisztikai alapú jellemzővektor kiválasztást végeztem. Megvizsgáltam, hogy a becslés mekkora RMSE hibaértékkel valósítható meg, ami alapján arra jutottam, hogy a több nyelvre érvényes modellemmel kapott eredmény majdnem olyan pontossággal képes becsülni a depressziós állapot súlyosságát, mint más kutatók nyelvspecifikus megoldásai. A magyar nyelvspecifikus regressziós modellemmel összehasonlítva pedig, hasonló pontosságot értem el, mint amikor a származtatott akusztikai-fonetikai jellemzőket beszédhangszintű szegmentálás felhasználása nélkül számítottam (7.1. fejezet 19. táblázat).

Megvizsgáltam még a magyar+német beszédmintákkal tanított regressziós modellem esetében, hogy a bemondók depressziós állapotuk súlyosság szerinti csoportosítása esetében, hogyan alakul a modellemmel elért becslés pontossága. A vizsgálat alapján megállapítottam, hogy a nem depressziós/depressziós állapot határán tapasztalható a legkisebb hiba a modellem által kapott becslések esetében.

A regressziós modellem depressziós állapot detektálására adaptált tesztelésével 78%-os pontosságot értem el. Az eredményeimet Alghowinem és társai 2016-ban publikált kutatásával lehetséges összehasonlítani. A kutatásban három (angol, amerikai angol és német) depressziós beszédadatbázis esetén vizsgálták a depressziós állapot több nyelvre érvényes gépi detektálási lehetőségeit. Azokban az esetekben, amikor a modellt két adatbázissal tanították, és LOOCV eljárással tesztelték, 76-85%-os pontossággal voltak képesek elkülöníteni a nem depressziós és depressziós beszélőket. Azonban fontos kiemelnem, hogy az általuk vizsgált depressziós beszédadatbázisok csupán nem depressziós és súlyosan depressziós beszélők beszédmintáit tartalmazták.

Az általam elvégzett akusztikai-fonetikai jellemzők kétnyelvű statisztikai vizsgálata, illetve a depressziós állapot súlyossága szerinti többnyelvű gépi becslése nemcsak a hazai, hanem a nemzetközi szakirodalomban is újnak, illetve újszerűnek számítottak publikálásuk idejében (2015-2017).

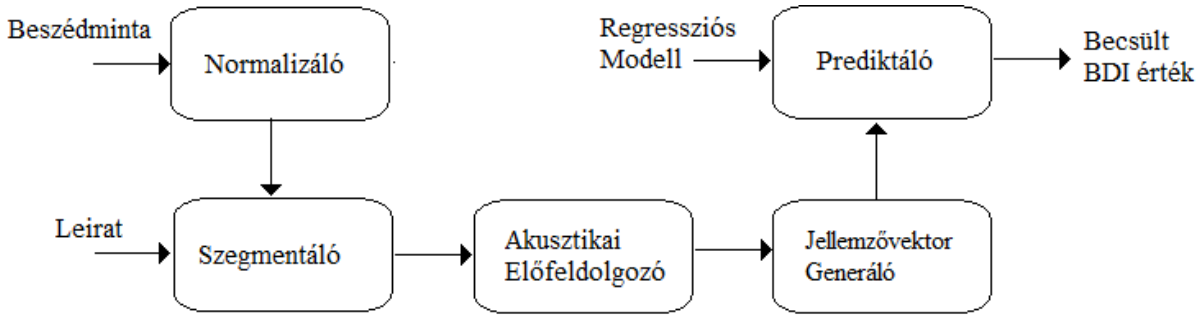
9. Az eredmények gyakorlati alkalmazhatósága

Részt vettem a European Space Agency (ESA) Coala projektjében. A projekt célja a Concordia kutatóállomáson áttelelő kutatók pszichológia állapotának beszédhang alapú megfigyelése volt, különösképpen a depressziós állapot kialakulásának jelzésére.

A projekt keretei közt kidolgoztam és megvalósítottam a beszélő depressziós állapotát beszédminta alapján becslő rendszert. A megvalósított rendszer többnyelvű modell segítségével francia, magyar, német és olasz nyelven használható. A megvalósított rendszer segítségével a Concordia kutatóállomáson dolgozó személyek depressziós állapotát monitoroztam.

Az állomáson dolgozó francia és olasz kutatók nagyjából kétheti rendszerességgel felmondtak és rögzítettek egy rövid mesét („Az északi szél és a Nap) az anyanyelvükön. A felvételek rögzítésére 2013-ban, 2014-ben és 2016-ban került sor. Minden évben 10-12 kutató önkéntesen vett részt a kísérletben.

A feladatom volt a hangminták elemzése, és az alapján a beszélők depressziós állapotának becslése. Mivel a kutatás során nem állt rendelkezésemre megfelelő francia és olasz nyelvű depressziós adatbázis, így fontos szempont volt, hogy az általam tervezett rendszer alkalmas legyen több nyelven is a depressziós állapot becslésére. Az elkészült rendszer folyamatábrája a 10. ábrán látható.



10. ábra. Depressziós állapot becslésére készült rendszer folyamatábrája

A rendszer bemenetként megkapja a beszédmintát, a beszélő nemét és anyanyelvét, és kimenetként megadja a beszélő becsült depressziós súlyosságát a BDI-skála szerint. A *Normalizáló* egység a vizsgált beszédmintát normalizálja amplitúdó szerint. A *Szegmentáló* egység a normalizált beszédminta és a beszélő nyelvének megfelelő leirat alapján elkészíti a beszédminta beszédhangszintű szegmentálását kényszerített illesztéssel. A szegmentálás alapján az *Akusztikai Előfeldolgozó* egység kinyeri a megfelelő akusztikai-fonetikai

jellemzőket. A *Jellemzővektor Generáló* egység a kiszámított jellemző értékeket a beszélő neme szerinti modell alapján transzformálja (mivel a modell tanítása során a jellemző értékek -1 és 1 közötti értékek közé lettek normalizálva), és előállítja a modellnek megfelelő jellemzővektort. A *Prediktáló* egység az elkészített jellemzővektor és a beszélő neme szerinti modell alapján elvégzi a depressziós állapot becslését a BDI-skála szerint, ami a rendszer végső kimenete.

A modellek tanítását a IV.2 tézisben bemutatott eljárással készítettem a Magyar és Német Depressziós Adatbázisok beszédmintái alapján, annyi módosítással, hogy a statisztikai jellemző kiválasztási algoritmus után FFS algoritmust alkalmaztam a becslés pontosságának növelésére, amivel a tanítómodell LOOCV eljárással való tesztelése esetén a hibaértékeket tovább tudtam csökkenteni (RMSE: 8,4; MAE 6,7).

A kialakított végleges rendszert 12 egészséges, különböző anyanyelvű személyen teszteltem (3 francia, magyar, német és olasz). A tesztelés során a 9,1 RMSE hibaértéket míg 7,3 MAE hibaértéket kaptam.

A végleges rendszer nemcsak a Concordia kutatóállomáson dolgozó kutatók pszichológiai állapotának a becslésére alkalmas, hanem általánosságban egy többnyelvű, a depressziós állapot súlyosság szerinti becslésének olyan diagnosztikát támogató rendszer megvalósításra, ami nem invazívan – egy mese felolvasásával –, a depressziós állapot meglétére figyelmeztet. Például jól alkalmazható lehet előszűrésre a körzeti orvosi rendelőkben.

10. Összegzés és téziseim összefoglalása

Dolgozatomban megvizsgáltam a depressziós állapot hatását az akusztikai-fonetikai jellemzők értékére, valamint a megváltozott akusztikai-fonetikai jellemzők alapján a depressziós állapot gépi detektálásának, továbbá súlyosság szerinti több kategóriájú elkülönítésének, illetve súlyosság szerinti becslésének lehetőségeit.

Statisztikai vizsgálatok segítségével igazoltam, hogy a depressziós állapot hatására szignifikánsan megváltoznak a beszéd egyes származtatott akusztikai-fonetikai jellemzői, és ez a változás több esetben a depresszió súlyosságával szignifikánsan korrelál is.

Megmutattam a jellemzővektor kiválasztásának fontosságát, illetve hogy a nemek szerint eltérő modellek kialakítása és használata jelentős javulást eredményez depressziós állapot becslésében.

Megmutattam, hogy a beszédhang szintű szegmentálás felhasználása az akusztikai-fonetikai jellemzők számításában jelentősen csökkenti a jellemzők által létrehozott modellek becslésének hibáját, illetve hogy a regressziós modellek segítségével nagyobb pontosság érhető el a depressziós állapot detektálásában, valamint a depressziós állapot súlyossága szerinti több kategóriájú elkülönítésében.

Megvizsgáltam a depressziós állapot hatásának az akusztikai-fonetikai jellemzőkre gyakorolt nyelvfüggőségét. Statisztikai vizsgálatok segítségével igazoltam, hogy kiválasztható az akusztikai-fonetikai jellemzők olyan részhalmaza, ami az általam vizsgált magyar és német nyelvű beszédadatbázisok esetében és a két nyelv egyesítésével létrehozott beszédadatbázis esetében is szignifikáns korrelációt mutat a depressziós állapot súlyosságával. Regressziós modellek segítségével bemutattam, hogy lehetséges a depressziós állapot súlyosság szerinti többnyelvű becslésének megvalósítása, és a többnyelvű modellem becslésének pontossága összemérhető más kutatók nyelvspecifikus megoldásainak eredményével.

Kidolgoztam és megvalósítottam a beszélő depressziós állapotát beszédminta alapján becslő rendszert. A megvalósított rendszer többnyelvű modell segítségével angol, francia, magyar, német és olasz nyelven használható. A megvalósított rendszer segítségével a Concordia kutatóállomáson dolgozó személyek depressziós állapotát monitoroztam. A végleges rendszer azonban nemcsak a Concordia kutatóállomáson dolgozó depressziós állapotának a becslésére alkalmas, hanem általánosságban egy többnyelvű, a depressziós állapot súlyosság szerinti becslését célzó, olyan diagnosztikát támogató rendszer megvalósításra, ami nem invazívan –

egy 8-10 mondatos mese felolvasásával –, a depressziós állapot meglétére figyelmeztet. Az általam megvalósított rendszer például jól alkalmazható lehet előszűrésre a körzeti orvosi rendelőkben.

A következőkben tételesen felsorolom a dolgozatomban összefüggésekben is bemutatott téziseimet, valamint megadom azok rövid összefoglalóját is.

I.1. tézis: [B1, C4, C6, C7, J1] Statisztikai úton bizonyítottam, hogy a depressziós állapot szignifikáns változást okoz a beszéd egyes akusztikai-fonetikai alapjellemzőiből származtatott jellemzők átlagos értékében, az egészséges átlaghoz képest. 99%-os szignifikancia szinten az alábbi alapjellemzők származtatott értékei mutattak szignifikáns eltérést mindkét nem esetén: shimmer, F1, F2, B1, B2, mel-sávos energiaértékek (65-390 Hz és 1330-5190 Hz közöttiek), intenzitás.

A tézis állítása szerint a depressziós állapot hatással bír a bemutatott akusztikai-fonetikai jellemzők értékére. A tézisem állítását statisztikai úton igazoltam a Magyar Depressziós Beszédadatbázis segítségével a nem depressziós és depressziós személyek beszédmintáiból számított akusztikai-fonetikai jellemzői alapján. A tézis további következménye, hogy ezáltal statisztikai úton feltehetőleg elkülöníthetők a nem depressziós és depressziós emberek a beszédproduktumuk vizsgálata alapján.

I.2. tézis: [J1] Statisztikai úton bizonyítottam, hogy a beszéd egyes akusztikai-fonetikai alapjellemzőinek származtatott értékei szignifikáns korrelációt mutatnak mindkét nem esetében a beszédminta depresszió súlyosságát leíró Beck Depression Inventory (BDI) skálán mért értékével. 99%-os szignifikancia szinten az alábbi alapjellemzők mutattak szignifikáns korrelációt mindkét nem esetén: shimmer, F1, B2, mel-sávos energiaértékek (65-390Hz; 1518-4688Hz), intenzitás.

A tézis azt állítja, hogy korreláció áll fenn a depresszió súlyossága és a bemutatott akusztikai-fonetikai jellemzők értékei között. A tézisem állítását statisztikai úton igazoltam a Magyar Depressziós Beszédadatbázis segítségével, az adatbázisban található bemondók beszédmintáiból számított akusztikai-fonetikai jellemzőik értéke és a bemondók BDI-skálán mért depressziós állapotuk szerinti súlyosságuk alapján. A tézis az I.1-es tézishez képest abban mond többet, hogy a depressziós állapot hatására nemcsak hogy megváltoznak a bemutatott akusztikai-fonetikai jellemzők értékei, hanem a változás mértéke korrelál a depresszió súlyosságával, aminek a következménye, hogy feltehetőleg nemcsak a depressziós állapot

detektálása lehetséges statisztikai úton, hanem a depressziós állapot súlyosság szerinti becslése is.

II.1. tézis: [C4, J1] Kimutattam, hogy SVM gépi tanuló eljárás alkalmazásával depressziós és egészséges beszélőktől származó beszédminták egymástól elkülöníthetők, továbbá a depresszió négyfokozatú súlyossága szerinti osztályozás is megvalósítható. Kétosztályos osztályozás esetén 68%-os pontosságot, négyosztályos osztályozás esetén 52%-os pontosságot értem el, a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel és nemek szerinti közös modellel.

A tézis az I.1-es tézis feltételezett következményének az igazolása. A tézis állítása szerint SVM alapú osztályozási modell alkalmazásával a depressziós állapot detektálása és súlyosság szerinti négykategóriájú elkülönítése lehetséges. A tézisem állítását a Magyar Depressziós Beszédadatbázis segítségével, SVM alapú osztályozási modellek tesztelésével igazoltam, a modellek építése során az általam vizsgált összes akusztikai-fonetikai jellemzőt felhasználtam. A modellek tesztelése során 68%-os pontosságot értem el a depressziós állapot detektálásában, 52%-os pontosságot értem el a depressziós állapot súlyosság szerinti Beck Depression Inventory által definiált négy kategória elkülönítésében.

II.2. tézis: [J1] Kimutattam, hogy statisztikai alapú jellemzővektor kiválasztással akár 12% relatív hibacsökkenés is elérhető a kétosztályú osztályozás esetében, és akár 11% relatív hibacsökkenés is elérhető a négyosztályú osztályozás esetében a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel kapott eredményekhez képest.

II.3. tézis:[J1] Kimutattam, hogy nemek szerint külön létrehozott és statisztikai jellemző kiválasztással optimalizált modellek alkalmazása jelentősen javítja a depresszió osztályozási pontosságát. Az optimalizált modellel 56% relatív csökkenést értem el a hibásan osztályozott személyek számában a kétosztályú osztályozás esetében, és 35% relatív csökkenést értem el a hibásan osztályozott személyek számában a négyosztályú osztályozás esetében, a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel kapott eredményekhez képest.

A depressziós állapot beszédjel feldolgozás alapú gépi detektálásának hasznosíthatóságát nagyban befolyásolja a gépi detektálás pontossága.

A téziseim állítása szerint a megfelelő jellemzővektor kiválasztásával és a nemek szerinti eltérő modellek létrehozásával és alkalmazásával nagyban javítható a depressziós állapot

detektálásának és súlyosság szerinti négykategóriájú elkülönítésének pontossága, így ezek a modellépítési stratégiák lényegesek és hasznosak egy valós diagnosztikát támogató rendszer megtervezésében, létrehozásában. A téziseim igazolását a Magyar Depressziós Beszédadatbázis segítségével végeztem, és az optimalizált modellek pontosságát az I.1-es tézisemben bemutatott modell pontosságához viszonyítottam.

III.1. tézis: [C5] Kimutattam, hogy SVR gépi tanuló eljárás alkalmazásával a beszélő depressziós állapotának súlyossága becsülhető egyszerű (beszédhangszintű szegmentálást nem igénylő) akusztikai-fonetikai jellemzőkkel. A modelleket nemek szerint elkülönítve, statisztikai jellemző kiválasztással optimalizált módon hoztam létre. Az elért átlagos négyzetes hiba gyökére (RMSE) 8,4, míg az átlagos hibára 6,7 (MAE) értékeket kaptam.

A II. téziscsoport téziseivel igazoltam, hogy lehetséges a depressziós állapot gépi detektálása, illetve súlyosság szerinti több kategóriájú elkülönítése. Azonban felmerül a kérdés, hogy lehetséges-e a depressziós állapot szerinti súlyosság becslése.

A tézis az I.2-es tézis feltételezett következményének igazolása. A tézis állítása szerint lehetséges a depressziós állapot súlyosság szerinti beszédjel feldolgozás alapú gépi becslése, a beszédmintából beszédhangszintű szegmentálás felhasználása nélkül kinyerhető, akusztikai-fonetikai jellemzők alapján. A tézisem állítását a Magyar Depressziós Beszédadatbázis segítségével, SVR alapú regressziós modellek tesztelésével igazoltam, a modellek optimalizálása során alkalmaztam a II.2-es és II.3-as téziseim eredményeit. A regressziós modellem becslésére 8,4 RMSE és 6,7 MAE hibaértékeket kaptam.

III.2. tézis: [C5] A beszédhangszintű feldolgozással végzett jellemzőkinyerés javít a becslés pontosságán. Az egyszerű előfeldolgozással mérhető (beszédhangszintű szegmentálás nélküli) akusztikai-fonetikai jellemzőkkel kialakított modellel elért eredményhez képest 25%-os relatív csökkenést értem el az RMSE hibaértékben, és 29%-os relatív csökkenést értem el a MAE hibaértékben.

A tézis állítása szerint, ha az akusztikai-fonetikai jellemzők számításához felhasználjuk a beszéd beszédhangszintű szegmentálását, úgy javulás érhető el a depressziós állapot súlyosság szerinti becslésében. A tézisem állítását a Magyar Depressziós Beszédadatbázis segítségével SVR alapú regressziós modellek tesztelésével igazoltam. Az akusztikai-fonetikai jellemzők számításánál felhasználtam beszédhangszintű szegmentálást, és így 25%-os relatív csökkenést értem el a RMSE hibaértékben a III.1-es tézisem eredményehez képest.

IV. 1. tézis: [B1, C3, J1] Statisztikai úton kimutattam, hogy az alkalmazott akusztikai-fonetikai jellemzőkészletből kiválasztható olyan részhalmaz, amely magyar és német nyelv esetén is szignifikáns korrelációt mutat a beszélők depressziós állapotuk szerinti súlyosságával. Az alábbi jellemzők mutattak 95%-os szignifikancia szinten szignifikáns korrelációt mindkét nem esetén: f0, jitter, shimmer, B2, intenzitás, AS, RSZH, RoT.

A beszédtechnológiában kifejezetten fontos annak a kérdésnek eldöntése, hogy mely megoldások alkalmazhatók a beszélő nyelvétől függetlenül is. A tézis állítása szerint megadhatók olyan akusztikai-fonetikai jellemzők magyar és német nyelv esetén, amelyek szignifikáns korrelációt mutatnak a beszélő nyelvétől függetlenül a beszélő depressziós állapota szerinti súlyosságával. A tézisem állítását a Magyar Depressziós Beszédadatbázis és a Német Depressziós Beszédadatbázis beszédmintáinak statisztikai elemzésével igazoltam. A tézis következménye, hogy feltehetőleg statisztikai úton több nyelvre érvényes regressziós modell használatával becsülhető a depressziós állapot szerinti súlyosság beszédjel feldolgozás alapján a beszélő nyelvétől függetlenül, több nyelv esetében is.

IV.2. tézis: [B1, C1, C2, C3, J1, J2] Megmutattam, hogy lehetséges olyan SVR alapú regressziós modell létrehozása, amely megfelelő jellemző kiválasztási algoritmus használatával, a tanító beszédmintáktól eltérő anyanyelvű beszélők (vizsgált nyelvek: magyar, német, olasz) depressziós állapotának becslésére is alkalmas.

A tézis az IV.1-es tézis feltételezett következményének igazolása. A tézis állítása szerint lehetséges a depressziós állapot súlyosság szerinti, beszédjel feldolgozás alapú, több nyelvre érvényes gépi becslése. A tézisem állítását a Magyar Depressziós Beszédadatbázis, a Német Depressziós Beszédadatbázis és az Olasz Depressziós Beszédadatbázis segítségével, SVR alapú regressziós több nyelvre érvényes modellek tesztelésével igazoltam, a modellek optimalizálása során alkalmaztam a II.2-es, II.3-as és III.2-es téziseim eredményeit.

11. Rövidítések, jelölések

AS	artikulációs sebesség
B1	első formánsfrekvencia sávszélessége
B2	második formánsfrekvencia sávszélessége
BDI	Beck Depression Inventory / Beck depressziós skála
DSM	The Diagnostic and Statistical Manual of Mental Disorders
f0	alapfrekvencia
F1	első formánsfrekvencia
F2	második formánsfrekvencia
FFS	fast forward selection
H0	null hipotézis
H1	alternatív hipotézis
HAM-D	Hamilton Depression Rating Scale / Hamilton depressziós skála
LOOCV	leave one out crossvalidation / teljes keresztvalidáció
MAE	mean absolute error / átlagos abszolút hiba
RMSE	root mean square error / átlagos négyzetes hiba gyöke
RoT	ratio of transients / tranziensek aránya
RSZH	relatív szünethossz
SAMPA	Speech Assessment Methods Phonetic Alphabet / fonetikus abc
SVM	support vector machine / szupport vektor gép
SVR	support vector regression / support vector regresszió
WHO	World Health Organization / Egészségügyi Világszervezet

12. Irodalomjegyzék

[1] Olaszy Gábor (2010), A beszéd komplex szerkezete. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 9-19.)

[2] Godino-Llorente, J. I., Gomez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering*, *51*(2), 380-384.

[3] Vicsi, K., Sztahó, D. (2012). Recognition of Emotions on the Basis of Different Levels of Speech Segments. *JACIII*, *16*(2), 335-340.

[4] El Ayadi, M., Kamel, M. S., Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, *44*(3), 572-587.

[5] Sztahó, D., Tulics, M. G., Vicsi, K., Valálik, I. (2017). Automatic estimation of severity of Parkinson's disease based on speech rhythm related features. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 11-16).

[6] Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S., ... Nöth, E. (2016). Automatic detection of Parkinson's disease in running speech spoken in three different languages. *The Journal of the Acoustical Society of America*, *139*(1), 481-500.

[7] Sztahó Dávid (2013). "Automatikus érzelem-felismerés akusztikai paraméterek alapján." BME, Informatikai Tudományok Doktori Iskola, Budapest.

[8] Olaszy Gábor (2010). A beszédképzés folyamata. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 19-26.)

[9] Kraepelin, E., (1921.) Manic depressive insanity and paranoia. *J. Nerv. Ment. Dis.* 53, 350.

[10] 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.

[11] Yang, Y., Fairbairn, C., Cohn, J. F. (2013). Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, 4(2), 142-150.

[12] Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Gedeon, T., Breakspear, M., Parker, G. (2013). A comparative study of different classifiers for detecting depression from spontaneous speech. *In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 8022-8026).

[13] Alpert, M., Pouget, E.R., Silva, R.R., (2001). Reflections of depression in acoustic measures of the patient's speech. *J. Affect. Disord*. 66, 59–69.

[14] dr. Simon Lajos pszichiáter szakorvos személyes közlése alapján. Semmelweis Orvostudományi Egyetem Pszichiátriai és Pszichoterápiás Klinika.

[15] Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current Directions in Psychological Science*, 10(5), 173-176.

[16] Hönig, F., Batliner, A., Nöth, E., Schnieder, S., Krajewski, J., (2014). Automatic modelling of depressed speech: relevant features and relevance of gender. *In Proceedings of INTERSPEECH 2014* (pp. 1248–1252).

[17] Alghowinem, S., Goecke, R., Epps, J., Wagner, M., Cohn, J. F. (2016). Cross-Cultural Depression Recognition from Vocal Biomarkers. *In Proceedings of INTERSPEECH 2016* (pp. 1943-1947).

[18] World Health Organization. (2017). *Depression and other common mental disorders: global health estimates* (No. WHO/MSD/MER/2017.2). World Health Organization.

[19] Mathers, C. D., Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, 3(11), e442.

[20] Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D., Saxena, S. (2012). Depression: A global public health concern (www.who.int)

[21] Friedrich, M. J. (2017). Depression is the leading cause of disability around the world. *Jama*, *317*(15), 1517-1517.

[22] Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H. U., Jönsson, B., CDBE2010 Study Group, European Brain Council. (2012). The economic cost of brain disorders in Europe. *European Journal of Neurology*, 19(1), 155-162.

[23] Hawton, K., i Comabella, C. C., Haw, C., Saunders, K. (2013). Risk factors for suicide in individuals with depression: a systematic review. *Journal of Affective Disorders*, 147(1-3), 17-28.

[24] Joiner Jr, T. E., Brown, J. S., Wingate, L. R. (2005). The psychology and neurobiology of suicidal behavior. *Annu. Rev. Psychol*., 56, 287-314.

[25] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

[26] Deckersbach, T., Dougherty, D. D., Rauch, S. L. (2006). Functional imaging of mood and anxiety disorders. *Journal of Neuroimaging*, 16(1), 1-10.

[27] Evans, K. C., Dougherty, D. D., Pollack, M. H., Rauch, S. L. (2006). Using neuroimaging to predict treatment response in mood and anxiety disorders. *Annals of Clinical Psychiatry*, 18(1), 33-42.

[28] Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., ... Kennedy, S. H. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, 45(5), 651-660.

[29] Nestler, E. J., Barrot, M., DiLeone, R. J., Eisch, A. J., Gold, S. J., Monteggia, L. M. (2002). Neurobiology of depression. *Neuron*, 34(1), 13-25.7

[30] Gilbody, S. M., Whitty, P. M., Grimshaw, J. M., Thomas, R. E. (2003). Improving the detection and management of depression in primary care. *BMJ Quality Safety*, 12(2), 149-155.

[31] Luscher, B., Shen, Q., Sahir, N. (2011). The GABAergic deficit hypothesis of major depressive disorder. *Molecular Psychiatry*, 16(4), 383.

[32] 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.

[33] Diekstra, R. F., Gulbinat, W. (1993). The epidemiology of suicidal behaviour: a review of three continents. *World Health Statistics Quarterly*. Rapport trimestriel de statistiques sanitaires mondiales, 46(1), 52-68.

[34] Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., ... De la Torre, F. (2009). Detecting depression from facial actions and vocal prosody. *In 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (pp. 1-7).

[35] Hall, J. A., Harrigan, J. A., Rosenthal, R. (1995). Nonverbal behavior in clinician patient interaction. *Applied and Preventive Psychology*, 4(1), 21-37.

[36] Sobin, C., Sackeim, H. A. (1997). Psychomotor symptoms of depression. *American Journal of Psychiatry*, 154(1), 4-17.

[37] Mitchell, A. J., Vaze, A., Rao, S. (2009). Clinical diagnosis of depression in primary care: a meta-analysis. *The Lancet*, 374(9690), 609-619.

[38] Sikorski, C., Luppa, M., König, H. H., van den Bussche, H., Riedel-Heller, S. G. (2012). Does GP training in depression care affect patient outcome? –A systematic review and meta-analysis. *BMC Health Services Research*, 12(1), 10.

[39] Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99(2), 143.

[40] Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36(3), 189-208.

[41] Kent, R. D. (2000). Research on speech motor control and its disorders: A review and prospective. *Journal of Communication Disorders*, 33(5), 391-428.

[42] Croarkin, P. E., Levinson, A. J., Daskalakis, Z. J. (2011). Evidence for GABAergic inhibitory deficits in major depressive disorder. *Neuroscience Biobehavioral Reviews*, 35(3), 818-825.

[43] Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition*, 56(1), 30-35.

[44] Ellgring, H., Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, 20(2), 83-110.

[45] Moore II, E., Clements, M. A., Peifer, J. W., Weisser, L. (2008). Critical analysis of the impact of glottal features in the classification of clinical depression in speech. *IEEE Transactions on Biomedical Engineering*, 55(1), 96-107.

[46] Quatieri, T. F., Malyska, N. (2012). Vocal-source biomarkers for depression: A link to psychomotor activity. *In Proceedings of INTERSPEECH 2012* (pp. 1059-1062)

[47] 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), 42.

[48] Murphy, F. C., Sahakian, B. J., Rubinsztein, J. S., Michael, A., Rogers, R. D., Robbins, T. W., Paykel, E. S. (1999). Emotional bias and inhibitory control processes in mania and depression. *Psychological Medicine*, 29(6), 1307-1321.

[49] 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.

[50] Enge, S., Fleischhauer, M., Lesch, K. P., Reif, A., Strobel, A. (2011). Serotonergic modulation in executive functioning: linking genetic variations to working memory performance. *Neuropsychologia*, 49(13), 3776-3785.

[51] Christopher, G., MacDonald, J. (2005). The impact of clinical depression on working memory. *Cognitive Neuropsychiatry*, 10(5), 379-399.

[52] Levelt, W. J., Roelofs, A., Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-38.

[53] 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.

[54] Hamilton, H., 1960. HAMD: a rating scale for depression. Neurosurg. Psych. 23, 56–62.

[55] Beck, A. T., Epstein, N., Brown, G., Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology*, 56(6), 893.

[56] Beck, A.T., Steer, R.A., Ball, R., Ranieri, W.F., (1996). Comparison of beck depression inventories-ia and-ii in psychiatric outpatients. *J. Pers. Assess*. 67, 588–597.

[57] France, D. J., Shiavi, R. G., Silverman, S., Silverman, M., Wilkes, M. (2000). Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions on Biomedical Engineering*, 47(7), 829-837.

[58] Moore, E. I. I., Clements, M., Peifer, J., Weisser, L. (2004, September). Comparing objective feature statistics of speech for classifying clinical depression. *In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 1, pp. 17-20).

[59] Low, L. S. A., Maddage, N. C., Lech, M., Allen, N. (2009, June). Mel frequency cepstral feature and Gaussian Mixtures for modeling clinical depression in adolescents. *In 8th IEEE International Conference on Cognitive Informatics* (pp. 346-350).

[60] Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., Parker, G. (2012). From Joyous to Clinically Depressed: Mood Detection Using Spontaneous Speech. *In FLAIRS Conference* (pp. 141-146)

[61] Yingthawornsuk, T., Keskinpala, H. K., France, D., Wilkes, D. M., Shiavi, R. G., Salomon, R. M. (2006). Objective estimation of suicidal risk using vocal output characteristics. *In INTERSPEECH 2006 – ICSLP* (pp. 649.652)

[62] 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.

[63] Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., ... De la Torre, F. (2009). Detecting depression from facial actions and vocal prosody. *In 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops* (pp. 1-7).

[64] 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.

[65] Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., ... Pantic, M. (2013, October). AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. *In Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge* (pp. 3-10)

[66] Valstar, M., Schuller, B., Smith, K., Almaev, T., Eyben, F., Krajewski, J., ... Pantic, M. (2014, November). Avec 2014: 3d dimensional affect and depression recognition challenge. *In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge* (pp. 3-10).

[67] Darby, J. K., Hollien, H. (1977). Vocal and speech patterns of depressive patients. *Folia Phoniatrica et Logopaedica*, 29(4), 279-291.

[68] Hollien, H. (1980). Vocal indicators of psychological stress. *Annals of the New York Academy of Sciences*, 347(1), 47-72.

[69] Nilsonne, A. (1988). Speech characteristics as indicators of depressive illness. *Acta Psychiatrica Scandinavica*, *77*(3), 253-263.

[70] 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.

[71] Darby, J. K., Simmons, N., Berger, P. A. (1984). Speech and voice parameters of depression: A pilot study. *Journal of Communication Disorders*, 17(2), 75-85.

[72] 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.

[73] Laukkanen, A. M., Björkner, E., Sundberg, J. (2006). Throaty voice quality: subglottal pressure, voice source, and formant characteristics. *Journal of Voice*, 20(1), 25-37.

[74] Moore II, E., Clements, M. A., Peifer, J. W., Weisser, L. (2008). Critical analysis of the impact of glottal features in the classification of clinical depression in speech. *IEEE Transactions on Biomedical Engineering*, 55(1), 96-107.

[75] Ooi, K. E. B., Lech, M., Allen, N. B. (2013). Multichannel weighted speech classification system for prediction of major depression in adolescents. *IEEE Transactions on Biomedical Engineering*, 60(2), 497-506.

[76] Cummins, N., Epps, J., Breakspear, M., Goecke, R. (2011). An investigation of depressed speech detection: Features and normalization. *In Proceedings of INTERSPEECH 2011* (pp. 2997-3000)

[77] Cummins, N., Epps, J., Ambikairajah, E. (2013). Spectro-temporal analysis of speech affected by depression and psychomotor retardation. *In IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 7542-7546).

[78] Cummins, N., Epps, J., Sethu, V., Breakspear, M., Goecke, R. (2013). Modeling spectral variability for the classification of depressed speech. *In Proceedings of INTERSPEECH 2013* (pp. 857-861).

[79] Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., Parker, G. (2013, May). Detecting depression: a comparison between spontaneous and read speech. *In IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 7547-7551). IEEE.

[80] Helfer, B. S., Quatieri, T. F., Williamson, J. R., Mehta, D. D., Horwitz, R., Yu, B. (2013). Classification of depression state based on articulatory precision. *In Proceedings of INTERSPEECH 2013* (pp. 2172-2176).

[81] Williamson, J. R., Quatieri, T. F., Helfer, B. S., Horwitz, R., Yu, B., Mehta, D. D. (2013, October). Vocal biomarkers of depression based on motor incoordination. *In Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, (pp. 41-48).

[82] Kiss, G., Sztahó, D., Vicsi, K. (2013). Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. In *IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 579-582).

[83] Sándor, B. (1997). Bevezetés a matematikai statisztikába. Kossuth Egyetemi Kiadó.

[84] Schultz, B. B. (1985). Levene's test for relative variation. Systematic Zoology, 34(4), 449-456.

[85] Blalock, H. M. (1972). Social statistics. New York: McGraw-Hill. (p. 625)

[86] IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

[87] Russell, S. J., Norvig, P., Antal, P., Dobrowiecki, T., Haidegger, G. (2000). Mesterséges intelligencia: modern megközelítésben. Panem Kiadó

[88] Hofmann, M., Klinkenberg, R. (Eds.). (2013). RapidMiner: Data mining use cases and business analytics applications. Chapman and Hall/CRC (p. 525)

[89] Cortes, C., Vapnik, V. (1995). Support vector machine. *Machine Learning*, 20(3), 273-297.

[90] Suykens, J. A., Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, *9*(3), 293-300.

[91] Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., Vapnik, V. (1997). Support vector regression machines. In *Advances in Neural Information Processing Systems* (pp. 155-161).

[92] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1--27:27, 2011.

[93] Hall, M. A. (1999). Correlation-based feature selection for machine learning (PhD Dissertation)

[94] 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).

[95] Björn W. Schuller, Anton Batliner, Christian Bergler, Florian B. Pokorny, Jarek Krajewski, Margaret Cychosz, Ralf Vollmann, Sonja-Dana Roelen, Sebastian Schnieder, Elika Bergelson, Alejandrina Cristià, Amanda Seidl, Anne Warlaumont, Lisa Yankowitz, Elmar Nöth, Shahin Amiriparian, Simone Hantke, Maximilian Schmitt: "The INTERSPEECH 2019 Computational Paralinguistics Challenge: Styrian Dialects, Continuous Sleepiness, Baby Sounds Orca Activity", *Preliminary in Proceedings of INTERSPEECH 2019, ISCA, Graz, Austria, 2019.*

[96] Olaszy Gábor (2010), A beszédképzés folyamata. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 19-26.)

[97] Vicsi Klára (2010). Érzelemfelismerés. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 399-401.)

[98] Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

[99] Horii, Y. (1982). Jitter and shimmer differences among sustained vowel phonations. *Journal of Speech, Language, and Hearing Research*, 25(1), 12-14.

[100] Vicsi Klára (2010). A beszéd fizikai jellemzése. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 39-56.)

[101] Childers, D. G. (Ed.). (1978). Modern spectrum analysis. *IEEE Computer Society Press*. (pp. 252-255)

[102] Vicsi Klára (2010). Pszichofizikai tényezők. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 56-69.)

[103] Olaszy Gábor (2010). A beszéd szupraszegmentális szerkezete. *A magyar beszéd*, szerk. Németh Géza, Olaszy Gábor. Akadémia Kiadó (pp. 171-205.)

[104] Vlasenko, B., Sagha, H., Cummins, N., Schuller, B. W. (2017). Implementing Gender-Dependent Vowel-Level Analysis for Boosting Speech-Based Depression Recognition. *In Proceedings of INTERSPEECH 2017* (pp. 3266-3270).

[105] Cummins, N., Vlasenko, B., Sagha, H., Schuller, B. (2017, June). Enhancing speech-based depression detection through gender dependent vowel-level formant features. *In Conference on Artificial Intelligence in Medicine in Europe* (pp. 209-214).

[106] Tulics, M. G., Vicsi, K. (2017, September). Phonetic-class based correlation analysis for severity of dysphonia. *In 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 21-26).

[107] Wang, J., Kothalkar, P. V., Cao, B., Heitzman, D. (2016). Towards Automatic Detection of Amyotrophic Lateral Sclerosis from Speech Acoustic and Articulatory Samples. *In Proceedings of INTERSPEECH 201*6 (pp. 1195-1199).

[108] Weiner, J., Herff, C., Schultz, T. (2016). Speech-Based Detection of Alzheimer's Disease in Conversational German. *In Proceedings of INTERSPEECH 201*6 (pp. 1938-1942).

[109] Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., ... Kálmán, J. (2015). Automatic detection of mild cognitive impairment from spontaneous speech using ASR. *In Proceedings of INTERSPEECH 201*5 (pp. 2694-2698)

13. A szerző tudományos közleményei

13.1. Értekezéshez szorosan kapcsolódó publikációk

Cikkek lektorált külföldi folyóiratban [12 p]

[J1] Kiss, G., Vicsi, K. (2017). Mono-and multi-lingual depression prediction based on speech processing. *International Journal of Speech Technology*, *20*(4), 919-935.
(BME-PA pontszám: 100% *6p = 6p.) Scopus / Web of Science

[J2] Kiss, G., Vicsi, K. (2015). Seasonal affective disorder speech detection on the base of acoustic phonetic speech parameters. *Acta Universitatis Sapientiae Electrical and Mechanical Engineering*, 7, 62-79.
(BME-PA pontszám: 100% *6p = 6p.)

Cikkek szerkesztett könyvben [1,5 p]

[B1] Kiss, G., Tulics, M. G., Sztahó, D., Esposito, A., Vicsi, K. (2016). Language independent detection possibilities of depression by speech. In *Recent Advances in Nonlinear Speech Processing* (pp. 103-114). Springer, Cham.
(BME-PA pontszám: 25% *6p = 1,5p.)

Konferenciacikkek [15,5 p]

[C1] Kiss, G., Módos, A., Vicsi, K. (2018) Comparison of the predicted depressed state of crew members with the results of their subjective psychological test at Concordia research station. *In Selected papers of 4th International Conference on Resarch, Technology and Educaion of Space* (p. 11).
(BME-PA pontszám: 50% *3p = 1,5p.)

[C2] Kiss, G., Vicsi, K. (2017). Follow-up Psychological Status Monitoring of the Crew Members of Concordia Research Station at Antarctica Based on their Speech. In *Selected papers of 3rd International Conference on Research, Technology and Education of Space* (p. 37).
(BME-PA pontszám: 100% *3p = 3p.)

[C3] Kiss, G., Vicsi, K. (2017). Investigation of cross-lingual depression prediction possibilities based on speech processing. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 97-102).
(BME-PA pontszám: 100% *3p = 3p.)

[C4] Kiss, G., Vicsi, K. (2017). Comparison of read and spontaneous speech in case of automatic detection of depression. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 213-218).
(BME-PA pontszám: 100% *3p = 3p.)

[C5] Kiss, G., Simon, L., Vicsi K. (2017). Depresszió súlyosságának becslése beszédjel alapján magyar nyelven. In: Vincze Veronika (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017)*. pp. 125-135.
(BME-PA pontszám: 50% *1p = 0,5p.)

[C6] Kiss, G., Vicsi, K. (2014). Physiological and cognitive status monitoring on the base of acoustic-phonetic speech parameters. In *International Conference on Statistical Language and Speech Processing* (pp. 120-131). Springer, Cham.
(BME-PA pontszám: 100% *3p = 3p.)

[C7] Vicsi, K., Sztahó, D., Kiss, G. (2012). Examination of the sensitivity of acoustic-phonetic parameters of speech to depression. In *3rd International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 511-515). IEEE.
(BME-PA pontszám: 50% *3p = 1,5p.)

13.2. A tématerületen megjelent egyéb publikációk

Cikkek lektorált külföldi folyóiratban [3 p]

[J3] Sztahó, D., Kiss, G., Vicsi, K. (2018). Computer based speech prosody teaching system. *Computer Speech Language*, *50*, 126-140.
(BME-PA pontszám: 50% *6p = 3p.) Scopus / Web of Science

Cikkek lektorált hazai folyóiratban [3,5 p]

[J4] Sztahó, D., Kiss, G., Tulics, M. G., Czap, L., Vicsi K. (2014) Számítógéppel támogatott prozódiaoktató program. *Alkalmazott Nyelvészeti Közlemények* 9:(1) 144-153. (2014)
(BME-PA pontszám: 25% *2p = 0,5p.)

[J5] Kiss, G., Vicsi, K. (2013). Akusztikai hangosztályok felismerésén alapuló, nemlineáris idővetemítés megvalósítása a mondathanglejtés és a szóhangsúlyozás oktatásához. *Beszédkutatás* 21: 247-260.
(BME-PA pontszám: 100% *2p = 2p.)

[J6] Vicsi, K., Sztahó, D., Kiss, G. (2011). Nem verbális hangjelenségek spontán társalgásban. *Beszédkutatás* 19: 134-147.
(BME-PA pontszám: 50% *2p = 1p.)

Konferenciacikkek [7,58 p]

[C8] Sztahó, D., Kiss, G., Tulics, M. G., Vicsi, K. (2018). Automatic Separation of Various Disease Types by Correlation Structure of Time Shifted Speech Features. In *41st International Conference on Telecommunications and Signal Processing (TSP)* (pp. 1-4).
(BME-PA pontszám: 33% *3p = 1p.)

[C9] Tündik, M. Á., Kiss, G., Sztahó, D., Szaszák, Gy. (2017) Automatikus frázisdetektáló módszereken alapuló patológiás beszédelemzés magyar nyelven. *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)* (pp. 113-124)
(BME-PA pontszám: 25% *1p = 0,25p.)

[C10] Tündik, M. Á., Kiss, G., Sztahó, D., Szaszák, Gy. (2017). Assessment of pathological speech prosody based on automatic stress detection and phrasing approaches. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 67-72).
(BME-PA pontszám: 25% *3p = 0,75p.)

[C11] Kovács, A., Kiss, G., Vicsi, K., Winkler, I., Coath, M. (2015). Comparison of skewness-based salient event detector algorithms in speech. In *6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 285-290).
(BME-PA pontszám: 25% *3p = 0,75p.)

[C12] Sztahó, D., Kiss, G., Vicsi, K. (2015). Estimating the severity of Parkinson's disease from speech using linear regression and database partitioning. *In Proceedings of INTERSPEECH 201*5 (pp. 498-502)
(BME-PA pontszám: 50% *3p = 1,5p.)

[C13] Kiss, G., Sztahó, D., Vicsi, K., Golemis, A. (2014). Connection between body condition and speech parameters-especially in the case of hypoxia. In *5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 333-336).
(BME-PA pontszám: 33% *3p = 1p.)

[C14] Sztahó, D., Kiss, G., Czap, L., Vicsi, K. (2014). A computer-assisted prosody pronunciation teaching system. In *WOCCI 2014* (pp. 45-49).

(BME-PA pontszám: 33% *3p = 1p.)

[C15] Kiss, G., Sztahó, D., Vicsi, K. (2013). Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. In *4th international conference on cognitive infocommunications (CogInfoCom)* (pp. 579-582). (BME-PA pontszám: 33% *3p = 1p.)

[C16] Vicsi, K., Sztahó, D., Kiss, G., Czira, A. (2010) Spontán beszédben rejlő nem verbális hangjelenségek – érzelmek, hanggesztusok – vizsgálata. In: Tanács Attila, Vincze Veronika (szerk.) *VII. Magyar Számítógépes Nyelvészeti Konferencia MSZNY 2010*. pp. 249-260.
(BME-PA pontszám: 33% *1p = 0,33p.)