

SEMMELWEIS EGYETEM
DOKTORI ISKOLA

Ph.D. értekezések

3202.

HAJDUSKA-DÉR BÁLINT

Pszichiátria

című program

Programvezető: Dr. Réthelyi János, egyetemi tanár

Témavezető: Dr. Simon Lajos, egyetemi docens

Depresszióra jellemző hangtani paraméterek vizsgálata, automatikus osztályozó rendszer megalkotása

Doktori értekezés

Dr. Hajduska-Dér Bálint

Semmelweis Egyetem
Mentális Egészségtudományok Doktori Iskola



Témavezető: Dr. Simon Lajos, egyetemi docens

Hivatalos bírálók: Dr. Hoffmann Ildikó, Ph.D., tudományos főmunkatárs
Dr. Balogh Livia, Ph.D., egyetemi adjunktus

Komplex vizsga szakmai bizottság:

Elnök: Dr. Purebl György, Ph.D., egyetemi tanár

Tagok: Dr. Györffy Zsuzsa, Ph.D., egyetemi docens
Dr. Vizin Gabriella, Ph.D., egyetemi docens

Budapest
2025

Tartalomjegyzék

<i>Rövidítések jegyzéke</i>	5
<i>1. Bevezetés</i>	7
1.1. A depresszió története	9
1.2. A depresszió típusai, tünetei	11
1.3. A depresszió epidemiológiája, társadalmi fontossága	11
1.4. Biomarkerek a depresszió diagnosztizálásában	12
1.5. A vokális csatorna, mint biomarker	15
1.6. A mesterséges intelligencia típusai.....	17
1.7. A mesterséges intelligencia főbb formái	18
1.8. A mesterséges intelligencia alkalmazása a pszichiátriában.....	22
1.9. A mesterséges intelligencia alkalmazása depresszióban	24
1.10. Depresszió és beszédelemzés mesterséges intelligenciával	26
1.11. A mesterséges intelligencia etikai vonatkozásai	27
<i>2. Célkitűzések</i>	30
2.1. A depresszióban végzett gépi tanuló eljárással támogatott hangelemzési módszerek kutatási gyakorlatának és eredményeinek áttekintése	30
2.2. Akusztikai rendszer létrehozása gépi tanuló eljárás segítségével a depresszió szűrésére	30
<i>3. Módszerek</i>	32
3.1. Az első vizsgálat módszerei	32
3.2. A második vizsgálat módszerei	33
3.2.1. A vizsgálatban résztvevő személyek	33
3.2.2. Alkalmazott mérőeszközök	34
3.2.3. Az adatbázis felépítése	35
3.2.4. A vizsgálat menete.....	37

3.2.5. Előfeldolgozás, jellemzők kinyerése	37
3.2.6. Modelltanítás, optimalizálás és statisztikai elemzés.....	38
4. <i>Eredmények</i>	41
4.1. Az első vizsgálat eredményei	41
4.1.1. Szisztematikus irodalomelemzés	41
4.1.2. Az akusztikai elemzés formái és módszerei	44
4.1.3. A depresszióval összefüggésbe hozható hangtani jellemzők	48
4.1.4. A prediktív modell teljesítménye	50
4.2. A második vizsgálat eredményei	51
4.2.1. A depressziós súlyosságának becslése BDI alapján	51
4.2.2. A depressziós súlyosságának becslése H2B alapján	55
4.2.3. A depressziós és egészséges alanyok elkülönítése	57
4.2.4. Az akusztikus modell és a BDI pontosságának összehasonlítása.....	58
5. <i>Megbeszélés</i>	60
5.1. Az első vizsgálat megbeszélése	60
5.1.1. Az első vizsgálat korlátai	62
5.2. A második vizsgálat megbeszélése.....	62
5.2.1. A második vizsgálat korlátai	65
6. <i>Következtetések</i>	66
6.1. Az első vizsgálat következtetései	66
6.2. A második vizsgálat következtetései	67
7. <i>Összefoglalás</i>	68
<i>Summary</i>	69
8. <i>Irodalomjegyzék</i>	70
9. <i>Saját publikációk jegyzéke</i>	78
9.1. A disszertációhoz tartozó közlemények	78
9.2. A disszertációtól független közlemények	78
10. <i>Köszönetnyilvánítás</i>	80

<i>Táblázatok és ábrák jegyzéke</i>	<i>81</i>
Táblázatok	81
Ábrák	81

Rövidítések jegyzéke

5-HTTLPR – szerotonin transzporter gén hosszúság-polimorfizmusa
AGI – Artificial General Intelligence
ANI – Artificial Narrow Intelligence
ASI – Artificial Superintelligence
AUC – Area Under the Curve
AVEC – Audio/Visual Emotion Challenge Workshop
BDI – Beck Depression Inventory
BDNF – agyi neurotrofikus faktor
CCC – Concordance correlation coefficient
CNN – Convolution Neural Network
CRP – C-reaktív fehérje
DAIC-WOZ – Distress Analysis Interview Corpus Wizard of Oz adatbázis
DALY – disability-adjusted life years
DNS – Deoxiribonukleinsav
DSM – Diagnostic and Statistical Manual of Mental Disorders
DT – Decision Tree
EEG – elektroencephalogram
fMRI – funkcionális mágneses rezonancia képalkotás
GDPR – General Data Protection Regulation
GMM – Gaussian Mixture Models
HAMD – Hamilton Depression Scale
HMM – rejtett Markov-modell
HNR – harmonic to noise ratio
HPA-tengely – hipotalamusz-hipofízis-mellékvese tengely
HRV – szívfrekvencia-variabilitás
IL-6 – Interleukin-6
LLD – low-level descriptors
LOOCV – leave-one-out cross-validation
MI – mesterséges intelligencia
MFCC – mel-frekvencia cepstrális együttthatója
ML – machine learning, gépi tanulás

MRI – mágneses rezonancia képalkotás
NER – named entity recognition
NLP – Natural Language Processing
PET – pozitron emissziós tomográfia
PHQ-9 – Patient Health Questionnaire-9
PTSD – Poszttraumás-stresszavar
RBF – radiális bázisfüggvény
REM – Rapid Eye Movement, gyors szemmozgás
RF – Random Forest
RMSE – átlagos négyzetes hiba négyzetgyöke
ROC – receiver operating characteristic
SAMPA – Speech Assessment Methods Phonetic Alphabet
SES – Socioeconomic status
SNRI – szelektív szerotonin és noradrenalin visszavétel gátlók
SSRI – szelektív szerotonin visszavétel gátlók
SVM – Support Vector Machine
TNF- α – tumor nekrosis faktor-alfa
WHO – World Health Organisation
XAI – explainable AI

1. Bevezetés

A depresszió a hangulat és az érdeklődés beszűkülésével járó pszichiátriai betegség, ami bármely életkorban jelentkezhets és jelentősen rontja a páciensek életminőségét.

Az életminőség romlása mellett a betegség nagy kockázati tényezőt jelent az öngyilkosságban, valamint kifejezetten nagy gazdasági teherrel is jár a munkaképesség csökkenésén keresztül. Mindezek miatt a mielőbbi szűrése és diagnosztikája a depresszív zavaroknak nagy fontossággal bír.

A depressziós zavarok diagnosztikájában a mindennapokban különböző szűrőtesztek alkalmazása mellett pszichiáter szakorvosi, klinikai szakpszichológusi vizsgálat szükséges, ahol a diagnosztikai kritériumok feltárásán keresztül történik meg a diagnózis alkotás. A legelterjedtebb szűrő eljárások között megkülönböztetünk önkitöltős kérdőíveket (mint pl. a Beck Depression Inventory, BDI), valamint klinikus által felvett interjúkat (mint pl. a Hamilton Depression Scale, HAMD). Az önkitöltős kérdőívek hátránya azok szubjektivitása, ami miatt a depressziós állapotot alul- vagy túlbecsülhetik. A klinikus által felvett interjúk szakképzettséget követelnek meg, így ember- és időigényesek.

Az elmúlt évtizedekben a depresszió területén is előtérbe kerültek a különböző biomarker kutatások, melyekkel a diagnosztika objektivizálhatóvá, felgyorsíthatóvá, valamint költséghatékonyá válhat. Klinikai és preklinikai tanulmányok számos, depresszióra jellemző biomarkerként alkalmazható faktort találtak, úm. növekedési hormon (GF), proinflammatorikus cytokinek, endokrin faktorok (hypothalamus-hypophysis-mellékvesekéreg tengely, pajzsmirigy, nemi hormonok), metabolikus markerek (pl. inzulin rezisztencia), azonban ezeknek az eredményeknek limitációja a szenzitivitás és specificitás hiánya, valamint egyikük sem szolgált egyedüli meghatározó szereppel. (1) A metabolikus és hormonális biomarkereken kívül az elektroencephalogram (EEG) kutatások is előtérben állnak a depresszió diagnózisalkotásában. Alvási EEG vizsgálatok során gyors szemmozgás (Rapid Eye Movement, REM) látencia csökkenése, valamint csökkent lassú hullámú alvás, 2. alvás fázis voltak mérhetőek a depressziós mintában. (2) Ezeknek a lehetséges biomarkereknek egyik nagy hátránya, hogy használatukhoz speciális laboratóriumi háttérre, alapellátásban nem hozzáférhető eszközökre (EEG) van szükség.

A vokális csatorna vizsgálata már évtizedek óta a kutatások tárgyát képezi depresszióban. A beszédjelfeldolgozás előnye, hogy nem invazív, valamint idő- és költséghatékony. A kezdeti kutatások a hangtani paraméterek kvalitatív elemzésére voltak alkalmasak, emellett a beszéd sebességének, szünetek hosszának mérése alapján vontak le következtetéseket. A számítástechnika fejlődése a beszédminták elemzésében, a depresszió gépi detektálásában nagy előrelépést hozott. Ehhez a mesterséges intelligenciát használó rendszerek elterjedése is nagy segítséget jelentett, főképp a gépi tanuló eljárásokon keresztül. A depresszió hangtani paraméterek alapján történő automatikus detektálása során lehetőség van a depressziós és nem depressziós alanyok osztályozással történő elkülönítésére, vagy a páciens depressziójának súlyosságának regressziós módszerrel történő becslésére. A szakirodalom alapján 60-95%-os pontossággal lehet megkülönböztetni az egészséges és a depressziós betegeket. (3)

Kutatásunkban egy olyan felügyelt gépi tanulást alkalmazó rendszer létrehozásán dolgoztunk, ami képes a beszéd alapján a depressziós és egészséges beszédet elkülöníteni egymástól. Vizsgálatunk tárgya volt az is, hogy a klinikus által felvett diagnosztikai interjú (Hamilton Depression Scale, HAMD) segítségével, milyen mértékben javítható a rendszer pontossága az önkitöltős klinikai teszten (Beck Depression Inventory, BDI) tanított rendszerhez képest. A vizsgálat során depressziós páciensek beszédét, előfeldolgozást követően elemeztük, az elkülönítésben leghatékonyabban használható akusztikai paraméterek azonosítása céljából. Ezt követően gépi tanuló eljárás segítségével egy automatikus osztályozó rendszert hoztunk létre, ami a BDI és HAMD pontszámokon történő betanítást követően képes a depressziós hanganyagok elkülönítésére. (4)

A depresszió esetében az akusztikai paraméterek és azok automatikus detektálása egyre inkább a figyelem középpontjába került az utóbbi években. A másik vizsgálatunkban arra voltunk kíváncsiak, hogy milyen kutatási eredmények és gyakorlatok állnak rendelkezésre jelenleg, valamint milyen konstruktív javaslatokat tudnánk megfogalmazni ezek alapján a jövőbeli kutatások számára. (5)

A disszertációban elsőként, a kutatás szempontjából kiemelt jelentőségű témaköröket tekintjük át. A bevezetésben a depresszió történeti, epidemiológiai és tünettani összefoglalása mellett, az aktuális biomarker kutatásokat, azon belül a vokális csatorna vizsgálatát tárgyaljuk. Emellett a kutatásunk szempontjából is lényeges mesterséges intelligencia elméleti hátterének, típusainak és formáinak elemzésén keresztül, a

pszichiátriában és azon belül a depresszióban és beszédelemzésben való alkalmazhatóságát foglaljuk össze szakirodalmi áttekintésre támaszkodva. A disszertáció második felében a témában végzett áttekintő tanulmány alapján levont következtetéseinket, valamint az általunk megalkotott automatikus osztályozó rendszer eredményeit ismertetjük részletesen, a témában megjelent közleményeink alapján.

1.1. A depresszió története

A hangulatzavarok vagy más néven affektív zavarok vizsgálata az emberiség történetével egyidős. Már az ókorban is születtek diagnosztikus leírások, melyek idővel pontosabbá, tudományosan egyre megalapozottabbá váltak a mai korunkig. A hangulatzavarok osztályozása az elmúlt évszázadokban több elméleten vonult végig, ennek áttekintésében a Robert J. DeRubeis és Daniel R. Strunk által szerkesztett, *The Oxford Handbook of Mood Disorders* című könyvet veszem alapul. (6)

Első írott anyagaink a depresszióval kapcsolatban az i.e. 5.századból származnak Hippocrates és tanítványainak leírása alapján. Melankóliaként fogalmazták meg a hosszantartó félelem érzésével vagy szomorúsággal együttjáró állapotot, ami álmatlansággal, ételtől való irtózással, nyugtalansággal, reménytelenséggel, irritáltsággal jár együtt. A hippocratesi elmélet szerint a betegségek háttérében álló négy nedv közül a fekete epe túltengése szerepel oki tényezőként a tünetek háttérében. A középkorban a melankólia háttérében erkölcsi és vallási okok kerültek leírásra, így az intenzív szomorúságot gyakran a bűn jelének tekintették. Egészen a reneszánsz korig tartott az elképzelés, ami a melankóliát az erkölcsiséggel állította párhuzamba. A reneszánsz azonban a tudományos gondolkodás virágzásával együtt, a depresszió kutatását is elmozdította.

Robert Burton *A melankólia anatómiája* című könyve (*The Anatomy of Melancholia*, 1621) volt az elindítója annak az új hullámnak, ami elkezdte a melankólia tüneteit, lehetséges okait és formáit megkülönböztetni. Ez az első olyan részletes mű, ami kifejezetten a hangulatzavarokkal foglalkozik és elindítója a további empirikus kutatásoknak.

A 17. századtól kezdődően a Hippocrates óta fennálló testnedv elmélet meghaladottá vált. Az újkori, empirikus megfigyelések és kutatások már az agy és idegrendszer zavariban kezdték el keresni az eltéréseket a hangulatzavarok háttérében. Emil Kraepelin német

pszichiáter (1856-1926) elsőként tett különbséget a hangulatazavarok és pszichotikus kórképek között, ezáltal elkülönítette a „mániás-depressziós tébolyt”, valamint a „dementia preacoxot”, mai megfogalmazásban szkizofréniát. Ez a Kraeplini felosztás (Kraeplini dichotómia) uralta a következő száz évben a pszichiátriai elméletet.

A 19. század vége, 20. század eleje, Freud pszichoanalitikus elméleteivel egy új, pszichológiai megközelítést hozta a depresszió elméletének. Freud a Gyász és melankólia című művében a depresszió háttérében a feldolgozatlan gyászt és veszteséget emelte ki oki tényezőként. A belsővé tett konfliktus fogalmával az önkritika és büntudat mögött az elvesztett tárgy vagy személy iránti harag befelé fordulását írta le. Az ilyen formában, a tudattalanban megjelenő érzések, a hangulatazavarok kiváltó okaként kerültek megfogalmazásra.

1952-ben a DSM-I (Diagnostical and Statistical Manual of Mental Disorders) létrejött, a pszichiátriai zavarok átfogó klasszifikációjának megalapozására törekedett. Ez a kiadás még nagyrészt a Kraepelini hagyományokat követve, a szkizofrénia és a paranoid reakciók mellett tárgyalta a melankóliával együtt járó súlyos hangulatazavarokat. A „neurotikus depresszió” elkülönítésre került, ami egy szorongásos állapot másodlagos jelenségeként, pszichés megküzdésként került azonosításra. A DSM-II (1968) a nem pszichotikus szintű depressziót továbbra is a neurózis formakörében tárgyalta, ami a szorongásos zavarokkal így egy csoportosítás alatt maradt a DSM-I-hez hasonlóan. A DSM-III (1980) megjelenése jelentette az úttörő pillanatot a depresszió diagnosztizálásában. Ekkor került először megfogalmazásra a major depressziós zavar fogalma, ami már külön entitásként határozta meg a depressziót, meghatározott tünetegyüttesrel, így a további kutatásokat is egyszerűsítette. A DSM-IV 1994-es megjelenésével tovább finomítottak a tünetek definícióját, újabb altípusok kerültek leírásra a legújabb tudományos kutatások alapján. A 2013-ban bevezetett DSM-5 klasszifikációs rendszerben már több új diagnózis került leírásra, mint pl. a diszruptív hangulatszabályozás zavar vagy a premenstruális diszfóriás zavar. A tartósan fennálló depresszív zavar diagnosztikai kategóriája az új kiadásban már magában foglalta a krónikus major depressziós zavart és a korábbi disztímiát is.

A depresszió jelenkori kutatásai, elméletei a betegség háttérében multikauzális, többdimenziós modellekben gondolkodnak, melyek magukba foglalják a genetikai, epigenetikai, biokémiai, pszichológiai és környezeti tényezők szerepeit.

1.2. A depresszió típusai, tünetei

A jelenleg is használt DSM-5 klasszifikációs rendszer alapján a depressziós epizód a bipoláris és kapcsolódó zavarok, valamint a depresszív zavarok csoportjába sorolódik. A depresszív zavarokon belül új diagnózisként jelent meg a DSM 5. kiadásában a diszruptív hangulatszabályozás zavar. Tünetei a súlyos, ismétlődő hangulati kitörések, amelyek intenzitása vagy hossza nincs arányban a kiváltó helyzettel. A hangulati kitörések átlagosan legalább heti három alkalommal megjelennek és a hangulati kitörések között szinte mindennap, a nap nagy részében folyamatosan irritált vagy dühös a hangulat. Másik új diagnózis a premenstruális diszfóriás zavar, amiben a tünetek fennállnak a menzesz kezdetét megelőző utolsó héten, a menzesz kezdetét követő néhány nap során kezdenek javulni és a menzesz utáni héten minimálisan csökkennek vagy megszűnnek. Tünetei az affektív labilitás, irritabilitás vagy düh, depressziós hangulat, reménytelenség érzés, szorongás, feszültség, felhúzottság érzés. A major depresszív zavar diagnózisához egy kéthetes időszakon belül legalább öt tünetnek kell fennállnia, melyekből legalább az egyik a depressziós hangulat vagy az érdeklődés, öröm elvesztése kell, hogy legyen. További tünetek a testsúly jelentős változása, alvászavar, pszichomotoros agitáció vagy gátoltság, fáradtság, értéktelenség érzés vagy büntudat, csökkent koncentrációs képesség, valamint a halállal kapcsolatos visszatérő gondolatok. A tartósan fennálló depresszív zavar, másnéven disztímia (ami a DSM-IV-ben leírt krónikus major depresszív zavar és disztímiás zavar egységesítése) definíciója szerint a nap legnagyobb részében fennáll, több napon át jelentkezik, mint ahány napon nem és legalább két éven át tart. Tünetei a major depresszív zavar tünetei, melyekből legalább kettőnek kell megjelennie a diagnózis alkotáshoz. (7)

1.3. A depresszió epidemiológiája, társadalmi fontossága

A depresszió az egyik leggyakoribb mentális betegség a világon, ami a lakosság körülbelül 3,8%-át érinti. (8) Ez világviszonylatban több mint 280 millió embert jelent, ami miatt is a depresszió jelentős társadalmi probléma napjainkban, ami kifejezetten negatív hatással van az országok egészségügyi és gazdasági rendszereire. WHO előrejelzések a probléma súlyosbodását vetítik elő, miszerint 2030-ra a depresszió lesz világszerte a vezető betegség, megelőzve a szív-érrendszeri kórképeket is. (9)

A depresszió a világ minden részén és demográfiai csoportjában jelen lévő probléma, de koronként, nemeként és társadalmi rétegenként eltérő a megjelenési aránya. A major depresszió élettartam-prevalenciája 14,6%, 1 éves prevalenciája 5,5%. (10) Magyarországon ezek az adatok 15,1% és 7,3 %. (11) Egy széleskörű európai kutatás alapján a depresszió prevalenciája Európában 6,38%, Magyarországon 7,98%. (12)

Az epidemiológiai vizsgálatok egységes elemzése alapján megfigyelhető, hogy a depresszió nagyobb arányban érinti a nőket, mint a férfiakat. A nők 1,5-2-szer nagyobb valószínűséggel szenvednek depresszióban, mint a férfiak, aminek a hátterében biológiai, pszichológiai és társadalmi tényezők multifaktoriális szerepe sejthető (13). A depresszió komoly kockázati tényezőt jelent az öngyilkosságban, az öngyilkosság kockázata aránya depressziós epizód alatt 15%. (14)

A depresszió gyakoriságának egyik jól körülírt jellemzője a társadalmi-gazdasági státusz (Socioeconomic status, SES). Az alacsony jövedelmű, kevésbé képzett társadalmi rétegekben a depresszió előfordulása nagyobb az olyan rizikótényezők miatt, mint a labilis munka státusz, a rossz lakhatási körülmények és az egészségügyi ellátáshoz való korlátozott hozzáférés. (15) A depresszió gazdaságra gyakorolt hatásai szerteágazóak, hatással vannak az egészségügyi kiadásokra, az iparra, és a szociális szférára is. A depressziós betegek ellátásával egyenes arányban nőnek az egészségügyi kiadások a különböző kezelési formákon (gyógyszerek, kórházi ágyak, pszichoterápia) keresztül. Egy tanulmány szerint például, az Egyesült Államokban a depresszió kezelésének költségei, közelítő becslések szerint meghaladják az évi 200 milliárd dollárt (16). A depresszió mindemelett jelentős közvetett költségekkel is együtt jár a munkabeli termelékenység hanyatlása, kiesése és a gyakori hiányzások miatt. Elértünk arra a pontra, hogy a mai világban már, a Global Burden of Disease Study vizsgálata alapján, a depresszió a gazdasági recesszió egyik vezető okává vált világszerte (17). Egy másik negatív gazdasági aspektusa a depressziónak a munkaképesség csökkenés mellett a leszázalékolások, rokkantsági nyugdíjak megnövekedett mértéke.

1.4. Biomarkerek a depresszió diagnosztizálásában

A depresszió diagnosztizálása jelenleg a DSM-5 által meghatározott tünetek alapján történik, klinikusok exploratív vizsgálata alapján, aminek hátránya az objektív, mérhető eltérések hiánya. Ebből kifolyólag, ez a szubjektív vizsgálati forma hozzájárul a

heterogén diagnosztikai eredményekhez, valamint az eltérő terápiás válaszokhoz. A depresszióval összefüggésbe hozható, mérhető biológiai eltérések, másnéven biomarkerek azonosítása forradalmasíthatja a depresszió elméleteit, diagnosztizálását és kezelését. A depresszió jelenleg legnagyobb figyelem alatt álló biomarker kutatásait molekuláris, képalkotó, genetikai és fiziológiai kategóriákba sorolhatjuk. A cél, hogy ezeknek a biomarkereknek a segítségével olyan szűrő és diagnosztikai eszközöket tudjunk létrehozni, melyek segíthetik a klinikusokat a diagnózis alkotásban, valamint az adekvát terápia kiválasztásában.

A molekuláris biomarker kutatások, főképp a nyálban, vérben, vagy a liquorban keresnek olyan biokémiai eltéréseket, melyek összefüggésben állhatnak a depressziós állapottal. Ilyen jellegű kutatások során például a pro-inflammatorikus citokinek (úm. interleukin-6 (IL-6), a tumor nekrosis faktor-alfa (TNF- α) és a C-reaktív fehérje (CRP)) emelkedett szintjét írták le depressziós betegeknél, különösen terápia rezisztens depresszió esetén. (18) A krónikus gyulladás ilyen formában feltehetőleg a neurotranszmitter-anyagcserét befolyásolva van hatással a hangulatra és a viselkedésre. (19) Egy másik lehetséges molekuláris biomarker, az agyi neurotrofikus faktor (BDNF) kulcsfontosságú szereppel bír a neuroplaszticitás szempontjából. A depresszió neurotrofikus elmélete a BDNF alacsony szintjét vizsgálja a depresszió hátterében, aminek értéke antidepresszáns terápia hatására növekszik, ami alapján a depresszió fennállásának és a terápiára adott válasz előrejelzőjeként szolgálhat. (20) A HPA-tengely (hipotalamusz-hipofízis-mellékvese tengely), ami a szervezet stresszre adott válaszszerének központi eleme, is a depresszió biomarker kutatásainak fő. Az kortizol szint emelkedését és a kortizol szekréciójának megváltozott napi ritmusát is összefüggésbe hozták a depressziós tünetek kialakulásával. Ezekkel a hormonális változásokkal magyarázhatóvá válik a megnövekedett stresszválasz, ami a depresszióban észlelhető, és ezáltal biomarkerként használható a szűrés és a krónikus állapot felmérése céljából. (21)

A biomarker kutatások másik fontos területe az agyi képalkotó leletek elemzésével próbál a depresszióval összefüggésbe hozható eltéréseket azonosítani, az agy strukturális és funkcionális aspektusain keresztül. Strukturális mágneses rezonancia (MRI) képalkotó vizsgálatok, a depresszióban nagy szereppel bíró területeken, mint a hippocampus és az amygdala írtak le morfológiai elváltozásokat. (22) Egy összesítő tanulmány alapján funkcionális MRI (fMRI) vizsgálatokkal a limbikus rendszer (anterior cinguláris kéreg és

az amygdala) aktivitásának növekedése, a frontális agykérgi területek (dorsolaterális prefrontális kéreg) aktivitásának csökkenése észlelhető depresszióban. Kognitív feladatok megoldása során a depressziós alanyoknál többnyire prefrontális aktivitásnövekedést észleltek a vizsgálatok során az egészséges kontrollokhoz viszonyítva. (23)

A genetikai biomarker kutatások a depresszióval összefüggésbe hozható specifikus génvariánsok és epigenetikai változások azonosítására törekednek, de jelentős limitációjuk, a többi biomarker kutatással egyhangúan, hogy önmagukban gyakran nem elegendőek a depresszió előrejelzésében. Ikervizsgálatok alapján a depresszió heritabilitása 37%. (24) Több, a neurotranszmitterek szekréciójában szerepet játszó gén (pl. a szerotonin transzporter gén hosszúság-polimorfizmusa (5-HTTLPR) rövid allél és az agyi neurotrofikus faktor gén (BDNF)) szerepét is kapcsolatba hozták a depresszióval. A szerotonin transzporter gén rövid allélja összefüggésben állhat a stresszt követően megnövekvő depresszióra való vulnerabilitással. (25) A génexpresszió befolyásolása a genetikai állomány változtatása nélkül is bekövetkezhet epigenetikai változásokon keresztül, mint például a DNS-metiláció és a hisztonmódosítás. Megfigyelték a BDNF gén hipermetilációját depresszióban, aminek a hátterében felmerül a korai életkorban történt, nagyfokú stresszre adott válaszreakció (26).

A fiziológiai vizsgálatokon alapuló biomarker kutatások, mint például az ébrenléti és alvási EEG, a szívfrekvencia-variabilitás (HRV) szintén a kutatások tárgyát képezik. A depresszió szubjektíven is az alvás romlásával jár együtt. EEG vizsgálatokkal a lassú hullámú alvás csökkenését, a REM-latencia lerövidülését és a REM fázisok frekvenciájának növekedését írták le. Poliszomnográfia segítségével vizsgálhatóvá válnak az alvási minták, ami segítséget nyújthat a depresszió diagnosztizálásában (27). Egy másik fiziológiai megfigyelés alapján, depressziós betegeknél csökken a HRV, ami a vegetatív idegrendszer szabályozási zavarával áll összefüggésben. Az alacsony HRV több vizsgálatban korrelációt mutat a depressziós tünetek súlyosságával, és így a jövőben a depresszió diagnosztikájának egyik biomarkerévé válhat (28).

A fentebb vázolt ígéretes eredmények ellenére, továbbra is nagy kihívással jár olyan biomarkerek meghatározása, melyek képesek önmagukban és megbízhatóan előrejelezni, vagy diagnosztizálni a depressziót. Az egyéni biológia különbségek, a társbetegségek és a változó környezeti tényezők összetett hatásai megnehezítik a nagy specificitású és

szenzitivitású biomarkerek azonosítását. Emellett a különböző pszichiátriai betegségekben gyakran átfedésbe kerülnek a vizsgálat alatt álló biomarkerek, ami ezáltal korlátozza specifikus diagnosztikai értéküket.

1.5. A vokális csatorna, mint biomarker

A beszéd egy olyan összetett jelenség, ami részeire bontva több akusztikai jellemző együttesét foglalja magába. A beszédképzésben az akaratlagos tevékenységen kívül, olyan nem tudatos változók is részt vesznek, mint pl. az érzelmek, feszültség, hangulat. Mindezek alapján a beszédelemzés egy újabb ígéretes eszköz a depresszió szűrésében és diagnosztizálásában, mivel objektív és nem invazív módon képes felmérni a hangulati státuszt. A korábbi kutatások alapján a depresszió több akusztikai paraméter együttes változásán keresztül befolyásolja a beszédet. (3) A depresszió hangelemzése során legfontosabb hangtani jellemzőket és azok rövid leírását az 1. táblázat foglalja össze.

1. táblázat: A depresszióban változást mutató főbb hangtani paraméterek és rövid leírásuk (28)

Név	Rövid leírás
Alapfrekvencia (F0)	A zöngéshang periódus idejének reciproka, mértékegysége: Hz
Jitter	A zöngés szakaszok periódus idejének kis elemzési ablakában vett átlagos ingadozását értjük, százalékban kifejezve
Shimmer	A zöngés szakaszok periódusonkénti amplitúdóértékeinek kis elemzési ablakában vett átlagos ingadozását értjük, százalékban kifejezve
Formánsfrekvenciák (F1, F2, ...)	Az artikulációs csatorna által felerősödött felhangcsoportok, rezonanciafrekvencia-tartományának burkológörbéjének maximum amplitúdóhelye
Formánsfrekvenciák sávszélessége (B1, B2, ...)	A jelfeldolgozáson belüli általános sávszélesség, vagyis az alsó és felső határfrekvencia tartománya, mértékegysége: Hz
Hangintenzitás	Egységnyi felületen merőlegesen átváramlott hangenergia mennyisége adott idő alatt

Intenzitásszint	Az aktuális hangintenzitás és egy adott referencia-hangintenzitás decibelskálán történő ábrázolása
Mel-sávós energiaértékek	A beszédjel sávszűrésével kiszámolható értékek, melyeket adott frekvenciaértéktől, adott frekvenciatartományokkal szokás elvégezni, mértékegysége: dB
Mel-frekvencia cepstrális együtthatója (MFCC)	Kiszámított mel-sávós energiaértékeken végrehajtott diszkrét koszinusz transzformációval kapott együttható
Artikulációs sebesség	Egy másodperc alatt kiejtett beszédhangok száma, mértékegysége: hang/s
Relatív szünethossz	A vizsgált beszédszakaszban található szünetek teljes hosszának aránya, a vizsgált beszédszakaszhoz viszonyítva, százalékban kifejezve
Tranziensarány	A vizsgált beszédszakaszban található változó szakaszok teljes hosszának aránya, a vizsgált beszéd szakasz teljes hosszához viszonyítva, százalékban kifejezve

A beszéd elemzésével a cél olyan biomarkerek azonosítása depressziós betegeknél, melyek a későbbiekben diagnosztikai eszközként szolgálhatnak. Egy beszéd elemzésen alapuló akusztikai rendszer megalkotása gyorsíthatja a szűrést, és javíthatja a diagnosztika pontosságát.

A depresszió beszédmintázatra gyakorolt hatása már évtizedek óta a kutatások tárgyát képezik. 1921-ben Emil Kraepelin leírása szerint, a depressziós betegek beszéde halk, lassú, monoton, időnként akadozásokkal teli, suttogó. (29) A kezdeti kutatások, a technológia hiányában csak a hangtani paraméterek kvalitatív elemzésére voltak alkalmasak. Newman és Mather a depressziós betegek beszédét élettelennek, közömbös minőségűnek írták le, ami szűk hangmagasság tartománnyal és redukált hangsúlyozással jár együtt. (30) Moses a depressziós hangot monotonként, egyneműként definiálta. (31) Darby és Hollien a depressziós betegek beszédét tompának, élettelennek írták le, amiben javulás volt észlelhető adekvát terápiát követően, főképp az artikuláció, hangmagasság tartomány és hang intenzitás terén (32). Ezt később Askenfelt and Sjolín is igazolta (33).

A 80-as években történt kutatások során, a depressziós betegek beszédében legfőbb változásokként a csökkent beszédsebesség és a megnőtt szünethossz volt mérhető, ami összefüggésben állhat a pszichomotoros retardációval. (34) Később az antidepresszáns terápia megjelenésével, a társuló klinikai állapotjavulás, a beszéd akusztikai paraméterekkel leírt modellezésével is társult. (35) Az 1990-es években a fonetikai elemzés fejlődése lehetővé tette a kutatók számára, hogy a beszéd akusztikai jellemzőit átfogóbban elemezzék. Ezáltal számszerűsíthetővé váltak olyan paraméterek, mint a hangmagasság, a hangintenzitás és az olyan idő dimenzióbeli változók is, mint a beszédsebesség és a szünetek hossza. Ezen jellemzők vizsgálatával olyan korai automatizált rendszerek váltak elérhetővé, melyek már képesek voltak az akusztikai eltérések elemzésére. Az ebben az időszakban végzett fonetikai vizsgálatok megállapításai szerint, a depressziós betegeknél általában alacsonyabb alapfrekvencia (hangmagasság), hosszabb szünethossz és lassabb beszédsebesség mérhető az egészségesekhez képest (36). Ezek a vizsgálatok úttörőként szolgáltak a depresszió hangtani elemzésében, mert képesek voltak számszerűsíthető adatokkal alátámasztani az évtizedek óta fennálló klinikai megfigyeléseket. A 90-es évek jelentették a kezdetét annak az új tudományos hullámnak, aminek hatására előtérbe került a beszéd hangtani elemzése a depresszió szűrése, nyomonkövetése céljából. A következő korszakot a mesterséges intelligencia megjelenése jelentette, amit a következő fejezetekben tárgyalok részletesebben.

1.6. A mesterséges intelligencia típusai

A mesterséges intelligenciát felhasználhatósága, illetve hatékonysága alapján három fő típusba sorolhatjuk. Az első típus a mesterséges keskeny intelligencia (Artificial Narrow Intelligence, ANI, más néven gyenge MI), ami célorientált módon képes elvégezni specifikus feladatokat, de nem tud az előre beállított paramétereken kívül dolgozni. Előnyei a hatékonyság, a gyorsaság és a monotonitástűrés. Az ANI legegyszerűbb formája a reaktív gép, ami memória hiányában nem képes még a korábbi műveleteket felidézni és ezáltal felhasználni azokat más feladatokhoz. A következő lépcsőfok a korlátozott memóriával rendelkező szint, ahol a rendszer már képes a tanulásra, ezáltal a megelőző tapasztalatok segítségével optimalizálja saját működését. A jelenleg is használt, mesterséges intelligenciát alkalmazó rendszerek nagy része ebbe a kategóriába tartozik.

A második típus a mesterséges általános intelligencia (Artificial General Intelligence, AGI, vagy erős MI), amiben már olyan intellektuális képességekkel rendelkezik a rendszer, melyek megközelítik az emberi gondolkodást és viselkedést, ezáltal képessé válik arra, hogy önállóan végrehajtsa komplex feladatokat. Mindemellett az AGI már képes a meghatározott feladatokat túlhaladni, a megszerzett tapasztalatait más feladatok kontextusaiban is alkalmazni. Mindez képessé teszi az innovatív és kreatív problémamegoldásra is. Jelenleg a mesterséges intelligenciának ezen szintje még nem elérhető, valamint az ezzel együttjáró szabályozási és etikai háttér sem került kidolgozásra. Hipotetikusán az AGI első fejlődési szakasza az ún. tudatelméleti szint, ami képes felismerni az érzelmeket és a tudatállapotokat, valamint reagálni azokra, ezáltal bizonyos fokú én-integritásra tesz szert. Az AGI legmagasabb szintje az öntudatra ébredés.

A harmadik típus a mesterséges szuperintelligencia (Artificial Superintelligence, ASI), amilyen szinten a gép, már az emberi intellektust már meghaladó funkciókkal rendelkezik. Az ASI szintjét jelenleg csupán a tudományos-fantasztikus irodalom tárgyalja, olyan entitásként, ami a saját cselekedeteinek tudatosításával képessé válik exponenciálisan fejleszteni és tökéletesíteni önmagát. (37)

1.7. A mesterséges intelligencia főbb formái

Gépi tanulás (machine learning)

A gépi tanulás (ML) a mesterséges intelligencián alapuló alkalmazásokon belül az egyik leggyakrabban használt eljárás, aminek segítségével a gépek elemezni képesek adathalmazokat és iteratív tanulási folyamatokon keresztül javítják a saját teljesítményüket. A konkrét utasításokon alapuló konvencionális programozással ellentétben, a gépi tanulás segítségével ezek a rendszerek képesek a betáplált adatokból tanulni és önállóan alkalmazkodni az új adatok elemzése érdekében.

A gépi tanulás célja, hogy mintákat ismerjenek fel adathalmazokban és következtetéseket tudjanak levonni az adatokról mindennemű programozás nélkül. A gépi tanuló eljárások, a hagyományos statisztikai módszereken alapulnak, azonban az idegtudományok és az emberi tanuláselméleti modellek is szerepet játszottak a megalkotásukban.

A gépi tanulásnak több típusa különíthető el, attól függően, hogy milyen adatokat használunk a rendszer betanítása során.

a) Felügyelt tanulás (Supervised learning)

A felügyelt tanulás a leggyakrabban használt gépi tanulási eljárás, aminek során előre címkézett adatokkal történik a rendszer betanítása. Ezzel a módszerrel a modellünk képes megtanulni a bemeneti és kimeneti adatok közötti összefüggéseket, aminek segítségével osztályozásra és a regresszióra lesz képes. Osztályozással diszkrét címkék predikciója, regresszióval folytonos érték előrejelzése válik lehetségessé. A betáplált adatokból való tanulás révén folyamatosan finomítható a rendszer, ezáltal a teljesítményének pontossága javítható.

Ebbe a csoportba tartozik az ún. Support Vector Machine (SVM). Ennek az eljárásnak a célja, hogy az adathalmazon belül feltárható lehessen az a vonal vagy „hipersík”, ami a legjobban elválasztja a különböző adatcsoportokat, így a lehető legszélesebb rés (vagy „margó”) válassza el a két adathalmazt, annak érdekében, hogy az osztályozás minél pontosabbá váljon. A „támogató vektorok”, az elválasztó vonalhoz legközelebb eső adatpontok segítik annak helyzetének meghatározását. A legtöbb mintában az adatokat nem lehet egy egyenes vonallal elválasztani, ilyen esetekben alkalmazzuk az ún. „kernel trükköt”, aminek segítségével összetett, több dimenziós határok hozhatók létre, nagy dimenziós adatok osztályozása céljából.

A felügyelt tanulás eszköze lehet az ún. decision tree (döntési fa), ami a választási lehetőséget és azok eredményeit fa formájában ábrázolja. A Random Forest (RF) eljárás során az adathalmaz véletlenszerűen kiválasztott részeiből döntési fákat hozunk létre, és az eredmények kombinálásával növeljük a modell pontosságát.

A bootstrapping (másnéven bagging, magyarul zsákolás) módszert a gépi tanulások algoritmusok pontosságának és stabilitásának növelésére használják, klasszifikáció és regresszió során.

b) Felügyelet nélküli tanulás (Unsupervised learning)

A felügyelet nélküli tanulás során a modell képzéséhez használt adatok nincsenek előre címkézett válaszokkal ellátva. Ez a megközelítés képes az adathalmazban a tendenciák és anomáliák azonosítására. Ezeket különösen a klaszterezési és asszociációs feladatoknál használják. A módszer nem csak elemzi az adatokat és mintákat talál, hanem potenciálisan hasznos jellemzőket is talál egy jövőbeli felügyelt tanulási feladathoz.

c) Félig felügyelt tanítás (Semi-supervised learning)

A félig felügyelt tanulást főként olyan feladatokban használják, ahol egyszerű hozzáférésünk van a jelöletlen adatokhoz, de nehéz hozzájutni a címkézett adatokhoz. Ez a technika lehetővé teszi, hogy a modell a címkézett adatokon keresztül képes legyen megtanulni az asszociációkat, majd ezt követően általánosítani tudjon a címkézetlen adatokon. A módszer olyan területeken alkalmazható, mint például a kép- és beszédfelismerés, ahol nehézségekbe ütközhet a címkézett adatok beszerzése.

d) Megerősítéses tanulás (Reinforcement learning)

A megerősítéses tanulás során a rendszer részét képező ágens a környezet felfedezésével tanul, amely során jutalmakon és büntetéseken keresztül visszajelzést kap. A próbálkozások és hibázások segítségével az ágens célja a kumulatív jutalmak maximalizálása az idő múlásával. A megerősítéses tanulási algoritmust elsősorban a robotikában, a játékfejlesztésben és az önvezető autókban alkalmazzák, amikor dinamikus környezetben kell cselekedeteket végezni, konkrét célok elérése érdekében.

(38)

Neurális hálózatok és mélytanulás

A mélytanulás egy olyan megközelítés, amely egy összetett probléma megoldására szolgál az idegtudományok ismeretein alapuló algoritmusok alkalmazásával. Ezek az úgynevezett neurális hálózatok úgy működnek, hogy az információt csomópontok (vagy neuronok) rétegein keresztül továbbítják, így mintákat tanulnak, és előrejelzéseket készítenek. A mélytanulás különösen hatékony az adatokban lévő minták felismerésében, például a képfelismerésben, a beszédfeldolgozásban és a nyelvi fordításban, ami azt jelenti, hogy olyan alkalmazások kulcsfontosságú részét képezi, mint az önvezető autók, a virtuális asszisztensek és az orvosi diagnosztika.

A neurális hálózat olyan algoritmusok sorozata, amelyek az emberi agy működését utánozva próbálják felismerni a mögöttes összefüggéseket egy adathalmazban. A neurális hálózat három alapvető komponensből áll: egy bemeneti rétegből, amely feldolgozatlan nyers adatokat (pl. képpixelet vagy szöveget) fogad, rejtett rétegekből, amelyek az adatokat feldolgozzák és átalakítják, egyre komplexebb jellemzők kinyerésével, és egy kimeneti rétegből, amely a végeredményt állítja elő, pl. azonosítja a képen lévő tárgyat

vagy osztályozza a szöveget. Minden csomópont, a hálózat más csomópontjához kapcsolódik, és ezek a kapcsolatok „súlyokkal” rendelkeznek, amelyek meghatározzák, hogy az adatok hogyan alakulnak át a hálózaton keresztül. Ezeket a súlyokat a tanulás során szabályozzák, hogy a hibák minimalizálhatóak legyenek az előrejelzésekben.

Itt két folyamatot különböztetünk meg, amelyekkel a neurális hálózatok tanulnak: a lejátszást (forward propagation) és a visszajátszást (backpropagation). A lejátszás során az adatokat rétegről rétegre áramoltatják a hálózaton keresztül, hogy egy kimenetet állítsanak elő. A visszajátszás lehetővé teszi a hálózat számára a tanulást azáltal, hogy meghatározza a hibát a megjósolt kimenet és a tényleges eredmény között, és ennek megfelelően módosítja a kapcsolati súlyokat, hogy csökkentse ezt a hibát. Ez az iteratív funkciója segít a hálózatnak abban, hogy idővel javuljon a teljesítménye.

A konvolúciós neurális hálózatok megalkotásánál, egy matematikai művelettel az adathalmaz egyes részeinek reprezentációja céljából súlykészletet hozunk létre. Ez a súlykészlet vagy szűrő kisebb, mint az adathalmaz egésze, annak áthelyezésével és a folyamat ismétlésével a teljes adathalmaz lefedése a cél. Ezt az eljárást főleg kép, hang és videóelemzésekben használják.

A mélytanulás egy olyan kifejezés, amelyet a több réteggel rendelkező neurális hálózatokra használnak, ezáltal mély neurális hálózatoknak is nevezik. Minél több réteg van a hálózatban, annál nagyobb a teljesítménye. A mély hálózatok különösen hasznosak a sok adatot vagy összetett mintákat tartalmazó feladatoknál, például a kép- és beszédfelismerésnél. (39)

Természetes nyelvi feldolgozás

A természetes nyelvfeldolgozás (Natural Language Processing, NLP) a mesterséges intelligenciának egy alkalmazása arra a célra, hogy egy számítógépes rendszer képes legyen megérteni, értelmezni és előállítani az emberi nyelvet. Az NLP lehetővé teszi a számítógépek számára, hogy a szövegből, beszédből és más emberi kommunikációból felismeréseket és jelentést nyerjenek. A technológiával olyan mindennapi technológiáinkban is találkozhatunk, mint például a különböző hangalapú asszisztensek, pl. Siri és Alexa, vagy az olyan nyelvi fordítóban, mint a Google fordító.

Az NLP-n belül többféle feladat is létezik. A szövegosztályozás, ahol a gépeket arra képezzük ki, hogy a szöveget különböző csoportokba sorolják, talán az egyik

legelterjedtebb feladat. Ezt használják például a spam-felismerésnél, az érzelmi osztályozásnál (annak ellenőrzése, hogy egy szöveg pozitív vagy negatív). A nevesített entitások felismerése (named entity recognition, NER) egy másik gyakori feladat, amikor a számítógép emberek, helyek vagy szervezetek neveit azonosítja a szövegben. Egy másik feladat a beszédrészek címkézése, ahol az NLP-eszközök azonosítják az egyes szavak funkcióját a mondatban (például, hogy főnév, ige, melléknév stb.).

A nyelvi modellezés az NLP egyik kulcseleme, annak megjósolása, hogy az aktuális kontextus alapján milyen szavak vagy szósorozatok következnek. Például egy üzenet megírásakor a nyelvi modellek segítségével generált prediktív szövegjavaslatokat kapunk. A transzformátorok olyan fejlettebb modellek, melyek célja, hogy megértsék a használt szavak kontextusát, a szavak közötti kapcsolatok alapján. Ez lehetővé teszi, hogy ezek a típusú modellek megértsék a bonyolult nyelvi struktúrákat, és emberhez hasonló válaszokat adjanak (40)

1.8. A mesterséges intelligencia alkalmazása a pszichiátriában

A mesterséges intelligencia számos területen, köztük a pszichiátriában is meghatározó szereppel bír. Egy ilyen technológiai integráció átalakíthatja a diagnosztikai és kezelési stratégiákat az adatvezérelt megközelítés, a prediktív analitika és a személyre szabott ellátási modellek segítségével.

A mesterséges intelligencia egyik legjelentősebb és legközvetlenebb felhasználási területe a pszichiátriában a korai diagnózisalkotás és a szűrés. Az eddig alkalmazott hagyományos diagnosztikai módszerek a betegek által jelzett tünetek szubjektív értékelésén alapulnak, ami hajlamos az elfogultságra. Ezzel szemben a mesterséges intelligencia algoritmusai mindenféle adathalmazt - elektronikus egészségügyi nyilvántartások és képalkotó vizsgálatok adatai, sőt még a közösségi média aktivitását is - képesek elemezni, hogy a pszichiátriai állapotok finom jelzőit megtalálják.

A téma elemzésében a korábban társszerzőként írt összefoglalómat használom alapul. (37)

A mesterséges intelligencia és a gépi tanulás jelentős előrelépést hozhat a mentális betegségek diagnosztikájában, különösen az orvosi döntéstámogató rendszerek fejlesztésében. Az MI képes az egyéni és családi kórtörténet, valamint a viselkedési

mintázatok elemzésével prediktív értékű mintázatokot azonosítani, amelyek segíthetnek a betegség lefolyásának előrejelzésében és az egyénre szabott kezelések kialakításában.

Az MI gyorsíthatja a diagnosztikai folyamatokat, miközben csökkenti a humánerőforrás és gazdasági terheket. Abd-Alrazaq és munkatársai szerint az MI 21–100%-os találati aránnyal diagnosztizálhat mentális zavarokat, mint például Alzheimer-kór, szkizofrénia, bipoláris zavar és PTSD. (41) Egy másik tanulmányban Tutun és munkatársai egy mindössze 28 kérdésből álló tesztbatteríát alkalmazó MI-rendszert hoztak létre, ami 89%-os pontossággal diagnosztizált mentális zavarokat emberi beavatkozás nélkül. (42)

Az MI-alapú rendszerek új távlatokat nyitnak a gyorsabb, precízebb diagnosztika és az egyénre szabott kezelések területén, miközben jelentős erőforrás-megtakarítást kínálnak. A szuicid rizikó becslése a mentális egészségügyi kutatások kiemelt területe, mivel hatékony eszközt nyújthat az öngyilkossági ráták csökkentéséhez. Passos és munkatársai gépi tanulással 72%-os pontossággal különítették el a szuicid kísérletet elkövetőket a nem elkövetőktől, különböző klinikai jellemzők alapján, úm. a depressziós állapot, pszichiátriai diagnózisok és szerhasználat az anamnézisben. (43) De Ávila Berni és kollégái Virginia Woolf naplóbejegyzéseinek és leveleinek szövegelemzése révén hoztak létre egy gépi tanuláson alapuló modellt, amely 80%-os pontossággal jelezte előre a szuicid magatartást. (44) Pestian és munkatársai a beszéd akusztikai és nyelvi jellemzőinek elemzésével 85%-os pontosságot értek el a szuicid páciensek azonosításában. (45)

A közösségi média is ígéretes, neurális hálók segítségével például Twitter-bejegyzések alapján felismerhetők a szuicid kijelentések, megkönnyítve a korai előrejelzést és a megelőzést. (46)

A mesterséges intelligencia alapú eszközök az agyi képalkotó leletek nagyszámú adatbázisokon történő elemzését is lehetővé teszik, amely a szkizofrénia előrejelzésére és felmérésére irányuló kutatások alapjául szolgál. Deep learning algoritmusok alkalmazásával agyi MR-felvételek, genomikai adatok és EEG-leletek elemzése révén meggyőző eredményeket értek el a szkizofrénia klasszifikációja és prognózisa terén, ami új távlatokat nyit a betegség diagnosztikájában és kezelésében. (47)

A mesterséges intelligencia és a gépi tanulás jelentős előrelépéseket hozhat a molekuláris biomarkerek fenotipizálásával, ami forradalmasíthatja a személyre szabott gyógyászatot. Ezek az eszközök lehetővé teszik a gyógyszeres kezelések pontosabb tervezhetőségét,

javítva a hatékonyságot, a terápiás válaszok előrejelzését és a mellékhatásprofil optimalizálását, ezáltal növelve a terápiahűséget. (48)

A pszichofarmakonok dozírozása és hatékonyságának előrejelzése továbbra is komoly kihívást jelent a klinikusok számára. Sheu és munkatársai 17 556 fős mintán vizsgálták, hogy elektronikus egészségügyi adatok felhasználásával négy antidepresszáns osztály (SSRI, SNRI, bupropion, mirtazapin) esetében a terápiás válasz 4-12 hét elteltével milyen pontossággal jelezhető előre. (49) Az eredmények azt mutatták, hogy a gépi tanuló modellek hatékonyan csökkenthetik az ineffektív kezelések arányát. A szűk terápiás indexű gyógyszerek, mint például a lítium, különösen alkalmasak az AI-alapú adatelemzés alkalmazására. A lítium esetében, amelynek szűk terápiás tartománya miatt rendszeres vérszint-monitorozás szükséges, Hsu és munkatársai gépi tanulási modellekkel olyan rendszert fejlesztettek, amely képes előre jelezni a lítium koncentrációját komplex klinikai adatok alapján. (50) Ezek az innovációk jelentősen hozzájárulhatnak a pszichotróp gyógyszerek pontosabb adagolásához és a betegek biztonságának növeléséhez.

1.9. A mesterséges intelligencia alkalmazása depresszióban

A mesterséges intelligencia különböző forrásokból származó nagy mennyiségű adatot képes elemezni, beleértve a beszédmintákat, az arc kifejezés elemzését, viselkedési adatokat és fiziológiai jeleket. A beszéd elemzése különösen fontos területnek bizonyult a depresszió felismerésében. A depresszióban szenvedő emberek beszédében például olyan jellegzetes vonások figyelhetők meg, mint a lassabb beszédtempó, a hangmagasság csökkent változékonysága és a gyakoribb szünetek. Több kutatás a beszélt nyelvi mintákból képzett akusztikai és nyelvi jellemzők alapján, gépi tanulási eljárások segítségével hatékonyan és nagy pontossággal azonosította a depressziós tüneteket (51). A természetes nyelvi feldolgozás eszközei képesek azonosítani a depresszió olyan jeleinek jelenlétét, mint a negatív érzelmek, az önreferenciális nyelvezet és a kisebb lexikális változatosság (52). Ezek az eszközök nemcsak a klinikai, hanem a szubklinikai stádiumban is segítenek a depresszió felismerésében, így megkönnyíthetik az időben történő beavatkozást.

A depresszióhoz társuló mikrokifejezéseket és viselkedésmintákat mélytanulósos arcfelismerő technológiák segítségével is lehet elemezni, ami tovább növeli a

diagnosztikai pontosságot. Az arckifejezések közötti apró változások (pl. kevésbé intenzív mosoly vagy a semleges arc hosszabb ideig való megtartása) az emberi megfigyelő számára észrevétlenek, de a mesterséges intelligencia rendszerek segítségével könnyen és nagy pontossággal felismerhetők (53). Amennyiben a diagnosztika multimodálissá válik, és az arcelemzést más adatmodalitással, például hanggal és szöveggel kombinálja, a mesterséges intelligencia rendszerek megbízhatósága tovább javulhat.

Hasonlóképpen ígéretes eszköz lehet az MI a depresszió diagnosztizálásában a mobil egészségügyi alkalmazások és a viselhető technológiák segítségével. Az ilyen típusú viselhető eszközök folyamatosan mérik a viselkedést (pl. alvás és aktivitási idő) és a fiziológiát (pl. szívritmus-variabilitás), ezáltal értékes adatbázisra tesznek szert a mindennapokban. Ezeket az adatokat mesterséges intelligencia-algoritmusok elemezhetik, hogy felismerjék a depressziós epizódot tükröző rendellenességeket. A csökkent aktivitási szint és a megzavart cirkadián ritmus mind az akut, mind a krónikus depresszió tipikus jellemzői, amelyek a viselhető szenzoros eszközökkel könnyen vizsgálhatók (54). Ezek az eszközök kritikus szerepet játszhatnak a preventív mentális egészségügyi ellátásban azáltal, hogy valós idejű adatokat szolgáltatnak, és mind a betegeket, mind a klinikusokat figyelmeztetik a korai figyelmeztető jelekre.

Egyéb szöveges tartalmak, például a Twitteren írt bejegyzések érzelmi mintáinak elemzése, szintén előre jelezheti a depressziót. (55) A Facebookról és más forrásokból származó közösségi média bejegyzések érzelmi és nyelvi elemzését is felhasználják a depressziós állapotok felismerésére. (56)

A mesterséges intelligencia a kezelés kiválasztása és a válaszreakció nyomon követése során is előrejelző funkcióval rendelkezhet. A depresszió heterogén természetű, ami hangsúlyozza a kezelés személyre szabottságának fontosságát. Klinikai vizsgálatok és elektronikus egészségügyi nyilvántartások adataiból tanított MI rendszerek képesek lehetnek előrejelzést nyújtani arról, hogy a beteg profilja alapján melyik kezelés (gyógyszeres vagy pszichoterápia) lesz a legmegfelelőbb (57). Ezenkívül a longitudinális adatok nyomon követése lehetővé teszi, hogy az MI-rendszerek figyelemmel kísérhessék a betegek kezelésének előrehaladását, azonosítsák a visszaeséseket, ezáltal a terápiás beavatkozás módosítását ajánlják, annak érdekében, hogy biztosítsák az ellátás folyamatosságát.

1.10. Depresszió és beszédelemzés mesterséges intelligenciával

A mesterséges intelligencia, különösen a gépi tanulás és a természetes nyelvi feldolgozás potenciálisan nagy hatékonysággal képes elemezni a beszédet a depressziós állapotok felismerése, osztályozása és nyomon követése céljából. Az MI-alapú eszközök a beszéd akusztikai, nyelvi és paralingvisztikai aspektusait vizsgálva skálázható és objektív megoldást kínálhatnak a depresszió diagnosztizálására és kezelésére.

A beszéd olyan aspektusai, mint például a hangmagasság, az intenzitás és az idő dimenzióbeli minták értékes jelzések lehetnek az egyének érzelmi és mentális állapotáról. Ezek a jellemzők a kognitív és érzelmi funkciókban bekövetkezett változásokon keresztül depresszióban gyakran megváltoznak. Az ilyen típusú akusztikai tulajdonságok gépi tanulási eljárásokkal elemezhetőek, ezáltal a depressziós állapotok objektív osztályozása és előrejelzése érhető el.

A depresszióban szenvedő egyéneknél megfigyelhető a monoton, csökkent hangmagasság-variabilitással és alacsonyabb intenzitással jellemezhető beszéd. Low és munkatársai kutatása alapján, az olyan akusztikai jellemzőkkel tanított ML-algoritmusok, mint az alapfrekvencia (F0) és annak változása, hatékonyan meg tudják különböztetni a depressziós egyént a nem depressziós egyéntől (3).

A beszéd időbeli jellemzői, mint például a beszédsebesség, a szünetek időtartama és a beszédritmus, a depresszió kulcsfontosságú mutatói. A depressziós egyének lassabban beszélnek és hosszabb szünetet tartanak a válaszaikban, ami a pszichomotoros retardáció jellemzői. (3)

A hangszalagok viselkedéséről és az artikulációról további információkat szolgáltatnak a spektrális jellemzők, például a formánsfrekvenciák. A beszéd rezonanciajellemzőinek definiálásán keresztül, a formánsfrekvenciák alapvető fontosságúak. Scherer és munkatársai a spektrális jellemzőket prozódikus adatokkal kombinálták, majd SVM-ek segítségével több mint 85%-os pontosságot értek el a depresszió felismerésében a beszédmintákból. (58) A spektrális jellemzők időbeli és prozódiai adatokkal való kombinálása bizonyítottan robusztusabbá teszi az ML-modelleket.

Az akusztikai jellemzők önállóan is nagy prediktív képességgel rendelkeznek, azonban a kombinációjukkal általában jobb eredmények érhetőek el. Cummins és munkatársai kimutatták, hogy a prozódikus, időbeli és spektrális jellemzők együttes elemzése a depressziós állapotok pontosabb osztályozásához vezet. Kutatásukban Random Forest

osztályozókat alkalmaztak és több jellemzőkészletet integráltak, így 86%-os osztályozási pontosságot értek el.áltak több jellemzőkészlet integrálására, és 86%-os osztályozási pontosságot értek el. (51) A vizsgálat ezáltal a depresszió beszédelemzése szempontjából a globális megközelítések előnyeit hangsúlyozza.

Az ML alkalmazása az akusztikus beszédelemzésben azonban a hatalmas előrelépések ellenére még mindig kihívást jelentő feladat. A modell teljesítménye nagymértékben függhet a felvételi körülményektől, az akcentusoktól és a nyelv változékonyságától. Mindezek mellett az érzékeny beszédadatok adatvédelmét és biztonságát is garantálni kell. A további kutatások fontos feladata az adatgyűjtés módjának szabványosítása, valamint általánosabb, a demográfiai és nyelvi sokféleséget figyelembe vevő modelleket kidolgozása.

1.11. A mesterséges intelligencia etikai vonatkozásai

Bizonyított, hogy a mesterséges intelligencia képes növelni a diagnosztikai pontosságot, a kezelés testre szabhatóságát és a működési hatékonyságot az orvostudományban. Használata azonban alapvető etikai kihívásokat is felvet, amelyeket körültekintően kell megvizsgálni. Ezek a megfontolások kiterjednek az adatvédelemre, az elfoglaltságra, az átláthatóságra és az elszámoltathatóságra, valamint az orvos-beteg kapcsolatra gyakorolt lehetséges következményekre is. Ezeknek az aggályoknak a kezelése szükséges ahhoz, hogy a mesterséges intelligenciát az orvostudományban felelősségteljesen és méltányosan lehessen alkalmazni.

Az egészségügyi mesterséges intelligencia számos olyan nagy adathalmazt használ, amely érzékeny betegadatokat, például kórtörténetet, genetikai és képalkotó eredményeket tartalmaz. Az ilyen adatok védelme valóban az egyik legfontosabb etikai megfontolás. Az adatok megsértésének nagyon súlyos következményei is lehetnek, mint például a személyazonosság ellopása vagy a személyes egészségügyi információkkal való visszaélés. E kockázatok minimalizálása érdekében fontos a hatékony titkosítási szabványok használata mellett az olyan adatvédelmi rendeletek szigorú betartása, mint pl. az európai általános adatvédelmi rendelet (GDPR). (59) Ugyanakkor kihívások merülnek fel azzal kapcsolatban is, hogy megtörténik-e a betegek beleegyezése az adataik rögzítésébe. Sok mesterséges intelligenciamodellt olyan adatokon képeznek ki, amelyeket a betegektől a kifejezett beleegyezésük nélkül gyűjtöttek. Ez kérdéseket vet fel

a területen az autonómiával és a megfelelő tájékoztatáson alapuló döntéshozatallal kapcsolatban. Annak érdekében, hogy a betegek bizalmát megőrizzük a rendszerben, átlátható iránymutatásokat kell kidolgozni az adatok felhasználhatóságával kapcsolatban. Egy másik nagy etikai kihívás a mesterséges intelligencia rendszerek elfogultsága. Ha az algoritmusokat elfogult adathalmazokon képzik ki, akkor azok végül megismétlik vagy súlyosbítják a jelenlegi egészségügyi egyenlőtlenségeket. Kutatások kimutatták, hogy az elsősorban egy demográfiai csoport adatain kiképzett mesterséges intelligenciamodellek gyakran rosszabbul teljesítenek, amikor más csoportokra alkalmazzák őket, ami egyenlőtlen eredményekhez vezethet. (60) A torzítások elkerülése érdekében biztosítani kell a sokszínű és reprezentatív adathalmazokat, valamint a modellek alapos tesztelését a különböző demográfiai csoportokon. Ezen túlmenően az algoritmusok folyamatos nyomon követése felfedheti és korrigálhatja az olyan torzításokat, amelyek a használat előtt nem voltak jelen vagy nem voltak előre jelezve.

A mesterséges intelligencia algoritmusok nagy része úgy képzelhető el, mint egy „fekete doboz”, ami megnehezíti az átláthatóságát és a megmagyarázhatóságát. A mesterséges intelligencia fekete doboz jellege miatt az orvosok és a betegek nehezen tudnak megbízni a kimenetekben, illetve nehezen tudják azokat validálni, mivel nem átlátható, hogy a mesterséges intelligencia modell hogyan jutott el egy adott ajánláshoz vagy döntéshez. A megalapozott döntéshozatal és az elszámoltathatóság is sérülhet az átláthatóság nélküli, ami főképp magas kockázatú környezetben lehet veszélyes. A megmagyarázható mesterséges intelligencia (explainable AI, XAI) kezdeményezések ezt a problémát próbálják megoldani azáltal, hogy átláthatóbbá teszik a modellek működését és döntéseit. Például a gépi tanulási modellek vagy a döntési fák, könnyebben értelmezhetőek, bár ez alacsonyabb pontossággal járhat, mint pl. a komplex mély tanulási modellek. (61) A mesterséges intelligencia orvoslásban való etikus alkalmazásának egyik fő nehezítő tényezője így továbbra is a teljesítmény és az értelmezhetőség közötti folytonos egyensúlyozás.

A mesterséges intelligencia által okozott műhibákért való felelősség megállapítása továbbra is etikai és jogi fejtörést okoz. Ha például egy mesterséges intelligenciával működő rendszer téves diagnózist vagy kezelési javaslatot javasol, ki a hibás, a rendszer fejlesztői, a rendszert használó klinikusok, mivel végső soron ők hozzák meg az orvosi döntéseket a mesterséges intelligencia ajánlásai alapján, vagy az egész rendszert használó

intézmény? Ez a jelenlegi jogi keretek között nem mindig egyértelmű, így a műhibaperek nehezen megoldhatók nem egyértelműek ebben a kérdésben, ami megnehezíti a műhibaperek megoldását. (62)

Ahogy a mesterséges intelligencia egyre nagyobb szerepet játszik az egészségügyben, aggodalmak merülnek fel a hagyományos orvos-beteg kapcsolatra gyakorolt hatásával kapcsolatban. Bár az MI növelheti a hatékonyságot és a diagnózis pontosságát, az ilyen technológiától való túlzott függés alááshatja a személyes orvosi találkozások fontos elemeit, mint pl. az empatikus és személyes ellátást. Az MI-vezérelt eszközökkel való interakciót a betegek személytelennek tekinthetik, ami a bizalom és az elégedettség csökkenéséhez vezethet. (63) E kockázatok mérséklése érdekében a mesterséges intelligenciát inkább az emberi ítélőképességet segítő, nem pedig azt helyettesítő eszközként kellene definiálnunk. A klinikusoknak szóló MI-képzési programoknak a technológia támogató jellegére kell összpontosítaniuk, hogy fenntartsák az orvoslás szerves részét képező betegközpontúságot.

2. Célkitűzések

2.1. A depresszióban végzett gépi tanuló eljárással támogatott hangelemzési módszerek kutatási gyakorlatának és eredményeinek áttekintése

A depresszió szűrésének egyre jobban kutatott területe a beszéd, mint biomarker vizsgálata gépi tanuló eljárások segítségével. A különböző kutatások eltérő módszertana és vizsgálati eszközei megnehezítik a standardizálását az eljárásnak. Umbrella review vizsgálatunk során áttekintettük az aktuális kutatási gyakorlatot és eredményeket annak érdekében, hogy a következő kérdésfelvetéseinkre választ kaphassunk:

- Mik az aktuális gyakorlatok a depresszió gépi tanuló eljáráson alapuló hangelemzéses módszereiben napjainkban?
- Milyen átfogó következtetlenségek nehezítik a téma kutatását?
- Milyen konstruktív javaslatokkal lehetne élni a jövőbeli kutatások egységesítésének javítása céljából?

A téma átfogó elemzésén és a leggyakrabban felmerülő következtetlenségek felismerésén keresztül, a jelenleg használt módszerek standardizálásához szükséges javaslatok fogalmazhatóak meg.

Jelen áttekintő tanulmányunkat a második vizsgálatunkhoz képest később végeztük el és publikáltuk, azonban didaktikai szempontból a tárgyalását elsőként végezzük.

2.2. Akusztikai rendszer létrehozása gépi tanuló eljárás segítségével a depresszió szűrésére

A disszertációban bemutatott vizsgálat célja, egy olyan gépi tanuló eljáráson alapuló rendszer létrehozása, ami képes a beszéd jellemzői alapján a depresszió elkülönítésére egészséges mintától. Kutatásunkat a BME Távközlési és Médiainformatikai Tanszékével közösen végeztük és létrehoztunk egy magyar nyelvű depressziós páciensektől gyűjtött beszéd adatbázist, amit egy már korábban gyűjtött egészséges adatbázissal tudunk összehasonlítani (n=218). A depressziós páciensek beválogatása a BDI pontszámuk alapján, valamint az adatbázis 20%-nál (n=43) HAMD interjúval való kiegészítéssel történt. A vizsgálat során az alábbi kérdésfelvetésekkel dolgoztunk:

- Milyen mértékben képes egy BDI-vel betanított rendszer felismerni a depressziós súlyosságát?

- Milyen mértékben lehetséges a BDI és HAMD pontszámok együttes használata és javítja-e a modell teljesítményét, amennyiben a BDI adatok helyett a HAMD pontszámokat használjuk a rendszer képzési minták egy részénél?

Feltevésünk szerint a beszéd akusztikai elemzése gépi tanuló eljárások segítségével, a jövőben egy könnyen elérhető és megbízható szűrőeszközként szolgálhat a depresszió diagnosztizálási folyamatában. (Etikai engedély száma: SE-TUKEB 197/2019)

3. Módszerek

3.1. Az első vizsgálat módszerei

Vizsgálati eljárásunk az ún. umbrella review, ami egy vizsgált tudományterületen feltárható korábbi eredmények összefoglalására és integrálására szolgál, korábban már publikált irodalmi áttekintések és meta-analízisek vizsgálatával.

Az elemzésünkhöz a PUBMED, Scopus és ProQuest adatbázisokat használtuk fel a vizsgálati témába vágó releváns közlemények azonosításához. Ehhez a PRISMA irányelv alapján szisztematikus irodalomkeresést végeztünk. Tekintettel a kutatási terület egyre dinamikusabb fejlődésére az elmúlt 5 évet határoztuk meg keresési intervallumként. A keresést 2024. szeptember 18-án végeztük el.

A következő keresőkulcsot használata történt a vizsgálat során: (acoustic*) OR (speech) OR (vocal) OR (audio) AND (depress*) AND („Machine learning”) OR („Deep learning”) OR („Neural network”) OR („automated detection”)

A duplumok, vagyis az adatbázisokban ismételten előforduló rekordok kizárását követően a tanulmányok címe, valamint absztraktja alapján határoztuk meg a vizsgálat tárgya alapján releváns közleményeket. Később az így fennmaradó publikációk teljes szövegét áttekintettük kutatótársammal párhuzamosan, egymástól függetlenül, a kizárásra került rekordok egyeztetésével és rögzítésével együtt. Amennyiben a vizsgálatba történő bevonhatóság alapján a döntésünk nem egyezett, két független értékelő bevonására került sor a döntési folyamatba.

Az alábbi bevonási és kizárási kritériumokat alkalmaztuk az így létrejött adatbázisunkban meglévő közlemények relevanciájának megítélése céljából:

Bevonási kritériumok:

1. A közlemény angol nyelven íródott
2. Az elmúlt 5 évben publikálták
3. PRISMA irányelv alapján készült irodalmi áttekintés a típusát tekintve (annak további típusától függetlenül)
4. legalább részben automatikus, gépi tanuló eljáráson alapuló hangelemzésre épülő detekciós rendszerek vizsgálatára irányul, depressziós betegpopulációban (de nem feltétlenül csak depresszióban)

Kizárási kritériumok:

1. A vizsgálat nem fogalmaz meg következtetéseket a depresszió automatikus hangelemzéssel történő detektálásáról, ideértve hatékonyságvizsgálati eredményeket vagy módszertani megfontolásokat
2. A közlemény teljes szövege nem elérhető

A vizsgálatba bevonásra került közlemények kvalitatív elemzését végeztük el, tekintettel arra, hogy csak egy meta-analízis került beválogatásra. AMSTAR 2 értékelőrendszerrel mértük és értékeltük az elemzésünkbe beválogatott szisztematikus áttekintések és meta-analízis minőségét a kvalitatív elemzést megelőzően, annak érdekében, hogy feltárhatóak legyenek azok a hibák, melyek az áttekintő tanulmányok megbízhatóságát és érvényességét befolyásolhatják. (64)

3.2. A második vizsgálat módszerei

A vizsgálati személyeink a BDI teszt kitöltését követően kerültek be a vizsgálati mintába, melyet az adatbázis 20%-nál HAMD interjú felvételével egészítettünk ki. A vizsgálat során a páciensek beszédét számítógéppel rögzítettük, majd elemeztük. A kutatást az Egészségügyi Tudományos Tanács Tudományos és Kutatásetikai Bizottsága (ETT-TUKEB) hagyta jóvá, az etikai engedély száma: SE-TUKEB 197/2019. Ennek megfelelően minden résztvevő előszóban és írásban történő tájékoztatást követően, hozzájárulási nyilatkozat aláírásával hitelesítette a vizsgálatban való részvételét.

3.2.1. A vizsgálatban résztvevő személyek

Vizsgálatunkban a páciensek a 18 év feletti életkor és a BDI-n depressziósként meghatározható 14 pont vagy afeletti érték esetén kerültek be a kutatási mintába. Kizárási kritériumként határoztuk meg a beszédképzést érintő szomatikus és neurológiai betegségeket az anamnézisben. Másik kizárási kritériumunk a fennálló antipszichotikus terápia, mivel az extrapiramidális tünetek hatással lehetnek az artikulációra, valamint a dopamin receptor blokkoló hatáson keresztül, az ilyen típusú gyógyszerek negatív hatással lehetnek a fogalomalkotásra és fogalmazásra. (65)

Kontrollcsoportunkat, az adatbázis egészséges (nem depressziós) alanyait szintén BDI alapján válogattuk be, 14 alatti pontszám alapján.

A létrehozott adatbázisunk, a Magyar Depressziós Beszédadatbázis (DEPISDA) 218 fő (144 nő, 74 férfi) beszédmintáit tartalmazza. A páciensek bevonása a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikáján, a Péterfy Sándor utcai Kórház pszichiátriai osztályán, a Nyíró Gyula Országos Pszichiátriai és Neurológiai Intézet és a Jahn Ferenc kórház pszichiátriai osztályán történt. A beszédminták gyűjtését a BME Távközlési és Médiainformatikai Tanszékének Beszédakusztikai Laboratóriumával közösen gyűjtöttük és elemeztük. A beszédminták függetlenek, minden páciens csak egyszer rögzítettünk.

3.2.2. Alkalmazott mérőeszközök

A vizsgálati alanyok bevonása során a BDI önkitöltős depressziós állapot felmérését szolgáló kérdőívet használtuk. A BDI az egyik legelterjedtebben használt kérdőív a depresszió szűrésére, ami a beszédelemzéses kutatások során is gyakran kerül felhasználásra. Az 1996-ban a DSM-IV által módosított 21 tételes változata, az elérhető pontszámok alapján négy súlyossági kategóriát különít el. (66) (2. táblázat)

2. táblázat: A BDI által definiált kategóriák és hozzájuk tartozó értékek

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

Az önkitöltős kérdőívek használata a gépi tanuló eljárásokat használó vizsgálatokban előnyös, mivel ennek segítségével gyorsabban, kisebb energiaráfordítással lehet több adatot gyűjteni, és nagy adatbázist létrehozni, ami a hatékony rendszerek létrehozásának előfeltétele. Egy tanulmány szerint a jelenlegi tanulmányok többségében (62%) önkitöltős kérdőíveket használtak a vizsgálati alanyok depressziójának súlyosságának meghatározása céljából. (3) Az önkitöltős kérdőívek gyakran vitatott hátránya azok szubjektivitása, aminek hatására a depressziós tünetek alul- és túlbecslése is megjelenhet. A szubjektivitás alapján torzított adatok következményesen rontják a gépi tanuló eljárások során alkalmazott modellek teljesítményét, ami a kis elemszámú adatbázisok

gyakorisága miatt az alacsony hatékonyság mellett, rossz általánosíthatóságú modelleket eredményez.

Az 1960-ban megalkotott HAMD általunk használt 21 tételes változata a depressziós szomatikus és viselkedésbeli tüneteit is figyelembe veszi, melyek egy része a klinikai általi megfigyelések alapján kerülnek rögzítésre. A NICE guideline alapján a klinikai interjút követően 5 súlyossági kategória kerül meghatározásra a pontszámok alapján. (3. táblázat)

3. táblázat: A HAMD által definiált kategóriák és hozzájuk tartozó értékek

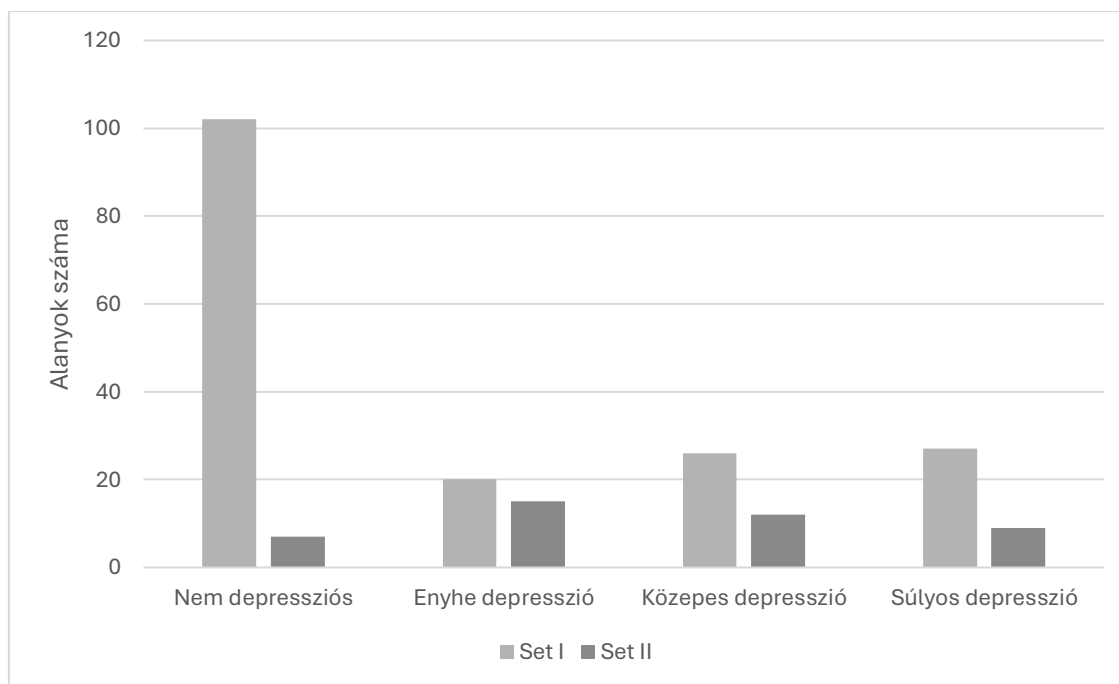
HAMD érték	Kategória neve
0-7	nem depressziós
8-13	enyhe depresszió
14-18	közepes depresszió
19-48 pont	súlyos depresszió

Vizsgálatok azt mutatják, hogy a kezelést megelőzően a BDI és HAMD értékek között a korreláció mérsékelt depresszióban, azonban terápia hatására két teszt közötti korrelációja megnő. Ennek hátterében a mérőeszközök eltérő fókuszú mintavételezése állhat, míg a BDI főképp a szubjektíven is megtapasztalt tüneteket, addig a HAMD a szomatikus és viselkedéses tüneteket hangsúlyozza. (67)

3.2.3. Az adatbázis felépítése

A teljes adatbázis (n=218) 20%-nál, 43 főnél a BDI felvétele mellett HAMD diagnosztikai interjú elvégzése is történt. A vizsgálati alanyok közel fele egészséges, a depressziós minta pedig szinte egyenletesen fedi le a depresszió súlyossági kategóriáit (1. ábra). Az adatbázist a felvett adatok alapján két csoportra bontottuk. Az I. készlet (Set I) olyan mintákat tartalmazott, melyeknél csak a BDI pontszámok álltak rendelkezésre, a II. készlet (SET II) pedig a BDI pontszámok mellett HAMD értékek is elérhetőek voltak. A BDI és HAMD pontszámok egységesítése céljából a két vizsgáló eszköz súlyossági kategóriáinak végpontjait illesztettük egymáshoz, majd az egyes kategóriákon belül lineáris skálázást használtunk, tekintettel arra, hogy az egyes kapcsolat a skálákon belül

lineárisnak tekinthető. Az átskálázott HAMD pontszámokra a továbbiakban H2B (Hamilton to Beck) pontszámokként hivatkozunk. A H2B pontszámok így a HAMD pontszámok átskálázott értékére utal amennyiben elérhető volt HAMD adat, annak hiányában a BDI érték jelzője. Az 4. táblázatban összesítettem az adatbázisfőbb leíró statisztikáit.



1. ábra: A depressziós beszédatbázis elemszámai a H2B általi súlyossági kategóriák szerint az összes, Set I és Set II. csoportosítás tekintetében (4)

4. táblázat: Az elemzésben használt adatbázis legfontosabb jellemzői (4)

		Elemszám	H2B (érték)		BDI (érték)		Életkor	
			Átlag	SD	Átlag	SD	Átlag	SD
Set I	Összes	175	13.5	12.8	H2B értékekkel egyező	H2B értékekkel egyező	44.3	16.0
	Férfi	62	13.3	11.8	H2B értékekkel egyező	H2B értékekkel egyező	43.2	17.9
	Nő	113	13.7	13.4	H2B értékekkel egyező	H2B értékekkel egyező	44.9	14.9
Set II	Összes	43	20.7	7.9	21.4	7.9	34.3	11.7
	Férfi	12	18.1	7.7	23.9	5.6	35.7	10.4
	Nő	31	21.6	7.8	27.4	8.5	33.8	12.4

3.2.4. A vizsgálat menete

A vizsgálat során minden résztvevőtől felolvasott beszédmintákat rögzítettünk, aminek keretében egy kb. 10 mondatból álló mesét („Az északi szél és a nap”) kellett felolvasni az alanyoknak (kb. 45 mp felvétel). A felvételeket csendes helyiségben rögzítettük Lexicon Alpha típusú külső usb audio interfész eszköz segítségével, valamint ATR3350 típusú csiptetős mikrofonnal, ami a vizsgált személyek mellkasához volt rögzítve. A vizsgálat részeként feljegyeztük a résztvevők BDI pontszámát, nemét, életkorát, valamint a dohányzási szokásokat, alkalmazott gyógyszereket és az egyéb beszéddel összefüggésben állható betegségeit, állapotát (pl. felső légúti gyulladás). A vizsgálati alanyok 20%-nál (n=43) HAMD interjú felvétele is történt, annak pontszámának rögzítésével.

3.2.5. Előfeldolgozás, jellemzők kinyerése

A hanganyagok elemzése a BME Beszédausztikai Laboratóriumában történt. A beszédminták csúcsamplitúdóhoz történő normalizálását követően, egy átíráson alapuló, automatikus szegmentációs módszerrel történt az annotációjuk. Ezt a nyelvfüggetlen, automatikus beszéd szegmentációs módszert a Beszédausztikai Laboratórium munkatársai fejlesztették ki, a folyamatos beszéd 9 fonéma osztályba történő szegmentálása alapján. Az osztályozást és szegmentálást rejtett Markov-modellek segítségével végezték. A beszéd és szünetek szegmentálását és annotációját követően fonéma szintű szegmentációt is végeztünk a beszédrészekben belül a SAMPA (Speech Assessment Methods Phonetic Alphabet) nyelvfüggő jelölés használatával. A SAMPA a nemzetközi fonetikus ábécé leírását teszi lehetővé latin betűkkel, megkönnyítve a számítógépes feldolgozást.

Korábbi tanulmányok eredményei a depresszió akusztikai elemzése során az intenzitás, alapfrekvencia, jitter, shimmer, első és második formáns frekvencia, valamint sáv szélességük, és a 13 mfcc (13 mel-frekvencia cepstrális együtthatója) együtthatók változását írták le, melyeket alacsony szintű leíró mutatóknak hívunk (low-level descriptors, LLD). (3, 68) Az alacsony szintű jellemzőket 30 ms Hamming-ablak hosszal és 10 ms időlépéssel számoltuk ki. Az alapfrekvencia, a jitter, a shimmer és a formánsfrekvenciák a beszédminta hanggal rendelkező részeiből lettek kiszámítva, míg a többi LLD-t a teljes beszédmintára számoltuk.

Az LLD mutatókból képzett tényleges leíró jellemzőket az átlag, a szórás és a percentilis tartományok (a rendezett értékek alsó és felső 1, 5, 15 és 25%-nak elhagyása után képzett tartomány) segítségével számítottuk ki. Néhány leíró jellemzőt helyfüggő módszerek alapján számoltunk ki, a beszéd részleten belül több helyen történt a mintavételezés, így az E magánhangzóból (a magyar "e" hangra vonatkozó SAMPA jelölés), az O magánhangzóból (a magyar „a” hangra vonatkozó SAMPA jelölés), az összes magánhangzóból és a teljes beszédrészletre vonatkozóan összesen 282 jellemzőt számoltunk ki az LLD-k segítségével. További jellemzőket is számítottunk, mint például az artikulációs arány, a szünetek aránya, valamint a tranziens részek aránya a beszédrészeken és a teljes beszédmintán. Így összesen 286 jellemzőt számoltunk ki minden egyes beszédmintából. (5. táblázat)

5. táblázat: Az LLD-kból kiszámított leíró jellemzők (4)

	Kiszámítva a következőkből	Összes jellemző száma
Alapfrekvencia	beszédrészek	6
Intenzitás	magánhangzók, teljes beszéd	12
Jitter	E, O, magánhangzók	18
Shimmer	E, O, magánhangzók	18
Első és második formáns frekvencia, és sáv szélességük	E, O, magánhangzók	72
13 MFCC	magánhangzók, teljes beszéd	156
Artikulációs arány	teljes beszéd	1
Szünet arány	teljes beszéd	1
Tranziens részek aránya	teljes beszéd, teljes minta	2
Összesen:		286

3.2.6. Modelltanítás, optimalizálás és statisztikai elemzés

Gépi tanuló eljárásként a Support Vector Regression (SVR) módszert választottuk, ami a Support Vector Machines (SVM) eljárások kifejezetten a regressziós problémák megoldására szolgáló változata. Választásunk amiatt is esett erre a módszerre, mert több tanulmány is kimutatta, hogy az SVM-ek használata jó eredményeket adhat a depresszió automatikus felismerésében, valamint kisebb adathalmazokon is nagyobb pontosság érhető el ezzel a módszerrel. (3, 51, 69) A LibSVM SVR implementációját használtuk radiális bázisfüggvény (rbf) kernellel. (70)

A modelleinket a teljes adatbázis alapján képeztük ki, mivel a II. készlet (Set II) mintaszáma nem lett volna elegendő a betanításhoz. Külön modelleket képeztünk ki az alanyok neme alapján, mivel korábbi tanulmányok alapján, ezzel a módszerrel javítható a depresszió automatikus detektálásának a pontossága. (69, 71)

Az adatbázis kis mérete miatt az SVR-modellek betanítását és tesztelését a leave-one-out cross-validation (LOOCV) módszerrel végeztük el, így minden mintát egyszer választottunk ki tesztkészletként, a modellt pedig a fennmaradó mintákkal képeztük és optimalizáltuk. Az SVR pontossága nagyban függ a bemeneti jellemzővektor és a hiperparaméterek (a modell általános szerkezetét és viselkedését szabályozó paraméterek, melyek befolyásolják a modellparaméterek megtanulását) megfelelő megválasztásától. Így a modellek képzését és optimalizálását is az LOOCV módszer segítségével valósítottuk meg. Ezt a kettős kereszt-validációs technikát 'nested loopnak' nevezzük, esetünkben 'nested LOOCV'-nek. A módszer előnye, hogy a betanító, fejlesztő és a tesztelő készletek teljesen függetlenek egymástól, ezáltal minimalizálható az overfitting (túlillesztés, amikor egy modell túlságosan illeszkedik a betanítás során használt adatokra, beleértve a zajokat és speciális mintázatokat, emiatt kevésbé teljesít jól ismeretlen adathalmazon) esélye, és így reális képet ad az adott módszer pontosságáról és általánosító képességéről.

A jellemzőkészlet méretének csökkentése érdekében kiválasztottuk a depresszió súlyosságával leginkább korreláló 20 jellemzőt (legmagasabb Pearson-korrelációt mutató), majd a redundancia csökkentése és a legjobban teljesítő jellemző készlet megtalálása érdekében 'fast forward' kiválasztási módszert (max. 20 jellemző kiválasztása) alkalmaztunk. Végül a hiperparaméterek optimalizálását rácsos keresési módszerrel valósítottuk meg.

Az statisztikai elemzéshez az átlagos négyzetes hiba négyzetgyöke (RMSE, a statisztikai modell előre jelzett értékei és a tényleges értékek közötti átlagos különbség) mellett az átlagos abszolút hiba (MAE, a tényleges és az előre jelzett értékek közötti különbség abszolút értékének átlaga), valamint az eredeti és az előre jelzett célértékek között a Spearman és Pearson korrelációs együtthatókat számoltuk ki. A regressziós módszer megjósolt értékei alapján osztályozási eljárásokat is végeztünk, ahol az osztályozási pontosságot, szenzitivitást és specifitást mértük. Mindezek mellett az osztályozási modellek ROC (receiver operating characteristic, a modell teljesítményének vizuális

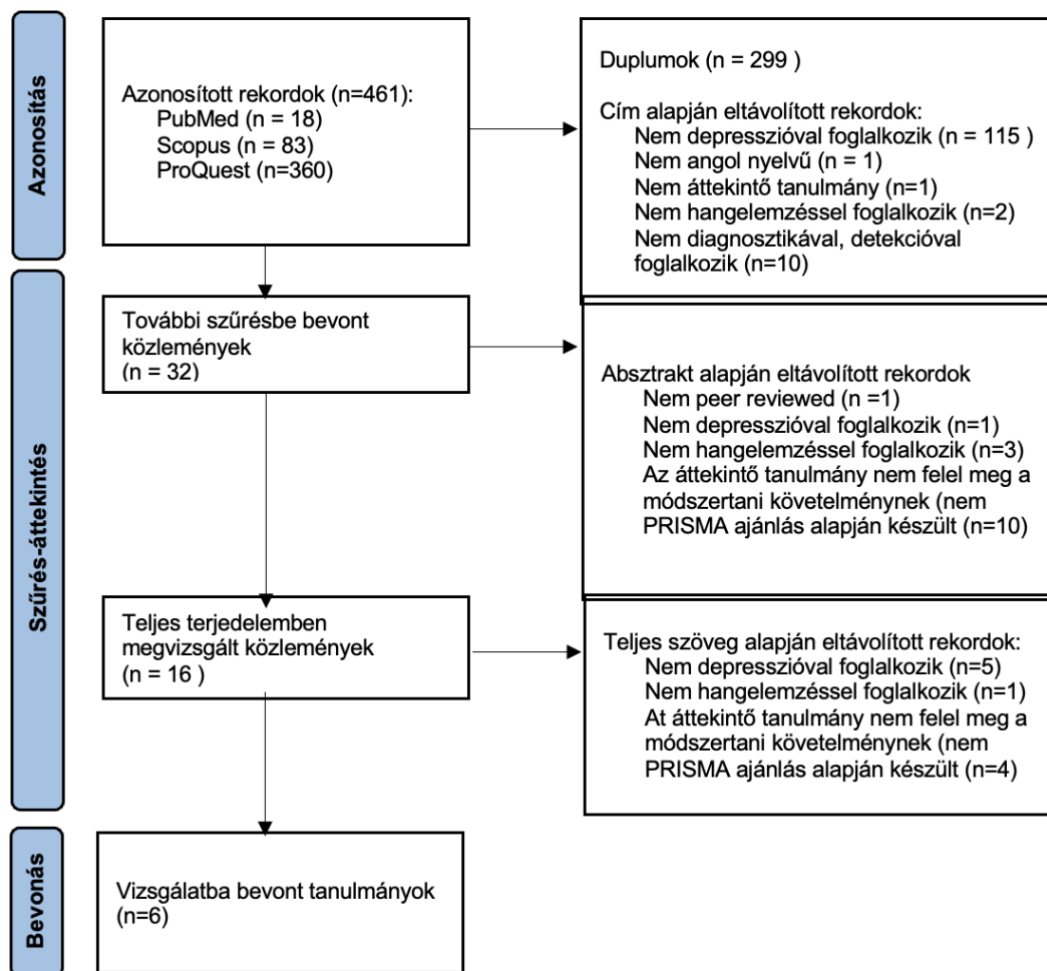
megjelenítése minden küszöbértéknél) görbét és annak görbe alatti területének (AUC, egy kvantitatív diagnosztikai teszt pontosságának mérőszáma) értékét is kiszámoltuk a teljesítményének mérése céljából. Fontos megjegyezni, hogy a 'nested loop' módszer alkalmazása miatt a tesztminták teljesen függetlenek voltak a modellek betanításától és optimalizálásától, így a kapott eredmények a módszerünk valós általánosító képességét tükrözik.

4. Eredmények

4.1. Az első vizsgálat eredményei

4.1.1. Szisztematikus irodomelemzés

A szisztematikus irodalomkutatás során, az azonosított 162 egyedi, nem duplikált rekord közül, a bevonási és kizárási kritériumok alapján 16 tanulmány teljes terjedelemben történő átvizsgálására került sor. Ez alapján további 10 közlemény került kizáráásra. A kiválasztási folyamatot az 2. ábra szemlélteti.



2. ábra: PRISMA irányelvek alapján végzett irodalomkutatás folyamatábrája (5)

A bevonásra került közleményeket AMSTAR 2 értékelőrendszer segítségével vizsgáltuk (6. táblázat). A vizsgálatba bevont közlemények, egy kivételével közepes vagy magas minőségűnek tekinthetők az elemzés alapján. Ez alapján az eredmények ismertetésére és

interpretálására irányuló követelményeknek a közlemények többsége megfelelőnek bizonyult. A továbbiakban így főként a tanulmányok elemzésének módszertani megfeleltethetősége képezte az értékelés alapját. 3 tanulmány egyetlen adatbázisban végzett keresést, és csak 1 alkalmazott független értékelőket az irodalomkutatás során. A protokoll előzetes regisztrációjáról csak 1 tanulmány nyilatkozott, valamint a bevonásra került közlemények irodalomjegyzékének áttekintését is csak 1 tanulmány végezte el.

6. táblázat: A bevont közlemények elemzése AMSTAR 2 értékelőrendszerrel (5)

Közlemény	PICO	Protokoll	RCT/NRSI	Keresési stratégia	Duplikált szelekció	Duplikált beválasztás	Kizárt lista	Találatok leírása	RoB felmérés	Funding elemzés	Meta-analízis -statisztika	Meta-analízis-RoB felmérés	RoB interpretálás	Heterogenitás	Kvantitatív -publikációs torzítás	Conflict of interest	Átfogó értékelés
Hecker P., 2022	I	RI	I	RI	N	N	N	I	N	N	NA	NA	N	I	NA	I	K
Liu L., 2024	I	I	I	RI	I	I	N	I	I	N	I	I	I	I	I	I	M
Low DM., 2020	N	RI	I	N	N	N	N	I	N	N	NA	NA	N	I	NA	I	K
Maaatoug R., 2022	I	RI	I	N	N	N	N	I	N	N	NA	NA	I	I	NA	I	K
Mao K., 2023	I	RI	I	N	N	N	N	I	N	N	NA	NA	N	I	NA	I	K
Oterogonzález I., 2024	N	RI	I	RI	N	N	N	I	N	N	NA	NA	N	I	NA	I	A

I: Igen, N: nem, RI: Részleges igen, NA: Nem alkalmazható, K: Közepes, M: Magas, A: Alacsony, PICO: Population, Intervention, Comparison, Outcome, RCT: Randomized controlled trial/Randomizált kontrollált vizsgálat, NRSI: Non-randomized study interventions/Nem randomizált elrendezés, RoB: Risk of Bias

A 6 vizsgálatba bevont összefoglaló tanulmány főbb jellemzőit a 7. táblázat összesíti. A bevonásra került közleményekben átlagosan 90 tanulmány szisztematikus áttekintését végezték el (min: 36, max: 264), melyekben a depresszió gépi tanuló eljárásán alapuló hangelemzéssel dolgozó detekciójával átlagosan 23 tanulmány foglalkozott. Emögött az áll, hogy a bevonási kritériumoknak megfeleltethető, de a depresszió kivül egyéb pszichiátriai zavar detekciójáról nem nyilatkozó tanulmányból csak 2 található. Egy tanulmány kivételével, ahol a bevonási időszak alsó határa nem került meghatározásra,

az áttekintési időszak 1-10 év között mozgott. A bevonásra került áttekintő közleményekről általánosságban elmondható, hogy a gépi tanulóeljárásokon alapuló predikciós modellek tesztelése szempontjából alacsony mintaelemszámú (50 fő alatt) tanulmányokat vizsgáltak, melyekben a női alanyok aránya általában felülreprezentált volt.

7. táblázat: A bevonásra került tanulmányok főbb jellemzői (5)

Szerző és év	Áttekintő tanulmány típusa	Regisztráció	Vizsgált intervallum	Bevonott tanulmányok száma	Ebből releváns	Vizsgált adatbázis	Vizsgált fókusz	Vizsgált paraméterek	Főbb eredmények
Hecker P, 2022	Szisztematikus áttekintés	Nem regisztrált	2015-2021	43	5	Web of Science, Pubmed, IEEE Xplore	Átfogó áttekintés a neurológiai és pszichiatriai zavarok hang alapú felismerésére vonatkozó tanulmányokról.	Hárgelemzés során használt feladatok típusa, alkalmazott algoritmusok, depresszióval kapcsolatba hozható akusztikai jellemzők	Leggyakrabban alkalmazott feladatok: felolvasás, spontán beszéd, leggyakoribb algoritmus: SVM, DT, RF, Differenciálást segítő akusztikai jellemzők: shimmer, jitter, HNR, össz-színvonal, f0 variabilitás
Liu L, 2024	Szisztematikus áttekintés meta-analízis	PROSPERO	2021-2024	25	25	Web of Science, Pubmed, IEEE Xplore, Medline, Embase, PsycINFO, Scopus	Deep-learning alapú beszédelemzésre épülő diagnosztikai rendszerek pontosságának felmérése depresszióban	A deep-learning modellek becslési pontossága (szenzitivitás, specificitás, pontosság)	Összesített pontosság: 0.87; Összesített szenzitivitás: 0.87 Legmagasabb pontosság CNN-transformer modell esetén: 0.94
Low DM, 2020	Szisztematikus áttekintés	Nem regisztrált	2010-2020	127	55	Google Scholar	Automatizált beszédelemzésre épülő pszichiatriai diagnosztikai rendszerek vizsgálata, elsősorban a beszéd megkülönböztető jegyeknek azonosítása céljával	Depresszióval kapcsolatba hozható akusztikai jellemzők	Differenciálást segítő akusztikai jellemzők: shimmer, jitter, HNR, f0 variabilitás, össz-színvonal. A modellek becslési pontossága a mérések komplexitásából adódó overfitting miatt nem általánosítható
Mantoung R, 2022	Konceptcionális és kritikai áttekintés	Nem regisztrált	2017-2021	45	2	PubMed	A hangulatzavarok digitális fenotípzálására irányuló törekvések elemzése, ide értve a beszédelemzésre épülő diagnosztikát is	Machine learning modellek becslési pontossága (szenzitivitás, specificitás, pontosság), Depresszióval kapcsolatba hozható akusztikai jellemzők	Pontosság: 0,65-0,82 Szenzitivitás: 0,65%; Specificitás: 0,65%; Differenciálást segítő akusztikai jellemzők: megnövekedett válaszadási latencia, szünetek gyakoriságának és hosszának növekedése
Mao K, 2023	Szisztematikus áttekintés	Nem regisztrált	2022-2023	264	18	Google Scholar	Gépi tanulás alapú diagnosztikai rendszerek pontosságának	Depresszióval kapcsolatba hozható akusztikai jellemzők	Differenciálást segítő akusztikai jellemzők: pozitív

4.1.2. Az akusztikai elemzés formái és módszerei

Az MI alapú detekciós modellek, retrospektív elemzések, melyek főképp a meglévő adatok elemzésére és a modellek optimalizálására összpontosítanak. A vizsgálatba bevont tanulmányok kétfajta megközelítés alapján különíthetők el a hangelemzésen alapuló gépi tanulóeljárással működő detekciós lehetőségek területén. A tradicionális megközelítést alkalmazó tanulmányok a kísérleti alanyok hanganyagait MI alapú algoritmusok alapján elemzik, a depresszió meglétére utaló akusztikai jellemzők azonosítása céljából, valamint az ezek által meghatározott predikciós modellek teljesítményének, tradicionális diagnosztikai eszközökkel való összehasonlítását végzik. (3, 72-75) Egy másik megközelítésben a tanulmányok mesterséges intelligencia alapú, ún. beszélgető ügynököket (virtuális asszisztenseket) alkalmaznak, melyek képesek a páciensekkel emberi beszédet szimulálni, interakcióba lépni és azt fenntartani. A tanulmányok többségénél a mesterséges intelligencián alapuló algoritmusok a beszélgetés fenntartásában játszanak szerepet, nem pedig a hangelemzés területén, az állapotfelmérés során pedig standardizált tesztek alkalmaznak. Mindössze 5 olyan tanulmány került azonosításra, ahol az ügynök a beszéd jellemzői alapján tett előrejelzést a depresszió fennállására. (76) Ennek az újfajta megközelítésnek a területén csak egy szisztematikus irodalmi áttekintést tudtunk azonosítani, ami azonban nem részletezi az alkalmazott mesterséges intelligencia alapú megközelítések hatékonyságát, főképp csak technikai leírásokba bocsátkozik, így a tradicionális megközelítést alkalmazó tanulmányok elemzésével foglalkoztunk.

A használt adatbázisokat áttekintve, a gépi tanulóeljárásokat alkalmazó kutatások 60-70%-a, a mélytanulással dolgozó kutatások 40%-a saját felvételekkel dolgozott. A fennmaradó részük nyilvánosan elérhető adatbázisokkal dolgozott, mint pl. az AVEC Audio/Visual Emotion Challenge Workshop vagy a DAIC-WOZ: Distress Analysis Interview Corpus Wizard of Oz adatbázis. Az önállóan gyűjtött hangmintákból épített adatbázisok esetén jelentős eltérés figyelhető meg a minta adatbázis nagysága, demográfiai eloszlása, a depresszió diagnosztikájához használt mérőeszközök, valamint a beszéd rögzítésének módjának területén. A minta adatbázisokon belül a komorbiditások nem kerültek jelzésre a tanulmányokban, ami tovább nehezíti az eredmények összehasonlíthatóságát. (3, 72-75)

A depressziós páciensek azonosításának legelfogadottabb módja a pszichiáter, szakpszichológus által végzett diagnosztika, ami azonban költségessé és időigényessé teszi a vizsgálatot. Ennek redukálása céljából a kutatások során, többnyire önkitöltős tesztel kerülnek alkalmazásra, leggyakrabban a BDI, valamint a PHQ-9 (Patient Health Questionnaire-9) kérdőívek, melyeket mélytanuló alkalmazások esetén több mint 90%-ban használnak, gépi tanuló eljárásoknál kisebb mértékben. Az áttekintő tanulmányokban csak pár eset került azonosításra, ahol klinikus által végzett interjút és önkitöltős tesztet szimultán végeztek. (3, 73-76)

A beszéd minták rögzítése során is több eltérő módszert alkalmaznak a különböző kutatások, leggyakrabban a strukturált interjú, szabad beszéd, szerepjáték, érzelmeket kiváltó történetek elmondása vagy felolvasási feladat szerepel a vizsgálati módszereknél. A bevonásra került tanulmányokban főképp a kötött szöveg, azon belül a felolvasási feladat szerepel leggyakrabban, de a kötetlen beszéd elemzése is gyakori, aminek elemzésével a nyelvi mintákat és akusztikai paramétereket vizsgálták. A két eltérő módszer prediktív teljesítményének összehasonlításáról egy áttekintő tanulmány nyilatkozott, ami a spontán beszéden alapuló elemzést általában jobbnak találta. (75) Kutatás módszertani szempontból fontos a hangminták ellenőrzött körülmények között történő rögzítése, megbízható, jó minőségű eszközök segítségével. Ennek hátránya, hogy nagy erőforrás igénye miatt költség- és időigényessé teszi a mintagyűjtési folyamatot. Alternatívaként több szerző említi az „off site”, okostelefonokkal, okos eszközökkel történő gyorsabb és hatékonyabb adatgyűjtést, azonban ez az adatminőség egyenetlensége miatt kockázatokat rejt magában. (3, 72, 75) Az általunk bevont tanulmányok egyike sem alkalmazta ezt a módszert.

A hanganyagok előfeldolgozása, ami a zajszűrést, a jel normalizálást, torzítások korrigálását jelenti digitális jelfeldolgozó eszközökkel végzik, azonban az általunk vizsgált összefoglaló tanulmányokban a kutatások ennek a módszertanát nem elemzik. A releváns akusztikai paraméterek kinyerése (feature extraction) során a tanulmányok leggyakrabban nyílt forráskódú platformokat (openSMILE, COVAREP és openEAR), valamint Python alapú könyvtárat alkalmaznak (pyAudioAnalysis). A kutatások során leggyakrabban vizsgálat akusztikai paramétereket és azok meghatározását a 8. táblázatban összesítem.

8. táblázat: Depresszióval összefüggésben vizsgált akusztikai jellemzők listája és azok rövid meghatározása (5)

Jitter	A zöngé periódusainak időbeli ingadozását méri, jelezve a hangadás stabilitását vagy instabilitását.
Shimmer	A zöngé periódusainak amplitúdóbeli ingadozását méri, jelezve a hangadás stabilitását vagy instabilitását.
Tremor	A hangszalagok rezgésének frekvenciabeli ingadozását méri, amely a hang stabilitásának periodikus változásait jellemzi.
Harmonics-to-noise ratio (HNR)	Az alaphfrekvencia (f0) és a zajenergia arányát méri a hangban, és a hang tisztaságát, valamint a vokális zaj jelenlétét jellemzi.
Quasi-open quotient (QOQ)	A hang minőségét és természetességét értékeli különböző paraméterek, például a hangszín és a tisztaság alapján.
Normalized Amplitude Quotient (NAQ)	A beszéd érthetősége és a háttérzaj arányát méri, ezzel segítve a beszéd minőségének értékelését.
Peak slope	A hanghullám csúcsának emelkedési sebességét méri, amely a hang intenzitásának gyors változását jelzi.
F1 átlag	Első formáns frekvencia átlaga.
F2 átlag	Második formáns frekvencia átlaga.
F1 variabilitás	Az (F1) frekvenciájának szórása.
F2 variabilitás	A második formáns (F2) frekvenciájának szórása.
F1 tartomány	Az első formáns frekvenciáinak intervalluma.
Vowel space	A vowel space (magánhangzó-tér) egy akusztikai tér, amely a magánhangzók formáns-frekvenciáinak (F1 és F2) alapján mutatja a nyelv helyzetét a beszéd során, és vizuálisan ábrázolja a különböző magánhangzók képzését.
Mel Frequency Cepstral Coefficients (MFCC)	A beszéd Mel-frekvenciák szerinti sávokra bontott intenzitás értékeinek a diszkrét koszinusz transzformációjával nyert együtthatói.
f0 átlag	A beszédhang alaphangjának átlagos értéke.
f0 tartomány	Beszédhang alaphangjának legkisebb és legnagyobb értékei közötti intervallum.
f0 variabilitás	Beszédhang alaphangjának szórása.
Intenzitás átlag	A hangenergia (hangnyomás) időbeli átlagolt mértéke.
Intenzitás variabilitás	A hang intenzitásának időbeli változékonyságát jelenti.
Energia sebesség	A beszédjel intenzitásának időbeli változását méri, vagyis azt mutatja meg, hogy a hangenergia milyen gyorsan ingadozik a beszéd során.
Energia változékonyság	A hanghullámok által közvetített energia változékonyságát jelenti.
Beszédsebesség	Másodpercenkénti beszédhangok száma egy adott beszédszakaszon a szüneteket is figyelembe véve.
Artikulációs sebesség	Másodpercenkénti beszédhangok száma egy adott beszédszakaszon a szünetek nélkül.
Átlagos szünethossz	A beszédben előforduló szünetek átlagos időtartama.
Szünet variabilitás	A beszédben előforduló szünetek időtartamának szórása.
Szünet gyakoriság	A beszédben előforduló szünetek és nem szünetek időbeli aránya.
Össz-szünethossz	A beszédben előforduló szünetek összhosszát jelenti.

A szükséges jellemzők kinyerését követően az adatok elemzéséhez a következő lépés a gépi tanuló eljárás kiválasztása és a modelltanítás. A depresszió akusztikai kutatásai között a leggyakrabban alkalmazott módszer az SVM, ami kis adathalmazokon is képes osztályozó feladatok elvégzésére, ellentétben a viszonylag ritkán alkalmazott

mélytanulással, annak nagyobb adatigénye miatt. Egyéb alkalmazott gépi tanulóeljárások még a Decision Tree (DT), ami klinikusok számára is könnyen átlátható és interpretálható modelleket kínál, valamint a Random Forest (RF), aminek segítségével a magas zaj arányú adatok is vizsgálhatóak, valamint előnyösebben csökkenti az overfittinget. (3, 72, 73, 75) A mélytanuló eljárások között a leggyakoribb a Convolution Neural Network (CNN) alkalmazása, aminek előnye, hogy nyers adatokból képes komplex mintázatokat kinyerni, előzetes kézzel történő jelölések nélkül is, valamint komplex összefüggések megtanulására és modellezésére is alkalmas. (73) A különböző gépi tanuló eljárások kiválasztása gyakran a meglévő audio jelek jellemzőitől függ, így hatékonyságuk nem feltétlenül tér el egymástól. A különböző mesterséges intelligencián alapuló eljárások előnyeit és hátrányait a 9. táblázatban részletezem.

9. táblázat: Gépi tanuló eljárások előnyei, hátrányai (77-79)

Eljárás típusa	Előnyök	Hátrányok
Támogató vektor gépek (SVM)	- Nagydimenziós adatokra is lehet használni - Jól kiválasztott hipersíkkal azonosítani tudja, akkor jobb teljesítmény érhető el	- Pontos kernelfüggvény azonosításának nehézsége lineárisan nem elkülöníthető adathalmazok esetén - Nagy zaj arányú adathalmaz esetén nem működik jól
Döntési fák (Decision Trees)	- Kategorikus és numerikus adatokra egyaránt használható - Könnyen értelmezhető - Kevesebb előfeldolgozásra van szükség - Gyorsabb adatfeldolgozással jár	- Nem kiegyensúlyozott - Globálisan optimális megoldások helyett lokálisan optimális megoldásokat ad
Véletlen erdők (Random Forest)	- Csökkenti a túlillesztés kockázatát - Nagyobb pontosság a több döntési fa miatt - Képes kezelni a hiányzó adatokat	- Magasabb számítási igény, mint egyetlen döntési fánál - Lassabb előrejelzési idő
Konvolúciós neurális hálózatok (CNN)	- Jó teljesítmény képfelismerési és vizuális problémákban - Automatikusan felismeri a releváns mintákat - Alkalmazható térbeli és időbeli adatokra	- Magas számítási erőforrás igény - Nagy adathalmazokra van szükség az eredményességhez - Nehezen értelmezhető modellek („fekete doboz”)

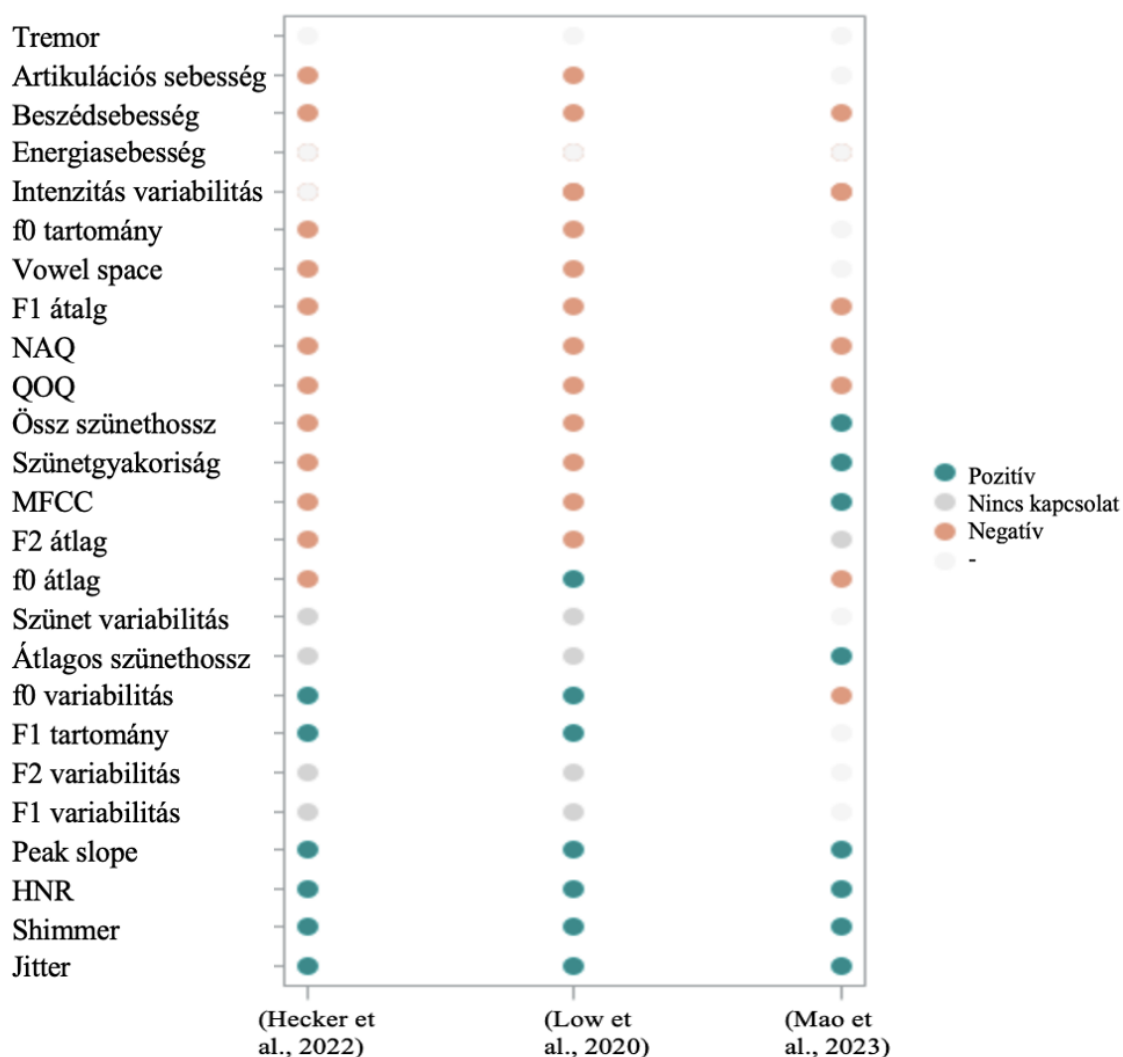
A beszédrészletek hossza is befolyásolhatja a predikciós teljesítményt, ennek a gyakorlatáról az általunk vizsgált összefoglaló tanulmányok közül egy nyilatkozott, ami alapján a legjobb teljesítményt (pontosság: 0,94, szenzitivitás: 0,92) 9 másodperces szegmensekkel érték el, általánosítva pedig az 5 másodpercnél hosszabb szegmensek alkalmazásának előnyét írták le. (73)

Az optimális teljesítmény elérése céljából az ún. hiperparaméterek alkalmazását használják a kutatások során, melyeket nem az adatokból, hanem interaktív tréningezési eljárásokon keresztül állítanak be különböző beállítások kipróbálásával. Ezek a beállítások befolyásolják a modell tanulási képességének gyorsaságát és hatékonyságát. Hiperparaméterre példa az epochszám (hány teljes ciklust végez a modell az adatokon a tréningezése során), a CNN modell esetén pedig a konvolúciós rétegek száma. Az általunk vizsgált tanulmányok többsége ugyan kitér a hiperparaméter beállítások fontosságára, azonban azok részletezését nem fejtik ki. (3, 72, 73, 75) Ezzel szemben egy tanulmány kiemeli, hogy fontossága ellenére több kutatásból hiányzik a hiperparaméterek pontos leírása. (73)

Az akusztikai modellek teljesítményének értékelése céljából a leggyakrabban használt mutatók a (az esetek több mint 80%-ban) a pontosság, érzékenység, specificitás és az F1 érték. Ennél ritkábban használják még az AUC, RMSE, MAE és CCC (Concordance correlation coefficient) mutatókat a teljesítmény jellemzésére. (72, 73, 75)

4.1.3. A depresszióval összefüggésbe hozható hangtani jellemzők

A vizsgálatunkba bevont közlemények összegzése alapján több hangtani paraméter depresszióra jellemző prediktív értéke megerősíthető volt. A leggyakrabban vizsgált jellemzőket a 3. ábrán mutatom be. Pozitív kapcsolat esetén az egyik változó növekedése a másik változó növekedésével, míg negatív kapcsolat esetén az egyik változó növekedése, a másik változó csökkenésével jár együtt. A nincs kapcsolat jelölésnél az egyik változó változása nem befolyásolja a másik változót. A '-' jelölés esetén az összefüggést a bevont tanulmányok nem vizsgálták



3. ábra: A leggyakrabban vizsgált akusztikai jellemzők és a depresszió összefüggéseinek grafikus összefoglalása (5)

A 3. ábra alapján jól látható, hogy az általunk vizsgált összefoglaló tanulmányok között nagy mértékű a korreláció, a depresszió és az akusztikai jellemzők közötti kapcsolat vonatkozásában. Az f0 értékének csökkenését, ami jól tükrözi a szubjektíven is tapasztalható monoton hanglejtést, következetesen megfigyelték a vizsgált kutatások depresszió esetén, hasonlóan az F1-átlag, a NAQ és QOO változókkal való negatív kapcsolattal, aminek háttérében a gondolkodás sebességének csökkenése, csökkent reakcióidő állhat. A gondolkodás, a mozgás és a reakcióidő sebességének csökkenése mellett a beszédminőség romlására, valamint a hangzásbeli variabilitás csökkenésére is utalhat a jitter, a shimmer, az úgynevezett Peak slope, valamint a HNR jellemzők depresszió súlyosságával együtt történő emelkedése. (3, 72, 75) Ellentmondások is

megfigyelhetőek a tanulmányok összehasonlítása során pl. az MFCC, a szünethossz és szünetgyakoriság jellemzők területén. Ennek hátterében fontos kiemelni, hogy a depresszió egy gyakran komorbid állapotokkal együttjáró, széles spektrumú, eltérő súlyosságú betegség, emiatt az akusztikai jellemzők a depressziós állapottal komplex interakciót mutathat. Az akusztikai jellemzők a vizsgált személy aktuális érzelmi állapotától, a beszédkörnyezetétől és a kommunikációs szándékától függően is változhat. Az MFCC jellemzőnél észlelt ellentmondás kapcsán például felmerül annak az aktuális érzelmi állapotra és kommunikációs kontextusra való érzékenysége, ami az eltérő eredmények hátterében állhat. Az össz-szünethossz és szünetgyakoriság esetén észlelt eltérésnél a növekedés utalhat a depressziós személyek nehezített gondolat formálására, a csökkenés a társuló feszültségre, valamint a gondolkodási nehezítettség gyorsult beszéddel, szünetek hiányával történő kompenzációjára is.

A klinikai magyarázatok mellett az eltérő módszertani megközelítések is az eltérések hátterében állhatnak. A leggyakrabban megjelenő probléma az eltérések hátterében a kisebb elemszámú adatbázisokon végzett vizsgálatok során tapasztalható instabil eredmények. Ez a vizsgálatunkban csak az f0 átlag eltérésénél merül fel, mert a többi változó esetén a legkisebb és legnagyobb adathalmazzal dolgozó összefoglaló tanulmányok eredményei konzisztensek egymással, az eltérések főképp a közepes mintaelemszámú tanulmányoknál észlelhetőek. Az eltérések hátterében felmerülhet a különböző hangfelvételi és elemzési technikák szerepe, azonban a bevonásra került tanulmányok közül csupán egy nyilatkozik a tárgyalt kutatások mérési módszereiről. (72) Különbséget adhat még a bevont alanyok klinikai jellemzőinek összessége is (pl. életkor, nem, beszélt nyelv, alkalmazott gyógyszeres kezelés, egyéb komorbiditások), azonban ezek elemzésére nem került sor a beválogatásra került tanulmányok esetén. (3, 72, 75, 76)

4.1.4. A prediktív modell teljesítménye

A vizsgált tanulmányok alapján a gépi tanuló eljárásokon alapuló prediktív modellek teljesítménye széles körben mozog. Gépi tanulási modelleknél a MAE 5,77- 7,08 között, a pontosság 0,75-0,86 között, a mélytanulás során a pontosság 0,57-0,99, a duplikált adatkészletek kizárását követően pedig a pontosság 0,75-0,9 között mozog. (3, 73, 74) A gépi tanuló eljárások hátterében többnyire két eltérő validációs eljárást alkalmaznak, melyek a hold-out (kihagyásos) és cross (kereszt) validációk (pl. K-fold, nested). A hold-

out validációk esetén az adathalmazt tanulási és tesztelő részekre bontják, és a tesztelő halmazt csak egyszer használják fel a modell teljesítményének értékeléséhez. A keresztvalidáció során a mintát több részre osztják és a mérést több alkalommal végzik el, mindig eltérő tanulási és tesztelő adathalmaz felhasználásával. A különbségek miatt a hold-out módszer egyszerűbb és gyorsabb, azonban főleg kis adatbázisokon kisebb prediktív teljesítményű a keresztvalidációhoz képest, ami részletesebb és megbízhatóbb értékeléssel is szolgál. A keresztvalidáció előnye még az overfitting csökkentése, azonban a túlságosan bonyolult modellek, valamint magas zajarányú tanulási adathalmazok esetén annak veszélye továbbra is fennállhat. (3, 73, 75) Low és mtsai tanulmányukban megjegyzik, hogy a kutatások nagy része keresztvalidációs technikával dolgozik, és gyakran ugyanazt a nyílt audio adatbázist (pl. AVEC) használják a modellek tesztelése céljából, ami segítheti a különböző modellek prediktív teljesítményének összehasonlíthatóságát. (3) A bevonásra került tanulmányok elemzése alapján a teljesítmény szempontjából a mintanagyság, a mintavétel körülményei és technikája, a zavaró tényezők kontrollálása (ami a legtöbb tanulmány esetében elmarad), a minta további jellemzői (nem, beszélt nyelv, életkor, kulturális háttér), valamint a minta előfeldolgozása egyaránt és együttesen bizonyult meghatározónak. Egyedi jellemzőkön mért különbségek ugyan nehezen mérhetőek, de az LLD és MFCC értékek elemzése mélytanuló eljárásokkal következetesen 80% fölötti predikciós pontosságot eredményezett. (73) Ugyan csak néhány tanulmány vizsgálta a multimodális elrendezéseket, de az alapján elmondható, hogy az a predikciós teljesítményre nézve pozitív hatással rendelkezik. Emellett a mélytanulást használó eljárások is nagyobb prediktív teljesítménnyel dolgoztak, mint a gépi tanuló eljárásokat alkalmazó vizsgálatok. (3, 73-75)

4.2. A második vizsgálat eredményei

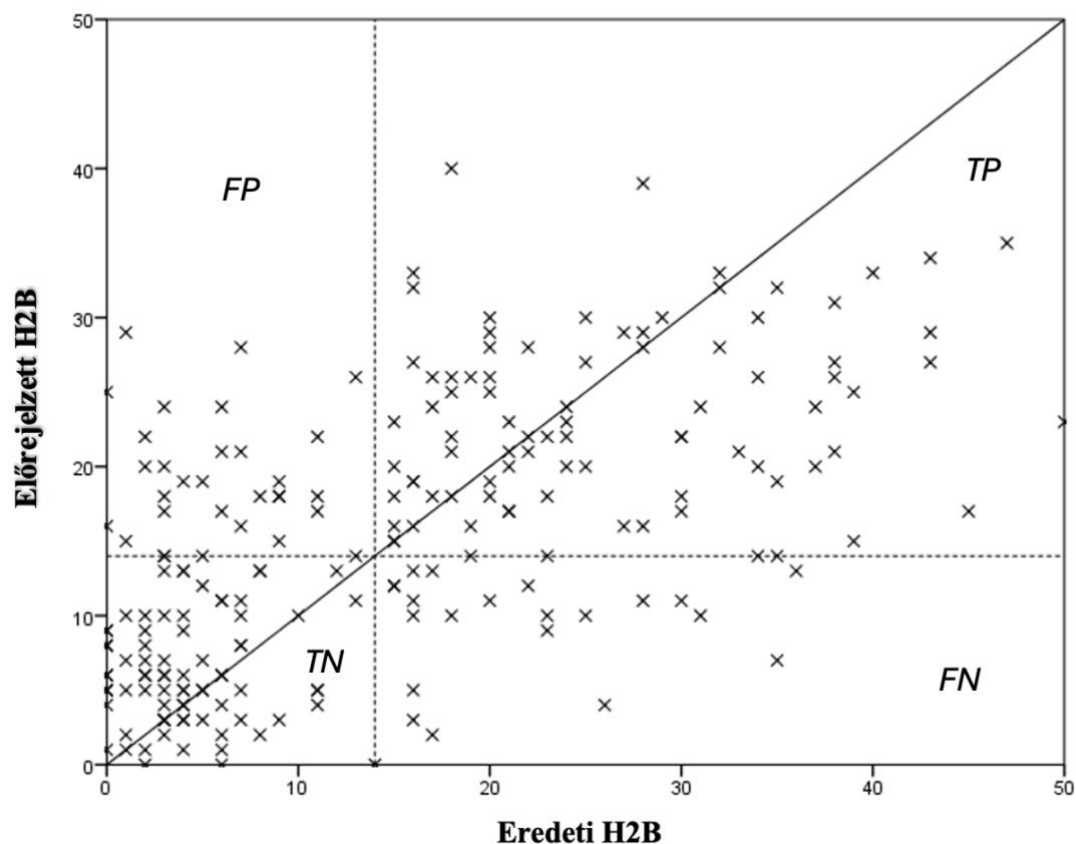
4.2.1. A depressziós súlyosságának becslése BDI alapján

Ebben a kísérleti elrendezésben a modelltanítást és optimalizálást a BDI-pontszámok alapján végeztük. A legjobb eredményeket a jellemzővektorokon belüli öt jellemzővel kaptuk, mind a férfiak, mind a nők esetében, nested LOOCV használatával (RMSE: 11,5 a nők esetében és 8,0 a férfiak esetében).

Mivel a modelloptimalizálás és a tesztelés nested loop eljárással történt, emiatt elképzelhető, hogy minden egyes tesztmintához eltérő modellt választott a rendszer. Emiatt azokat a jellemzőket alkalmaztuk, amelyeket az összes eset legalább 90%-ban kiválasztottak a modellek, azonban néhány jellemző erősen korrelált egymással, így könnyen fel is cserélhetőek. A női modellek esetében ezek a jellemzők a következők voltak: a tranziensek aránya a teljes beszédrészleteken, az mfcc6 szórása a magánhangzókon és az mfcc1 átlaga a teljes beszédrészleteken. A férfi modellek esetében ezek a jellemzők a tranziensek aránya a teljes beszédrészleteken és az mfcc4 átlaga a magánhangzókon voltak. A modelleket mind a BDI, mind a H2B célváltozókra teszteltük, az eredmények az 10. táblázatban láthatók. A férfi és női készletekre vonatkozó eredményeket „F” és „N” jelöli. Az előrejelzett és az eredeti H2B-értékek szórásdiagramja a 4. ábrán látható, ahol a depresszió határát szaggatott vonal jelzi (H2B = 14), és az $x = y$ (tökéletes döntés) egyenest is ábrázoltam.

10. táblázat: A depresszió előrejelzésének pontossága BDI-pontszámok felhasználásával a tanítás során (4)

	Cél	RMSE	MAE	Pearson	Spearman
Set I	BDI	10.1	7.7	0.63	0.59
	(=H2B)	(F: 7.7; N: 11.2)	(F: 6.2; N: 8.5)	(F: 0.76; N: 0.56)	(F: 0.73; N: 0.52)
Set II	BDI	11.6	8.7	0.19	0.24
		(F: 9.5; N: 12.4)	(F: 7.5; N: 9.2)	(F: 0.32; N: 0.14)	(F: 0.11; N: 0.15)
	H2B	9.3	7.1	0.21	0.27
		(F: 6.0; N: 10.3)	(F: 4.7; N: 8.0)	(F: 0.56; N: 0.15)	(F: 0.55; N: 0.25)
Összes	BDI	10.4	7.9	0.61	0.61
		(F: 8.0; N: 11.5)	(F: 6.4; N: 8.7)	(F: 0.73; N: 0.56)	(F: 0.71; N: 0.56)
	H2B	9.9	7.6	0.61	0.61
		(F: 7.4; N: 11.0)	(F: 5.9; N: 8.4)	(F: 0.76; N: 0.54)	(F: 0.74; N: 0.54)

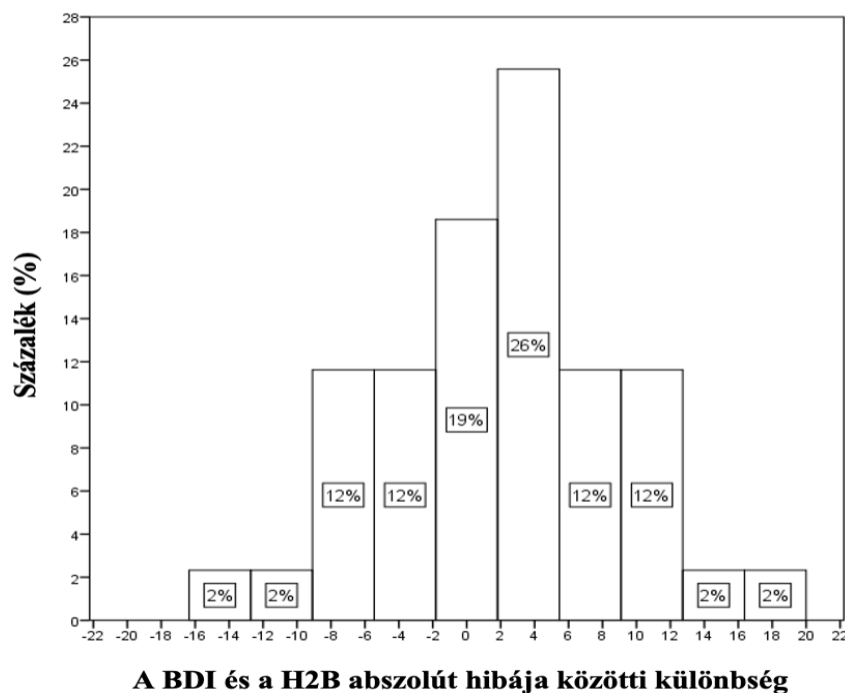


4. ábra: A depresszió előrejelzésének pontossága, BDI-pontszámok felhasználásával a tanítás során (FP: False positive, FN: False negative, TP: True positive, TN: True negative) (4)

Páros t-próbát alkalmaztunk annak vizsgálatára, hogy a BDI és a H2B tesztelése közötti átlagos abszolút hiba közötti különbség szignifikáns volt-e a II. készlet esetében, ennek eredményeként $p = 0,129$ -et kaptunk, ami alapján nem beszélhetünk szignifikáns különbségről. A férfiak esetében megfigyelhető, hogy a modell kisebb hibával tudta megbecsülni a depresszió súlyosságát ($p = 0,007$, két mintás t-próba alapján). Ennek egyik lehetséges magyarázata, hogy az eredeti BDI-pontszámok átlaga és szórása kisebbnek tűnik a férfiak esetében (férfiak: $15,0 \pm 11,7$; nők $16,7 \pm 13,7$), és a modell pontossága rosszabb volt a súlyosabb depresszió esetében, bár az eredeti BDI-pontszámok értékeiben nem mutatható ki szignifikáns különbség két mintás t-próbával ($p = 0,381$). Első pillantásra az I. és a II. készlet BDI-tesztelésénél kapott eredmények nem tűnnek szignifikánsan eltérőnek az RMSE hibaérték esetében (I. készlet: 10,7; II. készlet: 11,6, $p = 0,370$ két mintás t-próba alkalmazásával), azonban a II. készlet esetében a korrelációs értékeknél szignifikáns romlás figyelhető meg ($p < 0,001$ a Fisher-féle r to z

transzformációval), ebből arra lehet következtetni, hogy a modell teljesítménye eltérő az I. és a II. készlet esetében. Fontos megjegyezni, hogy míg az I. készlet a teljesen egészségesektől a súlyosan depressziósokig terjedő mintákat tartalmazott, a II. készlet elsősorban a depressziós alanyok beszédmintáit tartalmazza (1. ábra). Ezekből arra lehet következtetni, hogy az eljárás elsősorban a teljes skálát képes megfelelően elkülöníteni, egy szűkebb tartomány esetén kevésbé megbízhatóan differenciál.

A II. készlet esetében (ahol mind a BDI-, mind a HAMD-pontszámok rendelkezésre álltak) nem tudtunk szignifikáns különbséget kimutatni, az előrejelzett értékek BDI-vel vagy a H2B-vel való összehasonlítása során. Lehetséges azonban, hogy ez a minták kis száma miatt volt ($n=43$), ezért további leíró elemzéseket végeztünk. A MAE hibaérték esetében 8,7 hibaértéket kaptunk a BDI eredményekre vonatkozóan, szemben a 7,1-es H2B előrejelzett hibaértékkel (18%-os relatív javulás). Általánosságban ez átlagosan 1,7-tel közelebbi becslést jelent a H2B-pontszámhoz. Ennek alaposabb vizsgálatához megnéztük az abszolút hibák különbségét az egyes minták esetében, amikor a BDI vagy a H2B célváltozót használtuk. A különbségek hisztogramja az 5. ábrán látható. A negatív különbség a BDI-hez közelebbi eredményt jelent, míg a pozitív különbség a H2B-hez közelebbi eredményt.



5. ábra: Az abszolút hiba különbségeinek hisztogramja a BDI és H2B esetében a második készleten (4)

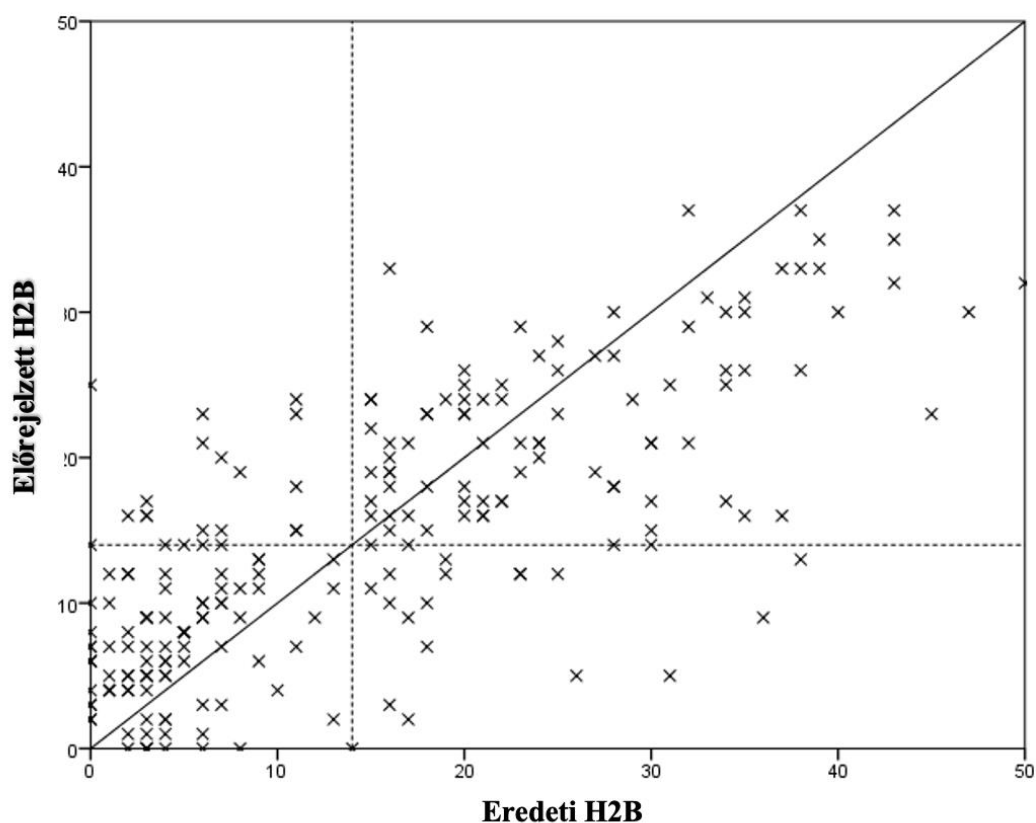
A 5. ábra azt mutatja, hogy az esetek 19%-ban az előre jelzett pontszám közel azonos eltérést mutat a BDI és a H2B pontszámokhoz képest. Az esetek 26%-ban azonban 2-6 értékkel közelítette meg a H2B-t, 24%-ban 6-13 értékkel, míg további 4%-ban magasabb értékkel, ezzel szemben csak az esetek 12%-ban közelítette meg a BDI-t 2-6 értékkel, 14%-ban 6-13 értékkel, míg további 2%-ban magasabb értékkel. Ezek alapján megállapítható, hogy a becsült pontszámok általában közelebb állnak a H2B pontszámhoz.

4.2.2. A depressziós súlyosságának becslése H2B alapján

Az előzőek alapján leírható, hogy a BDI-vel képzett modellek pontosabban képesek becsülni a H2B-t, mint maga a BDI. Ezért megvizsgáltuk, hogyan változik a módszer pontossága, ha a BDI-t a H2B-vel helyettesítjük a II. halmaz esetében (a HAMD csak ehhez a halmazhoz állt rendelkezésre, ami eltérő BDI- és H2B-pontszámokat eredményezett). A legjobb eredményeket a férfiak esetében 7-es, a nők esetében pedig 9-es méretű jellemzővektorral értük el. Az előző vizsgálatához hasonlóan azok a jellemzők, amelyeket a modellek az esetek legalább 90%-ban kiválasztottak, a női modellek esetében a következők voltak: a tranziensek aránya, az intenzitás szórása a teljes beszédrészekben, az mfcc1 átlaga a magánhangzókon, az mfcc2 átlaga a teljes beszédrészekben valamint az mfcc6 szórása a magánhangzókon. A férfi modellek esetében ezek a jellemzők a következők voltak: a tranziensek aránya a teljes beszédrészekben, az mfcc1 átlaga a teljes beszédrészekben és az mfcc4 a magánhangzókon. Az eredményeket a 11. táblázat mutatja be. Az előrejelzett pontszámok és az eredeti H2B pontszámok összehasonlítása a 6. ábrán látható, ahol a depresszió határát szaggatott vonal jelzi ($H2B = 14$), és az $x = y$ (tökéletes döntés) egyenest is ábrázoltuk.

11. táblázat: A depresszió előrejelzésének pontossága H2B-pontszámok felhasználásával a tanítás során (4)

	Cél	RMSE	MAE	Pearson Coef.	Spearman Coef.
Set I	H2B	8.2 (F: 5.7; N: 9.3)	6.3 (F: 4.4; F: 7.4)	0.77 (F: 0.88; N: 0.72)	0.72 (F: 0.85; N: 0.65)
Set II	H2B	8.2 (F:3.8; N: 9.4)	6.1 (F: 3.1; F: 7.3)	0.42 (F: 0.87; N: 0.25)	0.46 (F: 0.79; N: 0.32)
All	H2B	8.2 (F: 5.4; N: 9.3)	6.3 (F: 4.2; F: 7.4)	0.75 (F: 0.88; N: 0.69)	0.73 (F: 0.87; N: 0.67)



6. ábra: A depresszió előrejelzésének pontossága, H2B-pontszámok felhasználásával a tanítás során (4)

Az eredmények alapján az Összes készlet esetében 8,2 RMSE hibaértéket kaptunk, ami 17%-os relatív javulást jelent a modellek tisztán BDI-vel történő tanításához és optimalizálásához képest (9,9 RMSE).

Páros t-próbát alkalmaztunk annak eldöntésére, hogy ez a különbség szignifikáns volt-e vagy sem, ami alapján $p < 0,001$ értéket kaptunk, ez azt jelenti, hogy a szignifikáns különbség 99,9%-os szignifikanciaszinten megállapítható. Hasonló javulás volt tapasztalható az I. készlet (19%, $p < 0,001$) és a II. készlet (12%, $p < 0,001$) esetében. Fontos megjegyezni, hogy az eredmények nemcsak a II. készlet, hanem az I. készlet esetében is javultak, amelyeket a BDI pontszámok alapján kerültek kiképzésre (az I. készlet esetében $BDI = H2B$). Ez alapján kijelenthető, hogy a beszédjel alapú automatikus modellünk jobban tanítható a H2B adatokkal, mint a BDI adatokkal.

A javulás mértéke a 4. és 6. ábra összehasonlításával is megfigyelhető, mivel a második esetben a becsült pontszámok közelebb vannak az $x = y$ egyeneshez (tökéletes előrejelzés), és kevesebb minta került a bal felső és a jobb alsó téglalapokba, amelyek hibás felismeréseket jelentenek (bal felső: hamis pozitív, jobb alsó: hamis negatív).

4.2.3. A depressziós és egészséges alanyok elkülönítése

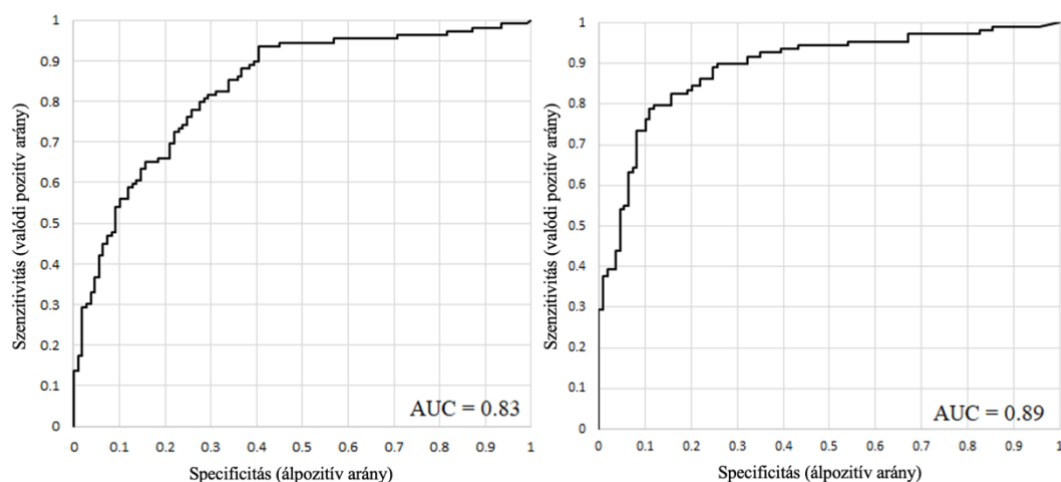
Kimutatható, hogy a beszédjelek feldolgozása alapján pontosan meg lehet jósolni a depresszió súlyosságát, azonban az RMSE és MAE hibaértékek nagyságából nehéz közvetlenül következtetni a módszer alkalmazhatóságára. Mivel a hosszú távú cél a depresszió diagnózisának támogatása, megvizsgáltuk, hogy a modell milyen pontosan képes különbséget tenni a depressziós és egészséges alanyok között. A regressziós módszer segítségével a klasszifikáció is megvalósítható, hiszen megadható egy összehasonlító érték, amely felett az adott személy depressziósnak, míg az alatt nem depressziósnak minősül. Ennek alapján megvizsgáltuk, hogy a BDI-vel és a H2B-vel betanított modellek segítségével milyen pontosan lehet megkülönböztetni a depressziós és a nem depressziós alanyokat. A 7. ábra a két modell ROC-görbéit és AUC-értékeit mutatja, a 12. táblázat pedig az osztályozási szenzitivitásukat, specificitásukat, valamint a maximális osztályozási pontosságukat, valamint amikor az összehasonlító határértéket úgy állítottuk be, hogy legalább 90%-os szenzitivitási értéket érjen el.

A Wilcoxon féle előjeles rangpróba segítségével megvizsgáltuk, hogy a különbség szignifikáns-e a maximális pontosság esetében, és $p = 0,022$ értéket kaptunk, ami azt jelenti, hogy a különbség szignifikáns. A vizsgálat alapján megállapítható, hogy a H2B használata a modell tanításban javítja az osztályozást, mivel magasabb AUC (görbe alatti terület) értéket (0,89), magasabb maximális osztályozási pontosságot (84%), magasabb

specifitást (73%) és osztályozási pontosságot (81%) értünk el 90%-os szenzitivitás mellett.

12. táblázat: A depressziós és egészséges alanyok osztályozásának pontossága, H2B vagy BDI pontszámok által történő betanítás esetén (4)

	Betanító változó	Osztályozási pontosság	Szenzitivitás	Specifitás
Maximális osztályozási pontosságnál	BDI	76%	80%	72%
	H2B	84%	79%	89%
90%-os szenzitivitásnál	BDI	75%	90%	62%
	H2B	81%	90%	73%



7. ábra: Az akusztikus modellek ROC-görbéje, amikor a modelleket BDI-vel (balra) és H2B-pontszámokkal (jobbra) tanítottuk (4)

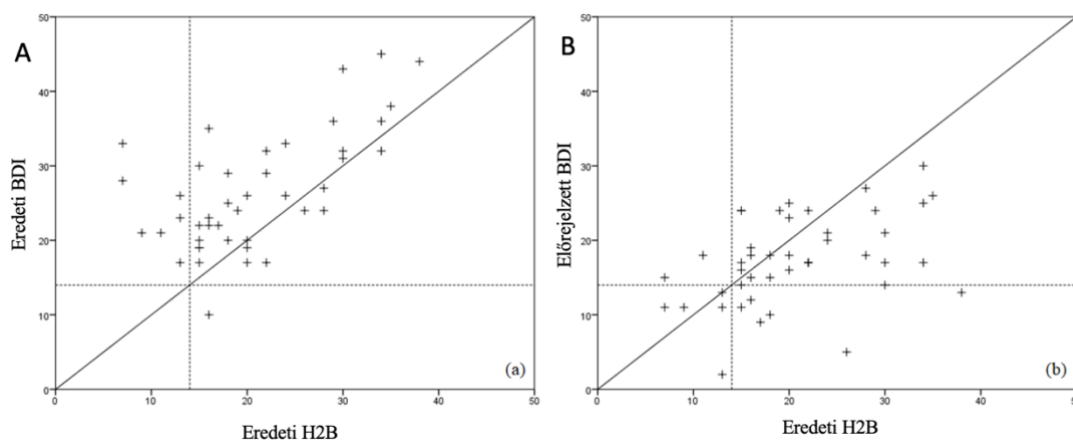
4.2.4. Az akusztikus modell és a BDI pontosságának összehasonlítása

Mivel a BDI és a HAMD pontszámok egyaránt rendelkezésre álltak a II. készlet esetében, ez lehetővé tette a BDI kérdőív és a kifejlesztett akusztikus modell pontosságának összehasonlítását. Tekintettel arra, hogy az adatbázis főként depressziós alanyokat tartalmaz, ebből az összehasonlításból csak korlátozott következtetéseket lehet levonni, és nem jó leírása a BDI-kérdőív pontosságának általában. Érdekes lehet azonban a két módszer összehasonlítása, mivel a hosszú távú cél egy olyan diagnosztikai segédeszköz kifejlesztése, amely akár felül is múlhatja a BDI kérdőív alkalmazhatóságát.

Az 13. táblázat és az 8. ábra alapján nem lehet egyértelműen meghatározni, hogy melyik módszer alkalmasabb a depresszió súlyosságának előrejelzésére. Míg az akusztikus modell valamivel alacsonyabb RMSE és MAE hibaértékekkel rendelkezik ($p = 0,513$, páros t-próba), a BDI kérdőív jobb korrelációt mutat a H2B pontszámmal ($p = 0,094$, Fisher-féle r to z transzformációval), így a depresszió súlyosságának ezen tartományában jobban képes differenciálni. Az átlagos hibákat összehasonlítva elmondható, hogy míg a BDI kérdőív túlbecsüli a depresszió súlyosságát, addig az akusztikus modell alábecsüli azt.

13. táblázat: Az akusztikus modell és a BDI kérdőív pontossága (4)

	RMSE	MAE	Mean Error	Pearson Coef.	Spearman Coef.
BDI kérdőív	8.8	6.9	5.7	0.63	0.55
Akusztikus modell	8.2	6.1	-3.3	0.42	0.46



8. ábra: A BDI-pontszámok (A) és az akusztikus modell előrejelzett pontszámainak (B) összehasonlítása a H2B pontszámokkal (4)

Az is megfigyelhető, hogy az akusztikus modell esetében a legtöbb szignifikáns hiba a súlyos depresszió ($H2B > 28$) esetén fordult elő, és a maximális előrejelzett pontszám 30 körül volt. Ennek oka lehet, hogy az adatbázisban nem volt elegendő súlyos depressziós beteg. Abban a 7 esetben azonban, amikor az alanyok a BDI kérdőív alapján közepes vagy súlyos depressziós pontszámot kaptak, a beszédalapú módszer helyesen jósolta meg, hogy nem depressziósnak vagy enyhén depressziósnak minősülnek.

5. Megbeszélés

5.1. Az első vizsgálat megbeszélése

A szakirodalmat áttekintve, ez az első szisztematikus umbrella review a depresszió diagnosztikájában alkalmazott automatizált hangelemzés területén, ami így segítséget nyújthat a kutatási eredmények értékelésében, valamint a meglévő trendek azonosításában.

A bevonásra került közlemények alapján több akusztikai jellemző (pl. f0 tartomány, jitter és a shimmer) prediktív értéke alátámasztható a depresszió objektívizálható diagnosztikájának területén. Az ilyen és hasonló hangtani jellemzőkkel, a korábban csak szubjektíven értékelt depressziós beszéd, annak monotonitása, csökkent reakcióideje, automatikus mérhetővé válik, ami felgyorsítja a szűrést és fontos része lehet a megelőzésnek. A depresszióhoz társított különböző hangtani jellemzők vonatkozásában a legtöbb tanulmány eredményei egyezést mutatnak, azonban az alkalmazott prediktív modellek teljesítményeiben eltérés észlelhető, ami alapján a betanító mintaadatbázisra ugyan jól illeszkedő modellek általánosíthatósága megkérdőjelezhető. Annak érdekében, hogy a modellek eltérő környezetben is hatékonyak legyenek, függetlennek kell lenniük az személy környezetétől, életkorától, akcentusától, a háttérzajtól, a felvételi eszközöktől vagy a komorbiditástól, azonban a jelenleg rendelkezésre álló modellek esetén ez generalizálhatóság még korlátozott mértékű. (73)

Vizsgálatunknak a korábbi kutatási eredmények összefoglalása mellett célja volt az ellentmondások, következetlenségek feltárása is, hogy azok hátterében feltárhatóak legyenek azok a tényezők, amelyek segítségével konstruktív javaslatokat lehet megfogalmazni a jövőbeli kutatások számára. Előtérbe került a standardizált módszertani keretek kidolgozásának a fontossága (úm. standard adatbázisok, mintavételi technikák, depresszió szűrő mérőeszközök, előfeldolgozási eljárások, modell választás, hiperparaméterek) melyekkel az eredmények összehasonlíthatósága növelhető.

Előtérbe került a klinikai vizsgálatok általános gyakorlata, aminek részeként a hangfelvételekre többnyire ellenőrzött környezetben (rendelők, orvosi helyiségek) kerül sor, ez azonban korlátozhatja a minta nagyságát és megnehezíti az általánosíthatóságot. Az elmúlt években ezzel ellentétben már ún. „off site” adatgyűjtés is elterjedőben van, ahol okostelefonok, okosórák segítségével történik a mintavétel. Ezzel a módszerrel

természetes környezetben, sokkal nagyobb mennyiségben lehet az adatokat begyűjteni, valamint a hosszútávú követésben is segítségül szolgálhatnak egyszerű és könnyen ismételtető alkalmazásukból kifolyólag. Hátránya az „off site” adatgyűjtésnek a változó környezet, valamint eltérő hangrögzítő technológiák okozta labilis adatminőség, ami a magas zaj-jel arány miatt a modellek teljesítményét ronthatják, valamint a standardizálást is megnehezítette. Ugyan ennek a mintavételi módszernek az elterjedtsége a depresszió hangtani elemzésén alapuló detekciós eszközök kutatásában még csekély, a nehézségektől függetlenül, a téma kutatói között egyetértés van azzal kapcsolatban, hogy a természetes környezetből származó adatok elemzését érdemes fókuszba helyezni a jövőben. (3, 72, 80)

A multimodális elrendezésű vizsgálatok, ahol az akusztikai adatok mellett pl. video adatokat is használnak az elemzés során növelheti a gépi tanuló eljárásokon alapuló modellek predikciós teljesítményét. Ezzel a megközelítéssel azonban viszonylag ritkán találkozunk a kutatások során, mivel a különböző adatforrásból származó információk eltérő minőséget, valamint reprezentációt hordozhatnak magukban, ami megnehezíti az adatok egységes elemzését és értelmezését. (81) A különböző modalitások közötti eltéréseket kezelni képes technikák és algoritmusok nagyobb technológiai kihívás elé is állítják a kutatókat, ami a terület alulkutatottsága hátterében is állhat. (3, 72-74)

Az általunk bevont közlemények elemzése alapján előtérbe került az adatelemzési eszközök megválasztásának a fontossága is. Az overfitting elkerülése céljából javasolt a tanító és tesztelő készletek különválasztása, azok egymástól való függetlenségének megtartása. (3) Az akusztikai jellemzők kinyerésének (feature extraction) standardizálásával javítható a tanulmányok egymással való összehasonlíthatósága. A hiperparaméterek részletes leírása a vizsgálatok reprodukálhatóságát, valamint átláthatóságát javíthatják. (3, 73) A modellek algoritmusainak számítógépes kódjainak elérhetővé tétele a hatékonyabb finomhangolást segítheti. (73) Kisebb mintaméretű adatbázisok esetén a hold-out validációs módszer helyett a bootstrapping módszer, nagyobb mintaelemszám, összetett modellek, valamint neurális hálók esetén pedig a keresztvalidáció alkalmazása javasolt. (3, 72, 73, 75) A becslési pontosságot célszerű több mutatóval jellemzi az összehasonlíthatóság érdekében, emellett az akusztikai jellemzőkben észlelt eltéréseket, pszichiátriai tünetekkel asszociálni az interpretálhatóság érdekében. (3, 72, 75)

Mindemellett a mesterséges intelligenciát használó eljárásoknál kiemelt szerepe van az adatvédelmi szempontoknak, valamint az adatvédelmi előírások betartásának. Az „off site” adatgyűjtések esetén például adatvédelmi nehézséget okoz, hogy személyes és érzékeny adatok kerülnek begyűjtésre. Emiatt is fontos, hogy a jövőbeli kutatások már a tervezés során a hatályos adatvédelmi jogszabályokkal összhangban, világos adatkezelési irányelvek kidolgozása mellett alkossák meg a mintavételi módszertanukat, valamint lehetőség szerint dedikált platformokon keresztül (weboldalak, engedélyezett alkalmazások) gyűjtsék az adatokat, melyek segítségével növelhető az adatok transzparenciája. (3, 72, 82, 83)

5.1.1. Az első vizsgálat korlátai

Az umbrella review módszeréből kifolyólag, a vizsgálatunk eredményeinek figyelembevételénél, annak korlátait is fontos tárgyalnunk. Ugyan a bevonásra került áttekintő tanulmányok mind a PRISMA guideline-t alkalmazták, a keresési kulcsok és a keresési stratégia eltérései befolyásolhatják az eredmények összegzését, értelmezését. Fontos azt is megjegyezni, hogy csak az angol nyelven rendelkezésre álló tanulmányok kerültek a látókörünkbe, ami a más nyelveken íródott releváns tanulmányok potenciális szerepét kizárta. Az általunk készített áttekintő tanulmány, a 2024. szeptember 18-ig közzétett szisztematikus áttekintő tanulmányokat és meta-analíziseket tartalmazza a témában, az azóta publikált eredményeket nem, ami a befolyásolhatja a megállapítások relevanciáját. A vizsgálatunk kifejezetten a depresszió automatizált hangelemzésével foglalkozik, így a témához kapcsolódó, de szélesebb körű vizsgálatok kizárásra kerültek, ami korlátozhatja az átfogóképességét a tanulmánynak.

5.2. A második vizsgálat megbeszélése

A depresszió gyakoriságának világszerte észlelhető növekedésére, valamint az egészségügyi ellátórendszerekre és a globális gazdaságra gyakorolt hatására tekintettel, egyre nagyobb szükség van könnyen hozzáférhető, szaktudást nem igénylő korai diagnosztikai szűrő eszközök alkalmazására. A beszédminták elemzése egyszerű, olcsó és széles körben elérhető eszköz lehet a depresszió szűrési folyamatában, akár a háziorvosi rendelőkben. Alapvető célunk egy pontosabb, érzékenyebb akusztikai modell

létrehozása, aminek segítségével a depresszió szűrése beszédminták alapján is lehetővé válik.

Vizsgálatunkban elsősorban arra kerestük a választ, hogy egy korlátozott célváltozó milyen mértékben használható a depresszió tényleges súlyosságának becslésére vagy a depressziós és nem depressziós alanyok megkülönböztetésére. A BDI-pontszám jó leírója lehet a depresszió súlyosságának, azonban önbevallásos jellegéből kifolyólag, a tünetek lehetséges eltúlzása vagy elrejtése miatt nem tekinthető a depresszió súlyosságának teljesen objektív leíró eszközének. A BDI-pontszám meghatározása azonban gyors és kevésbé erőforrás-igényes, így könnyebb nagy beszédadatbázisokat létrehozni, ami a hatékony gépi tanulási modellek előfeltétele.

A HAMD egy klinikus által felvett interjú, ami a depresszió külső nézőpontból meghatározható tüneteit (mint például a szomatikus és viselkedési tünetek) is figyelembe veszi, emellett a tüneti változás érzékenyebb mérőeszköze, azonban befolyásolhatja a klinikus előképzettsége, valamint szubjektív megítélése is.

A kutatás során létrehoztunk egy nagyobb beszédadatbázist, ami depressziós és egészséges alanyok beszédmintáit tartalmazta. Az adatbázis tartalmazza az egyes alanyok BDI-pontszámát és az alanyok 20%-nak HAMD-pontszámát is. Ez a tény lehetőséget biztosított számunkra, hogy összehasonlítsuk a BDI és a HAMD használhatóságát a depresszió automatikus felismerésének beszédjel-feldolgozáson alapuló modelljeinek tanításához. Az összehasonlításhoz a HAMD pontszámokat a két skála súlyossági kategóriáinak végpontjain történő illesztéssel, valamint az egyes kategóriákon belül lineáris skálázással konvertáltuk át BDI pontszámokká, a konvertált HAMD-pontszámokat, H2B-ként jelölünk.

Vizsgálatunkban kimutattuk, hogy amikor az akusztikus modelleket BDI pontszámokkal képeztük, az előrejelzett pontszámok átlagosan 1,7 ponttal közelebb álltak a konvertált HAMD pontszámokhoz, mint az eredeti BDI pontszámokhoz, annak ellenére, hogy maguk a BDI pontszámok meglehetősen pontosan megközelítik a HAMD pontszámokat (MAE 6,9).

Kimutattuk, hogy a BDI pontszámok HAMD pontszámokra történő helyettesítése a tanítás során javítja a depresszió automatikus előrejelzését és felismerését. Ennek vizsgálataként a rendelkezésünkre álló 43 fős minta BDI pontszámai helyett a konvertált HAMD pontszámokat (H2B) használtuk fel. Így a regresszió esetében az RMSE hibaérték

9,9-ről 8,2-re csökkent, ami 17%-os relatív javulást jelent, míg a MAE hibaérték 7,6-ról 6,3-ra csökkent, ami szintén 17%-os relatív javulást jelent, mindez 99,9%-os szignifikancia szinten szignifikáns. Az osztályozás esetében a pontosság 81%-ról 84%-ra nőtt, ami 10%-os relatív javulást jelent a hibásan eldöntött alanyok számában.

Azt is megvizsgáltuk, hogy az akusztikus modell vagy a BDI-skála nyújt-e pontosabb előrejelzést a HAMD pontszámok tekintetében. A kísérlet során azt találtuk, hogy míg az akusztikus modell kisebb RMSE és MAE hibaértékekkel becsüli a HAMD pontszámokat (azaz a depresszió súlyosságát), mint a BDI skála, a BDI pontszámok magasabb korrelációt mutatnak a HAMD pontszámokkal, mint az előrejelzett pontszámokkal. Így nem lehet egyértelműen meghatározni, hogy melyik alkalmasabb a depresszió súlyosságának pontosabb becslésére, azonban megállapítható, hogy az akusztikus modell használhatósága a BDI skáláéhoz hasonló.

Vizsgálatunk eredményei azt mutatják, hogy az akusztikus biomarkerek alapján történő szűrés révén a depresszió diagnózisa korábban felismerhető, és így egy gépi tanuláson alapuló automatikus detekciós rendszer, széles körben alkalmazható lehet diagnosztikai eszközként a háziorvosi rendelőkből. Okostelefonokon elérhető applikációk és weboldalak segítségével, egy ilyen szűrő eszköz széleskörűen elérhetővé, és könnyen hozzáférhetővé válhat. A beszédelemzés módszerével dolgozó MI eszközök így segíthetik a depresszió korai felismerését, annak érdekében, hogy a betegek minél előbb pszichiáterhez kerüljenek a megfelelő kezelés megkezdése céljából. A depresszió korai felismerése javíthatja az életminőséget, csökkentheti a halálozást és a kórházi tartózkodás átlagos hosszát, így módon csökkentheti a gazdasági költségeket és tehermentesítheti az egészségügyi ellátórendszereket. A beszédelemzés másik alkalmazási területe lehet emellett, a depresszió diagnózissal rendelkező páciensek tüneteinek kezelés alatt történő nyomon követése és a különböző terápiák hatékonyságának összehasonlítása is.

A közeljövőben szeretnénk az általunk kifejlesztett módszert a gyakorlatban tesztelni az alapellátásban, valamint tovább bővíteni az adatbázist, hogy még pontosabb modelleket hozzunk létre, és jobban össze tudjuk hasonlítani az akusztikus modell használhatóságát a BDI skálával.

5.2.1. A második vizsgálat korlátai

A vizsgálat előzőekben tárgyalt eredményei mellett a módszerből adódó korlátokról is érdemes említést tenni. A gépi tanuló eljárásokon alapuló modellek predikciós teljesítménye nagyban függ a tanításra alkalmazott mintaadatbázis nagyságától, minél több adatunk van, annál pontosabb eszközöket tudunk létrehozni. Jelen vizsgálatunkban a HAMD értékekkel rendelkező adatbázisunk 43 alany beszédmintájából áll, ami kis mintaelemszámnak tekinthető ezen a területen, emellett viszonylag nagy eltérést mutat a csak BDI értékkel rendelkező készlet elemszámához viszonyítva ($n=175$). Az elemszám mellett, a depresszió súlyosságának megoszlásában is eltér a két adatbázisunk, a 2. készletben (HAMD értékekkel rendelkező minták) nagyobb arányban fordultak elő depressziós páciensek (83,72%) az 1. készlethez képest (41,71%).

Vizsgálatunk során, az adatgyűjtés egyszerűsítése céljából nem tettünk különbséget a depresszió hátterében álló klinikai kórképek között. Ebből kifolyólag a vizsgálat korlátának tekinthető az adatbázisunkban szereplő depressziós alanyok heterogenitása, ami befolyásolhatja az eredmények értékelhetőségét.

Depressziós páciensek vizsgálata során gyakoriak a komorbid állapotok, melyek szintén hatással lehetnek a beszédképzésre. Vizsgálatunk során ezek felmérésére, valamint külön csoportosítására a kis elemszám miatt nem térünk ki, azonban ez szintén hatással lehet az eredményeinkre.

6. Következtetések

A disszertációban bemutatott két vizsgálat célja a depresszió területén a létező beszéd alapú automatikus diagnosztikai lehetőségek áttekintése, valamint egy új, pontosabb szűrő eszköz létrehozása, a betegség mielőbbi felismerése, és ezáltal a terápia mihamarabbi elkezdése céljából.

6.1. Az első vizsgálat következtetései

A depresszió hangtani paramétereit vizsgáló, és az alapján MI alapú diagnosztikai rendszereket kidolgozó kutatások száma az elmúlt években jelentősen megnőtt, azonban az eltérő mintavételi technikák és módszertanok miatt az általánosításuk nehézségekbe ütközik.

Vizsgálatunkban a szisztematikus umbrella review módszerével tekintettük át a szakirodalmat az optimális gyakorlatok és meglévő következtetlenségek feltárása céljából, iránymutatást adva a jövő kutatásainak. A vizsgálat során kimutatásra került, hogy ugyan a depresszióhoz társított hangtani paraméterek tekintetében a legtöbb tanulmány eredményei megegyeznek, a modellek prediktív teljesítménye eltérő, aminek hátterében a nem egységes mintaadatbázisok használata, valamint az inkonzisztens módszertan állhat. A vizsgálat rávilágított a standardizált módszertani keretek kidolgozásának fontosságára, a nagyobb generalizálhatóság érdekében. Emellett előtérbe került a mintaadatok gyűjtésének növelése és könnyítése érdekében a természetes közegben, könnyen elérhető eszközök által történő mintavétel, ami ugyan technikailag nagyobb kihívást jelent, de az MI rendszerek ismertén nagy adatigényét tudná kiszolgálni. A gépi tanuló eljárás típusának megválasztása, a szükséges beállítások, valamint validációs eljárások pontos és szakszerű kiválasztása is az eredmények javítását szolgálják, ezek részletes leírása, standardizálása segítheti a további egységesítést a kutatásokban.

Feltárható volt, a jelenleg még kevés tanulmány által vizsgált, multimodális rendszerek további vizsgálatának szükségessége is, amely módszerrel tovább növelhető a prediktív teljesítmény. Mindezek mellett nem elhanyagolható szempont az adatvédelem kérdésköre, ami megköveteli, hogy a jövőbeli kutatások már a tervezés szakaszában figyelembe vegyék a fennálló adatvédelmi előírásokat és jogszabályokat.

6.2. A második vizsgálat következtetései

Eredeti vizsgálatunkban depressziós és egészséges alanyok beszédét elemeztük gépi tanuló eljárás segítségével. Az ilyen jellegű vizsgálatok estén többnyire önkitöltős kérdőívek (pl. BDI) eredményei alapján történik a mesterséges intelligencián alapuló modellek tanítása, ami azonban a túl- és alulbecslések miatt pontatlansággal járhat együtt, rontva a rendszer prediktív teljesítményét. Vizsgálatunkban egy a depresszió diagnosztikájára szolgáló, klinikus által felvett félig strukturált interjút (HAMD) is felvettünk a bevont alanyoknál, mellyel a depresszió viselkedéses és szomatikus tüneteinek értékelésén keresztül pontosabb tüneti kép értékelhető.

Kimutatható volt, hogy a BDI-vel tanított modell esetén, az automatikus osztályozó rendszer által előrejelzett pontszámok közelebb álltak a HAMD pontszámokhoz, annak ellenére, hogy az eredeti BDI pontszámok meglehetősen pontosan megközelítették a HAMD értékeket. Ez alapján elmondható, hogy a modellünk a klinikus által felvett, objektív tünetek alapján értékelt diagnosztikai értéket jobban közelíti, mint az önkitöltős teszt során szolgáltatott szubjektív eredményeket.

A HAMD pontszámokkal tanított modellünk esetén a depresszió automatikus előrejelzésének és felismerésének pontosságában 10%-os javulása volt megfigyelhető. Ugyan a vizsgálat korlátai között is részletezett a HAMD adatokkal rendelkező adatbázis kis elemszáma, valamint a depressziós súlyosságának eloszlásában észlelhető különbség befolyásolhatja az eredményeket, de mindemelett elmondható, hogy HAMD klinikus által felvett interjú alkalmazása a modell tanítása során növeli a predikciós teljesítményt. Arra is kíváncsiak voltunk, hogy az általunk megalkotott gépi tanuló eljáráson alapuló modell alkalmasabb-e a depresszió súlyosságának pontosabb becslésére, a BDI kérdőívhez képest, itt azonban vegyes eredményeket kaptunk. Megállapítható azonban, hogy az akusztikus modell használhatósága a BDI skálához hasonló. Mindezek alapján elmondható, hogy a klinikus által végzett diagnosztikai interjúk használata a gépi tanuló eljáráson alapuló rendszerek tanítása esetén javítja a pontosságot. Ennek fontosságát a depresszió egészségre, gazdasági és társadalmi helyzetre gyakorolt hatása adja, ami miatt is hasznos lenne egy gyors, egyszerű és könnyen elérhető szűrőeszköz létrehozása.

7. Összefoglalás

A depresszió az egyik leggyakoribb mentális betegség, ami jelentős negatív hatással van az életminőségre, valamint a gazdasági és társadalmi folyamatokra. A szűrése céljából jelenleg alkalmazott kérdőíves és klinika interjúk eszközök ember és időigények, valamint hozzáférhetőségük nagyon eltérő lehet. A depresszió diagnosztizálására szolgáló biomarker kutatások egyike a beszéd elemzése, amiben megfigyelhetőek hangtani jellemzők változása a betegségben érintett egyéneknél. A mesterséges intelligencia fejlődése, a gépi tanulóeljárások alkalmazásával, a pszichiátriai kórképek, azon belül a depresszió diagnosztikájában is új távlatokat nyit. A gépi tanuláson alapuló, akusztikai diagnosztikai rendszerek, jó prediktív teljesítménnyel képesek a beszéd alapján elkülöníteni a depressziót. A legtöbb ilyen irányú kutatás a mintavétel során, inherensen nagyobb szubjektivitású önkitöltős kérdőívek alapján méri fel a depressziós állapotot, amit a mesterséges intelligencia alapú modellek betanítására használnak. Az **első vizsgálatunkban** a témában elérhető szakirodalomban végeztünk átfogó áttekintő tanulmányt, az összefüggések és következtetések feltárása céljából. Eredményeink alapján kimutatható, hogy a meglévő kutatások a depresszióhoz társított hangtani paraméterek tekintetében megegyeznek, azonban az eltérő módszertan miatt predikciós teljesítményük változó. A megfigyelt inkonzisztenciák alapján javaslatokat tettünk a kutatások egységesítésére szolgáló irányelvekre, melyek segítségével könnyebbé válhat az eredmények összehasonlíthatósága is. A **második vizsgálatunkban** egy klinikus által felvett interjú eredményeivel tanítottuk az akusztikai modellünket, így eredményeink alapján 10%-os javulás volt megfigyelhető a prediktív pontosságban az önkitöltős kérdőíven tanított rendszerhez képest. Megállapításra került az is, hogy az akusztikus modellünk használhatósága az általunk használt önkitöltős skálához hasonló. A vizsgálat fontos korlátai a klinikai interjú eredményét is tartalmazó mintánk kisebb elemszáma, valamint az esetlegesen fennálló komorbid állapotok beszédet befolyásoló hatásának szerepe. **Összefoglalva** a disszertációban bemutatott vizsgálataink alapján, a depresszió hangtani jellemzői alapján, gépi tanulóeljárások segítségével, létrehozhatóak nagy prediktív teljesítménnyel rendelkező akusztikai modellek, melyek segíthetik a diagnózis alkotás folyamatát, ezáltal felhasználhatóak lehetnek a betegellátás területén.

Summary

Depression is one of the most common mental disorder with a significant negative impact on quality of life, economic and social trends. The self-reported questionnaires and clinical interviews currently used for screening are manpower-intensive and time-consuming methods and their availability may vary. One of the biomarker studies for the diagnosis of depression is the analysis of speech, in which changes in phonological characteristics of individuals affected by the disorder can be observed. Advances in artificial intelligence, with the use of machine learning, are opening new horizons in the diagnosis of psychiatric disorders, including depression. Acoustic diagnostic systems based on machine learning can differentiate depression on the basis of speech with good predictive performance. Most research on this subject measure depression based on self-reported questionnaires, with inherently higher subjectivity, for training AI-based models. In our **first study**, we conducted a systematic umbrella review of the available literature on the topic to identify correlations and inconsistencies. Our results show that existing research is consistent with respect to phonological parameters associated with depression, but their predictive power varies due to different methodologies. Based on the observed inconsistencies, we proposed guidelines to standardize the research, which should also facilitate the comparability of results. In our **second study**, we trained our acoustic model with the results of a clinical interview recorded by a specialist, and our results showed a 10% improvement in predictive accuracy compared to a system trained with a self-reported questionnaire. It was also found that the usability of our acoustic model was similar to the scale of the self-reported questionnaire we used. Important limitations of the study are the smaller sample size of our dataset including clinical interview results and the role of the impact of any comorbid conditions on speech. **In conclusion**, based on our studies presented in this dissertation, acoustic models with high predictive performance can be generated based on the phonological features of depression using machine learning methods, which can aid the diagnosis process and thus be used in patient care.

8. Irodalomjegyzék

1. Strawbridge R, Young AH, Cleare AJ. Biomarkers for depression: recent insights, current challenges and future prospects. *Neuropsychiatric Disease and Treatment*. 2017;13(null):1245-1262.
2. Dev A, Roy N, Islam MK, Biswas C, Ahmed HU, Amin MA, Sarker F, Vaidyanathan R, Mamun KA. Exploration of EEG-Based Depression Biomarkers Identification Techniques and Their Applications: A Systematic Review. *IEEE Access*. 2022;10:16756-16781.
3. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*. 2020;5(1):96-116.
4. Hajduska-Dér B, Kiss G, Sztahó D, Vicsi K, Simon L. The applicability of the Beck Depression Inventory and Hamilton Depression Scale in the automatic recognition of depression based on speech signal processing. *Frontiers in Psychiatry*. 2022;13.
5. Hajduska-Dér B, Simon L, Réthelyi J, Haluska-Vass E. Automatizált hangelemzés és depresszió: Szisztematikus umbrella review. *Ideggyógyászati Szemle*. 2025;78(03-04):87-99.
6. Horwitz AV, Wakefield JC, Lorenzo-Luaces L. History of Depression. In: DeRubeis RJ, Strunk DR, editors. *The Oxford Handbook of Mood Disorders*: Oxford University Press; 2017. p. 11-23.
7. American Psychiatric Association. Depressive disorders. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC2013.
8. WHO. Depression Fact Sheet Geneva, Switzerland: WHO; 2020 [Available from: <https://www.who.int/news-room/fact-sheets/detail/depression>].
9. WHO. The global burden of disease: 2004 update. 2008.
10. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, de Graaf R, Demyttenaere K, Hu C, Iwata N, Karam AN, Kaur J, Kostyuchenko S, Lépine J-P, Levinson D, Matschinger H, Mora MEM, Browne MO, Posada-Villa J, Viana MC, Williams DR, Kessler RC. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine*. 2011;9(1):90.

11. Torzsa P, Rihmer Z, Gonda X, Sebestyén B, Szokontor N, Kalabay L. A depresszió prevalenciája az alapellátásban Magyarországon. *Neuropsychopharmacologia Hungarica*. 2008;10.(5. sz.):265-270.
12. Arias-de la Torre J, Vilagut G, Ronaldson A, Serrano-Blanco A, Martín V, Peters M, Valderas JM, Dregan A, Alonso J. Prevalence and variability of current depressive disorder in 27 European countries: a population-based study. *The Lancet Public Health*. 2021;6(10):e729-e738.
13. Albert PR. Why is depression more prevalent in women? *J Psychiatry Neurosci*. 2015;40(4):219-221.
14. Orsolini L, Latini R, Pompili M, Serafini G, Volpe U, Vellante F, Fornaro M, Valchera A, Tomasetti C, Fraticelli S, Alessandrini M, La Rovere R, Trotta S, Martinotti G, Di Giannantonio M, De Berardis D. Understanding the Complex of Suicide in Depression: from Research to Clinics. *Psychiatry Investig*. 2020;17(3):207-221.
15. Li W, Ruan W, Peng Y, Lu Z, Wang D. Associations of socioeconomic status and sleep disorder with depression among US adults. *Journal of Affective Disorders*. 2021;295:21-27.
16. Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry*. 2015;76(2):155-162.
17. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789-1858.
18. Young JJ, Bruno D, Pomara N. A review of the relationship between proinflammatory cytokines and major depressive disorder. *Journal of Affective Disorders*. 2014;169:15-20.
19. Amodeo G, Trusso M, Fagiolini A. Depression and Inflammation: Disentangling a Clear Yet Complex and Multifaceted Link. *Neuropsychiatry*. 2017;07.
20. Rana T, Behl T, Sehgal A, Srivastava P, Bungau S. Unfolding the Role of BDNF as a Biomarker for Treatment of Depression. *Journal of Molecular Neuroscience*. 2021;71(10):2008-2021.

21. Mikulska J, Juszczak G, Gawrońska-Grzywacz M, Herbet M. HPA Axis in the Pathomechanism of Depression and Schizophrenia: New Therapeutic Strategies Based on Its Participation. *Brain Sciences*. 2021;11(10):1298.
22. Yao Z, Fu Y, Wu J, Zhang W, Yu Y, Zhang Z, Wu X, Wang Y, Hu B. Morphological changes in subregions of hippocampus and amygdala in major depressive disorder patients. *Brain Imaging and Behavior*. 2020;14(3):653-667.
23. Pilmeyer J, Huijbers W, Lamerichs R, Jansen JFA, Breeuwer M, Zinger S. Functional MRI in major depressive disorder: A review of findings, limitations, and future prospects. *Journal of Neuroimaging*. 2022;32(4):582-595.
24. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *American journal of psychiatry*. 2000;157(10):1552-1562.
25. Shadrina M, Bondarenko EA, Slominsky PA. Genetics Factors in Major Depression Disease. *Frontiers in Psychiatry*. 2018;9.
26. Park C, Rosenblat JD, Brietzke E, Pan Z, Lee Y, Cao B, Zuckerman H, Kalantarova A, McIntyre RS. Stress, epigenetics and depression: A systematic review. *Neuroscience & Biobehavioral Reviews*. 2019;102:139-152.
27. Baglioni C, Nanovska S, Regen W, Spiegelhalter K, Feige B, Nissen C, Reynolds CF, Riemann D. Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychol Bull*. 2016;142(9):969-990.
28. Koch C, Wilhelm M, Salzmann S, Rief W, Euteneuer F. A meta-analysis of heart rate variability in major depression. *Psychol Med*. 2019;49(12):1948-1957.
29. Kraepelin E. Manic-depressive insanity and paranoia: E. & S. Livingstone; 1921.
30. Newman SS, Mather VG. Analysis of spoken language of patients with affective disorders. *American Journal of Psychiatry*. 1938;94:913-942.
31. Moses PJ. The Voice of Neurosis: Grune & Stratton; 1954.
32. Darby JK, Hollien H. Vocal and speech patterns of depressive patients. *Folia Phoniatri (Basel)*. 1977;29(4):279-291.
33. Askenfelt A. Voice analysis in depressed patients: Rate of change of fundamental frequency related to mental state. 1980.
34. Nilsson A. Speech characteristics as indicators of depressive illness. *Acta Psychiatr Scand*. 1988;77(3):253-263.

35. Stassen HH, Kuny S, Hell D. The speech analysis approach to determining onset of improvement under antidepressants. *European Neuropsychopharmacology*. 1998;8(4):303-310.
36. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological Psychiatry*. 2012;72(7):580-587.
37. Wernigg R, Hajduska-Dér B. A mesterséges intelligencia szerepe a pszichiátriában. *Psychiatria Hungarica*. 2024;39(1).
38. Mahesh B. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*[Internet]. 2020;9(1):381-386.
39. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.
40. Chowdhary KR. Natural Language Processing. In: Chowdhary KR, editor. *Fundamentals of Artificial Intelligence*. New Delhi: Springer India; 2020. p. 603-649.
41. Abd-Alrazaq A, Alhuwail D, Schneider J, Toro CT, Ahmed A, Alzubaidi M, Alajlani M, Househ M. The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review. *NPJ Digit Med*. 2022;5(1):87.
42. Tutun S, Johnson ME, Ahmed A, Albizri A, Irgil S, Yesilkaya I, Ucar EN, Sengun T, Harfouche A. An AI-based Decision Support System for Predicting Mental Health Disorders. *Inf Syst Front*. 2023;25(3):1261-1276.
43. Passos IC, Mwangi B, Cao B, Hamilton JE, Wu M-J, Zhang XY, Zunta-Soares GB, Quevedo J, Kauer-Sant'Anna M, Kapczinski F. Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. *Journal of affective disorders*. 2016;193:109-116.
44. de Ávila Berni G, Rabelo-da-Ponte FD, Librenza-Garcia D, V. Boeira M, Kauer-Sant'Anna M, Cavalcante Passos I, Kapczinski F. Potential use of text classification tools as signatures of suicidal behavior: A proof-of-concept study using Virginia Woolf's personal writings. *PloS one*. 2018;13(10):e0204820.
45. Pestian JP, Sorter M, Connolly B, Bretonnel Cohen K, McCullumsmith C, Gee JT, Morency LP, Scherer S, Rohlfis L, Group SR. A machine learning approach to identifying the thought markers of suicidal subjects: a prospective multicenter trial. *Suicide and Life-Threatening Behavior*. 2017;47(1):112-121.

46. Sinha PP, Mishra R, Sawhney R, Mahata D, Shah RR, Liu H, editors. # suicidal-A multipronged approach to identify and explore suicidal ideation in twitter. Proceedings of the 28th ACM international conference on information and knowledge management; 2019.
47. Cortes-Briones JA, Tapia-Rivas NI, D'Souza DC, Estevez PA. Going deep into schizophrenia with artificial intelligence. *Schizophrenia Research*. 2022;245:122-140.
48. Chen ZS, Kulkarni PP, Galatzer-Levy IR, Bigio B, Nasca C, Zhang Y. Modern views of machine learning for precision psychiatry. *Patterns (N Y)*. 2022;3(11):100602.
49. Sheu Y-h, Magdamo C, Miller M, Das S, Blacker D, Smoller JW. AI-assisted prediction of differential response to antidepressant classes using electronic health records. *NPJ Digital Medicine*. 2023;6(1):73.
50. Hsu C-W, Tsai S-Y, Wang L-J, Liang C-S, Carvalho AF, Solmi M, Vieta E, Lin P-Y, Hu C-A, Kao H-Y. Predicting serum levels of lithium-treated patients: A supervised machine learning approach. *Biomedicine*. 2021;9(11):1558.
51. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Communication*. 2015;71:10-49.
52. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*. 2010;29(1):24-54.
53. Pampouchidou A, Simos PG, Marias K, Meriaudeau F, Yang F, Pediaditis M, Tsiknakis M. Automatic Assessment of Depression Based on Visual Cues: A Systematic Review. *IEEE Transactions on Affective Computing*. 2019;10(4):445-470.
54. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, Mohr DC. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *J Med Internet Res*. 2015;17(7):e175.
55. Chen X, Sykora MD, Jackson TW, Elayan S, editors. What about mood swings: Identifying depression on twitter with temporal measures of emotions. Companion proceedings of the the web conference 2018; 2018.
56. Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. *Health information science and systems*. 2018;6:1-12.

57. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, Cannon TD, Krystal JH, Corlett PR. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry*. 2016;3(3):243-250.
58. Scherer S, Stratou G, Gratch J, Morency LP. Investigating Voice Quality as a Speaker-Independent Indicator of Depression and PTSD. In: Bimbot F, Cerisara C, Fougerson C, Gravier G, Lamel L, Pellegrino F, et al., editors. 14th Annual Conference of the International Speech Communication Association. *Interspeech2013*. p. 847-851.
59. Taddeo M, Floridi L. How AI can be a force for good. *Science*. 2018;361:751–752.
60. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453.
61. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;3(11):e745-e750.
62. Price I, Nicholson W. Artificial intelligence in health care: Applications and legal issues. 2017.
63. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *Jama*. 2017;318(6):517-518.
64. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008.
65. de Boer JN, Voppel AE, Brederoo SG, Wijnen FNK, Sommer IEC. Language disturbances in schizophrenia: the relation with antipsychotic medication. *NPJ Schizophr*. 2020;6(1):24.
66. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment*. 1996;67(3):588-597.
67. Schneibel R, Brakemeier EL, Wilbertz G, Dykieriek P, Zobel I, Schramm E. Sensitivity to detect change and the correlation of clinical factors with the Hamilton Depression Rating Scale and the Beck Depression Inventory in depressed inpatients. *Psychiatry Res*. 2012;198(1):62-67.

68. Kiss G, Vicsi K, editors. Comparison of read and spontaneous speech in case of automatic detection of depression. 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom); 2017: IEEE.
69. Kiss G, Vicsi K. Mono-and multi-lingual depression prediction based on speech processing. *International Journal of Speech Technology*. 2017;20(4):919-935.
70. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*. 2011;2(3):1-27.
71. Cummins N, Vlasenko B, Sagha H, Schuller B, editors. Enhancing speech-based depression detection through gender dependent vowel-level formant features. *Conference on artificial intelligence in medicine in Europe*; 2017: Springer.
72. Hecker P, Steckhan N, Eyben F, Schuller BW, Arnrich B. Voice Analysis for Neurological Disorder Recognition-A Systematic Review and Perspective on Emerging Trends. *Front Digit Health*. 2022;4:842301.
73. Liu L, Liu L, Wafa HA, Tydeman F, Xie W, Wang Y. Diagnostic accuracy of deep learning using speech samples in depression: a systematic review and meta-analysis. *J Am Med Inform Assoc*. 2024;31(10):2394-2404.
74. Maatoug R, Oudin A, Adrien V, Saudreau B, Bonnot O, Millet B, Ferreri F, Mouchabac S, Bourla A. Digital phenotype of mood disorders: A conceptual and critical review. *Front Psychiatry*. 2022;13:895860.
75. Mao K, Wu Y, Chen J. A systematic review on automated clinical depression diagnosis. *npj Mental Health Research*. 2023;2(1):20.
76. Otero-González I, Pacheco-Lorenzo MR, Fernández-Iglesias MJ, Anido-Rifón LE. Conversational agents for depression screening: A systematic review. *International Journal of Medical Informatics*. 2024;181:105272.
77. Rahaman MJ. A Comprehensive Review to Understand the Definitions, Advantages, Disadvantages and Applications of Machine Learning Algorithms. *International Journal of Computer Applications*. 2024;186:43-47.
78. Krichen M. Convolutional Neural Networks: A Survey. *Computers*. 2023;12(8):151.
79. GeeksforGeeks. What are the Advantages and Disadvantages of Random Forest? 2024 [updated 2024.02.15. Available from: <https://www.geeksforgeeks.org/what-are-the-advantages-and-disadvantages-of-random-forest/>.

80. Ermers NJ, Hagoort K, Scheepers FE. The Predictive Validity of Machine Learning Models in the Classification and Treatment of Major Depressive Disorder: State of the Art and Future Directions. *Frontiers in Psychiatry*. 2020;11.
81. Yang L, Jiang D, Xia X, Pei E, Oveneke MC, Sahli H. Multimodal Measurement of Depression Using Deep Learning Models. *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*; Mountain View, California, USA: Association for Computing Machinery; 2017. p. 53–59.
82. Karizat N, Vinson AH, Parthasarathy S, Andalibi N. Patent Applications as Glimpses into the Sociotechnical Imaginary: Ethical Speculation on the Imagined Futures of Emotion AI for Mental Health Monitoring and Detection. *Proc ACM Hum-Comput Interact*. 2024;8(CSCW1):Article 106.
83. Vitale F, Carbonaro B, Cordasco G, Esposito A, Marrone S, Raimo G, Verde L. A Privacy-Oriented Approach for Depression Signs Detection Based on Speech Analysis. *Electronics*. 2021;10(23):2986.

9. Saját publikációk jegyzéke

9.1. A disszertációhoz tartozó közlemények

Hajduska-Dér B, Kiss G, Sztahó D, Vicsi K, Simon L. The applicability of the Beck Depression Inventory and Hamilton Depression Scale in the automatic recognition of depression based on speech signal processing. *Frontiers in Psychiatry*. 2022;13.

Hajduska-Dér B, Simon L, Réthelyi J, Haluska-Vass E. Automatizált hangelemzés és depresszió: Szisztematikus umbrella review. *Ideggyógyászati Szemle*. 2025;78(03-04):87-99.

9.2. A disszertációtól független közlemények

Wernigg R, **Hajduska-Dér B**. A mesterséges intelligencia szerepe a pszichiátriában. *Psychiatria Hungarica*. 2024;39(1).

Sztahó D, Kiss G, Tulics MG, **Hajduska-Dér B**, Vicsi K, editors. Automatic discrimination of several types of speech pathologies. 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD); 2019 10-12 Oct. 2019.

S. Nagy Z, Salgó E, Bajzát B, **Hajduska-Dér B**, Unoka ZS. Reliability and validity of the Hungarian version of the Personality Inventory for DSM-5 (PID-5). *PLOS ONE*. 2022;17(7):e0266201.

Kenézli E, Rákár-Szabó N, Szabó D, Lévy E, Szabó G, Szegő A, **Hajduska-Dér B**, Seres I, Császár I, Hordósi J, Zseli K, Réthelyi J, Unoka ZS. Nature-Adventure based experiential methods for enhancing psychotherapeutic efficacy. *European Psychiatry*. 2024;67(S1):S51-S52.

Hajduska-Dér B, Egervári L. Evészavarok. In: Réthelyi J, editor. *Pszichiátria jegyzet orvostanhallgatóknak*. Budapest: Oriold és Társai Kiadó; 2020. p. 174-84.

Hajduska-Dér B, Simon L. Alapvető pszichoterápiás beavatkozások és pszichoterápiás módszerek. In: Réthelyi J, editor. Pszichiátria jegyzet orvostanhallgatóknak. Budapest: Oriold és Társai Kiadó; 2020. p. 300-8.

10. Köszönetnyilvánítás

Szeretném kifejezni köszönetemet témavezetőmnek, Dr. Simon Lajosnak szakmai támogatásáért és segítségéért.

Köszönöm Prof. Dr. Réthelyi Jánosnak, a Pszichiátriai és Pszichiátriai Klinika igazgatójának, hogy támogatta a kutatást.

Köszönöm a segítséget a kutatásban közreműködő kollégáimnak: dr. Kiss Gábornak, dr. Sztahó Dávidnak és Prof. dr. Vicsi Klárának.

Külön köszönöm dr. Haluska-Vass Editnek a sok támogatást és útmutatást.

Valamennyi önkéntes résztvevőnek köszönöm, hogy vizsgálatunkban részt vettek.

Táblázatok és ábrák jegyzéke

Táblázatok

1. TÁBLÁZAT: A DEPRESSZIÓBAN VÁLTOZÁST MUTATÓ FŐBB HANGTANI PARAMÉTEREK ÉS RÖVID LEÍRÁSUK (28)	15
2. TÁBLÁZAT: A BDI ÁLTAL DEFINIÁLT KATEGÓRIÁK ÉS HOZZÁJUK TARTOZÓ ÉRTÉKEK. 34	
3. TÁBLÁZAT: A HAMD ÁLTAL DEFINIÁLT KATEGÓRIÁK ÉS HOZZÁJUK TARTOZÓ ÉRTÉKEK	35
4. TÁBLÁZAT: AZ ELEMZÉSBEN HASZNÁLT ADATBÁZIS LEGFONTOSABB JELLEMZŐI (4) ..	36
5. TÁBLÁZAT: AZ LLD-KBŐL KISZÁMÍTOTT LEÍRÓ JELLEMZŐK (4).....	38
6. TÁBLÁZAT: A BEVONT KÖZLEMÉNYEK ELEMZÉSE AMSTAR 2 ÉRTÉKELŐRENDSZERREL (5).....	42
7. TÁBLÁZAT: A BEVONÁSRA KERÜLT TANULMÁNYOK FŐBB JELLEMZŐI (5)	43
8. TÁBLÁZAT: DEPRESSZIÓVAL ÖSSZEFÜGGÉSBEN VIZSGÁLT AKUSZTIKAI JELLEMZŐK LISTÁJA ÉS AZOK RÖVID MEGHATÁROZÁSA (5)	46
9. TÁBLÁZAT: GÉPI TANULÓ ELJÁRÁSOK ELŐNYEI, HÁTRÁNYAI (77-79).....	47
10. TÁBLÁZAT: A DEPRESSZIÓ ELŐREJELZÉSÉNEK PONTOSSÁGA BDI-PONTSZÁMOK FELHASZNÁLÁSÁVAL A TANÍTÁS SORÁN (4).....	52
11. TÁBLÁZAT: A DEPRESSZIÓ ELŐREJELZÉSÉNEK PONTOSSÁGA H2B-PONTSZÁMOK FELHASZNÁLÁSÁVAL A TANÍTÁS SORÁN (4).....	56
12. TÁBLÁZAT: A DEPRESSZIÓS ÉS EGÉSZSÉGES ALANYOK OSZTÁLYOZÁSÁNAK PONTOSSÁGA, H2B VAGY BDI PONTSZÁMOK ÁLTAL TÖRTÉNŐ BETANÍTÁS ESETÉN (4)	58
13. TÁBLÁZAT: AZ AKUSZTIKUS MODELL ÉS A BDI KÉRDŐÍV PONTOSSÁGA (4).....	59

Ábrák

1. ÁBRA: A DEPRESSZIÓS BESZÉDADATBÁZIS ELEMSZÁMAI A H2B ÁLTALI SÚLYOSSÁGI KATEGÓRIÁK SZERINT AZ ÖSSZES, SET I ÉS SET II. CSOPORTOSÍTÁS TEKINTETÉBEN (4)	36
2. ÁBRA: PRISMA IRÁNYELVEK ALAPJÁN VÉGZETT IRODALOMKUTATÁS FOLYAMATÁBRÁJA (5).....	41

3. ÁBRA: A LEGGYAKRABBAN VIZSGÁLT AKUSZTIKAI JELLEMZŐK ÉS A DEPRESSZIÓ ÖSSZEFÜGGÉSEINEK GRAFIKUS ÖSSZEFOGLALÁSA (5)	49
4. ÁBRA: A DEPRESSZIÓ ELŐREJELZÉSÉNEK PONTOSSÁGA, BDI-PONTSZÁMOK FELHASZNÁLÁSÁVAL A TANÍTÁS SORÁN (FP: FALSE POSITIVE, FN: FALSE NEGATIVE, TP: TRUE POSITIVE, TN: TRUE NEGATIVE) (4).....	53
5. ÁBRA: AZ ABSZOLÚT HIBA KÜLÖNBSÉGEINEK HISZTOGRAMJA A BDI ÉS H2B ESETÉBEN A MÁSODIK KÉSZLETEN (4)	54
6. ÁBRA: A DEPRESSZIÓ ELŐREJELZÉSÉNEK PONTOSSÁGA, H2B-PONTSZÁMOK FELHASZNÁLÁSÁVAL A TANÍTÁS SORÁN (4).....	56
7. ÁBRA: AZ AKUSZTIKUS MODELLEK ROC-GÖRBÉJE, AMIKOR A MODELLEKET BDI-VEL (BALRA) ÉS H2B-PONTSZÁMOKKAL (JOBBRA) TANÍTOTTUK (4)	58
8. ÁBRA: A BDI-PONTSZÁMOK (A) ÉS AZ AKUSZTIKUS MODELL ELŐREJELZETT PONTSZÁMAINAK (B) ÖSSZEHASONLÍTÁSA A H2B PONTSZÁMOKKAL (4)	59



OPEN ACCESS

EDITED BY

Lucas Drucaroff,
Consejo Nacional de Investigaciones
Científicas y Técnicas (CONICET),
Argentina

REVIEWED BY

András Horváth,
Pázmány Péter Catholic University,
Hungary
Mariusz Stanisław Wiglusz,
Medical University of Gdańsk, Poland
Andrzej Rynkiewicz,
University of Warsaw, Poland

*CORRESPONDENCE

Bálint Hajduska-Dér
hajduska-der.balint@semmelweis.hu

SPECIALTY SECTION

This article was submitted to
Digital Mental Health,
a section of the journal
Frontiers in Psychiatry

RECEIVED 20 February 2022

ACCEPTED 18 July 2022

PUBLISHED 04 August 2022

CITATION

Hajduska-Dér B, Kiss G, Sztahó D,
Vicsi K and Simon L (2022) The
applicability of the Beck Depression
Inventory and Hamilton Depression
Scale in the automatic recognition
of depression based on speech signal
processing.
Front. Psychiatry 13:879896.
doi: 10.3389/fpsy.2022.879896

COPYRIGHT

© 2022 Hajduska-Dér, Kiss, Sztahó,
Vicsi and Simon. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

The applicability of the Beck Depression Inventory and Hamilton Depression Scale in the automatic recognition of depression based on speech signal processing

Bálint Hajduska-Dér^{1*}, Gábor Kiss², Dávid Sztahó²,
Klára Vicsi² and Lajos Simon¹

¹Department of Psychiatry and Psychotherapy, Semmelweis University, Budapest, Hungary,

²Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary

Depression is a growing problem worldwide, impacting on an increasing number of patients, and also affecting health systems and the global economy. The most common diagnostic rating scales of depression are self-reported or clinician-administered, which differ in the symptoms that they are sampling. Speech is a promising biomarker in the diagnostic assessment of depression, due to non-invasiveness and cost and time efficiency. In our study, we try to achieve a more accurate, sensitive model for determining depression based on speech processing. Regression and classification models were also developed using a machine learning method. During the research, we had access to a large speech database that includes speech samples from depressed and healthy subjects. The database contains the Beck Depression Inventory (BDI) score of each subject and the Hamilton Rating Scale for Depression (HAM-D) score of 20% of the subjects. This fact provided an opportunity to compare the usefulness of BDI and HAM-D for training models of automatic recognition of depression based on speech signal processing. We found that the estimated values of the acoustic model trained on BDI scores are closer to HAM-D assessment than to the BDI scores, and the partial application of HAM-D scores instead of BDI scores in training improves the accuracy of automatic recognition of depression.

KEYWORDS

depression, machine learning, diagnosis, speech, Support Vector Regression

Introduction

The growing prevalence of depression is a severe problem in our society, which does not only affect the individual, but it is also burdening health systems and the global economy. WHO studies show that 264 million people are affected by depression today (1). Projections show that depression will be the first cause of burden of disease worldwide by 2030 (2).

Depression is often difficult to diagnose due to the variety of symptoms and the stigma on mental illnesses (3). The most common assessment methods in depression are the Hamilton Rating Scale for Depression (HAMD) and the Beck Depression Inventory (BDI). While the HAMD is a clinician-administered assessment, the evaluation by this instrument can vary depending on the expertise of the clinician, BDI is a self-reported rating scale and patients can exaggerate or conceal symptoms (4).

In the past few decades, several studies attempted to identify biological markers, which could improve the diagnosis and classification of psychiatric disorders (5). Altered levels of serum growth factors and pro-inflammatory cytokines, changes in endocrine factors and metabolic markers can be associated with depression but limited by a lack of sensitivity and specificity (6). Non-invasive EEG biomarkers are also under examination, such as gamma-band power and signal complexity (7).

Speech is a promising biomarker for automatic recognition of depression (8–10). The advantage of speech signal processing is that it is non-invasive and time and cost efficient. Several studies have shown that it is possible to automatically detect depression (11–15) and estimate the severity of depression (9, 16–20) based on speech signal processing.

To support the automatic recognition of depression, it is possible to separate depressed and non-depressed subjects by classification, or to estimate the severity of the subject's depression using a regression method. Based on literature, it is possible to distinguish between healthy subjects and depressed patients with an accuracy of 60–95%. However, it is difficult to compare the results in the literature in terms of both classification and regression, because the results obtained are influenced by the number of subjects in the examined database (unfortunately, the examined database is generally not public) and the distribution of their severity according to depression. In addition, the possible differences in the scale used to describe the severity of depression (HAMD, BDI, Patient Health Questionnaire 9 – PHQ-9, Montgomery-Åsberg Depression Rating Scale – MADRS) in each experiment further complicate the comparison of regression results. The currently known depressed speech databases are small: the median of the size of the databases (depressed and healthy together) is 123 (10).

Low et al. (10) comprehensive study highlights that in the majority of current studies (62%), self-reported questionnaires were used to determine the severity of depression of examined subjects. The use of self-reported questionnaires is completely

understandable, as they allow more data to be obtained more quickly, and a large database is a prerequisite for creating precise machine learning models. Clinician-administered scales, like HAMD and MADRS can assess different symptoms of depression, MADRS can be more sensitive to change in symptoms, but both need more time and professional personnel (21). With the use of self-reported rating scales, the exaggeration and concealment of symptoms as a noise is introduced into the target feature, which degrades the performance of the trained model. The problem of incorrect targets is further exacerbated by the fact that during model optimization, the evaluation of individual models will also be incorrect, which can result in low efficient models, with poor generalizability, due to small size of databases.

The present study examined the extent to which rating scale (HAMD) improves the accuracy of depression estimation. Previous studies show that the correlation between pre-treatment BDI and HAMD scores in depression is between $r = 0.4$ and $r = 0.7$ (22). Several other studies show an increase in the correlation of the rating scales in longitudinal studies, the correlations between observer-rated and self-report scales are 0.4 at the beginning and 0.7 at the end of the study (23). The modest association at the first assessment can be related to the difference between the symptoms that the scales are sampling, HAMD focuses on the somatic and behavioral symptoms, whereas BDI underlines subjective symptoms (24). Longitudinal studies show that the HAMD is a more sensitive measure of symptom change than BDI (25).

In the present research, we sought to answer two questions: To what extent can a model taught with BDI accurately learn the severity of depression? To what extent is it possible to use BDI and HAMD scores together and does using HAMD instead of BDI for some of the training samples improve model performance? Answering the first question is important because the BDI questionnaire is self-reported, so it may have limits in the operation of the model. The answer to the second question is important from a practical point of view. On the one hand, if there are two depression databases, but the severity of depression was measured on a different scale, it is necessary to transform them into a common scale. On the other hand, it is faster to collect a database based on BDI, but at the same time the HAMD scores assess different symptoms of depression, so maybe it would be possible to improve the performance of the acoustic model by adding and using some samples labeled with HAMD scores.

Materials and methods

Database

In the present research we used the Hungarian Depressed Speech Database (DEPISDA) (19). The database currently

contains speech samples of 218 (144 females and 74 males) Hungarian subjects (depressed and healthy subjects). The collection of speech samples and the selection of subjects are carried out jointly by the Laboratory of Speech Acoustics, Budapest University of Technology and Economics and the Department of Psychiatry and Psychotherapy, Semmelweis University. The speech samples are independent, all patients were recorded only once.

Study inclusion criteria for patients with depression required study participants to be at least 18 years old and BDI score of 14 or higher. Exclusion criteria was ongoing antipsychotic medication use because extrapyramidal symptoms can negatively affect the articulatory system, and conceptualizing and formulating are known to be negatively affected by antipsychotic drugs that block dopamine receptors (26).

Read speech samples were recorded from each subject. Each speaker had to read a tale of about 10 sentences (“The North Wind and the Sun”). BDI score, age, gender, smoking habits, medications taken and other speech-related illnesses and conditions of the subjects were also noted in each case. HAMD score of examined subjects was recorded in 20% of cases, in a total of 43 speakers. Nearly half of the subjects in the database are healthy, while depressed patients cover different degrees of severity of depression almost evenly. In order to achieve our research goals, we divided the database into two sets. *Set I* included samples where only BDI scores were available. *Set II* included samples where both BDI and HAMD scores were available.

To compare HAMD scores and BDI scores, the endpoints of the Hamilton scale categories (minimum 0–7, mild 8–13, moderate 14–18, severe 19–48) were fitted to the endpoints of BDI categories (minimum 0–13, mild 14–19, moderate 20–28, severe 29–63), and then linear scaling was used within each category, since the relationship is presumably linear within each scale.

The rescaled HAMD scores are hereinafter referred to as H2B. An H2B score was then assigned to each sample, which, if a HAMD score was available, was obtained by scaling it, otherwise the BDI score was assigned. Thus, both BDI and H2B scores are available for each sample in each set during further experiments. The main descriptive statistics of the subjects are shown in [Table 1](#), and the distribution of the subjects by depression is shown in [Figure 1](#).

Preprocessing

Speech samples were then normalized to peak amplitude to eliminate any different recording gains. Annotation of speech samples was created using a transcription based force alignment automatic segmentation method (27)

valid for several languages. Segmentation and annotation of speech and silence parts were performed, as well as phoneme-level segmentation of speech parts using the SAMPA alphabet. Phoneme-level segmentation is important because calculating certain characteristics at phoneme level improves the accuracy of automatic recognition of depression (9, 28).

The following low-level descriptors (LLDs) were calculated: intensity, fundamental frequency, jitter, shimmer, first and second formant frequencies and their bandwidths and 13 mfcc coefficients. Low-level features were calculated with a 30 ms Hamming window length and with 10 ms timestep. The fundamental frequency, jitter, shimmer, and formant frequencies were calculated from the voiced parts of the speech sample, while the other LLDs were calculated over the entire speech sample. These LLDs were calculated because previous studies have shown that these speech characteristics alter as a result of depression (10, 15).

Feature extraction

From the LLDs, the actual descriptive features were calculated using the mean, standard deviation and percentile ranges (range formed after leaving the lower and upper 1, 5, 15, and 25% of the ordered values). Some descriptive characteristics were calculated based on more place-dependent methods, from vowel E (Sampa notation for Hungarian “e”) and vowel O (Sampa notation for Hungarian “a”), from all vowels, and over the whole speech part ([Table 2](#)), so a total of 282 features were calculated from LLDs.

Additional features were also calculated, such as articulation rate, pause ratio and the ratio of transient parts on speech parts and on total speech sample. Thus, a total of 286 features were calculated from each speech sample ([Table 2](#)).

Model training and optimization

Support Vector Regression (SVR) (29) was selected as the machine learning method, which is a version of the Support Vector Machines (SVM) family for solving a regression problem (30, 31). We selected this because several studies have shown that using SVMs can give good results in the automatic recognition of depression (8–10). However, their use is not necessary, similar results can be obtained with other machine learning methods (18, 19, 32, 33). The LibSVM (34) SVR implementation was used with radial basis function (rbf) kernel.

The models were trained based on the whole database, as the sample number of *Set II* would not have been sufficient for training purposes. Separate models were trained based on

TABLE 1 Main descriptive statistics of subjects of the applied Depressed Speech Database (DEPISDA).

		Count	H2B (score)		BDI (score)		Age (year)	
			Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Set I	Both	175	13.5	12.8	same as for H2B	same as for H2B	44.3	16.0
	Males	62	13.3	11.8	same as for H2B	same as for H2B	43.2	17.9
	Females	113	13.7	13.4	same as for H2B	same as for H2B	44.9	14.9
Set II	Both	43	20.7	7.9	21.4	7.9	34.3	11.7
	Males	12	18.1	7.7	23.9	5.6	35.7	10.4
	Females	31	21.6	7.8	27.4	8.5	33.8	12.4
All	Both	218	15.0	12.3	16.1	13.1	42.3	15.8
	Males	74	14.1	11.4	15.0	11.7	42.0	17.1
	Females	144	15.4	12.8	16.7	13.7	42.4	15.1

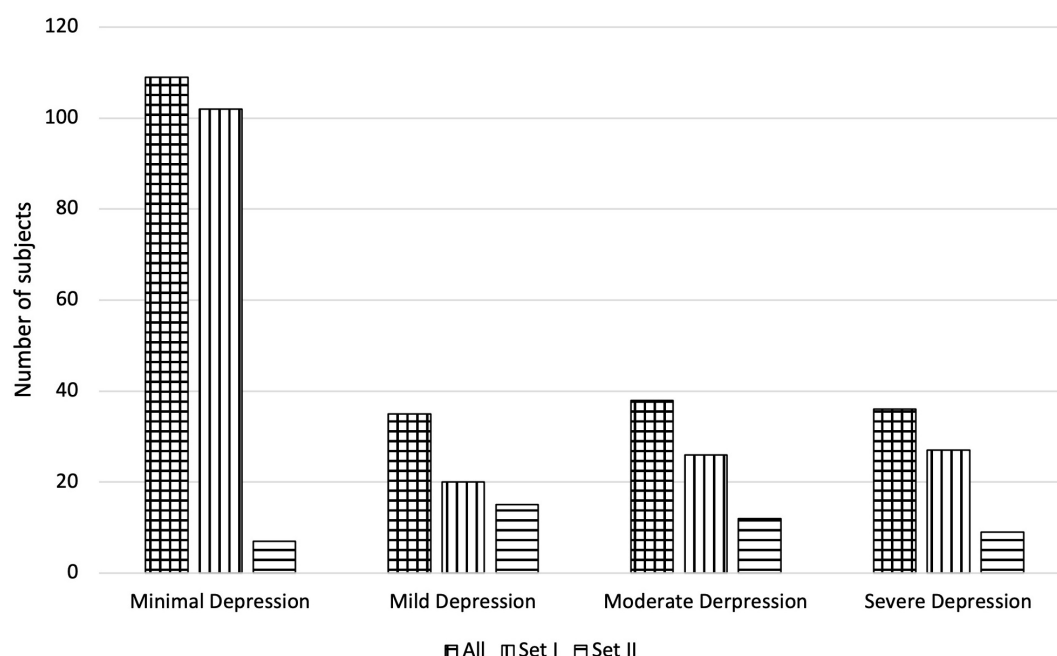


FIGURE 1

Number of subjects of the applied Depressed Speech Database by H2B category in the case of All, Set I, and Set II.

the gender of the subjects, because it improves the automatic recognition of depression (9, 28). Due to the small size of the database, the training and testing of SVR models were implemented with leave-one-out cross-validation (LOOCV), so each sample was selected once as a test set and the model were trained and optimized with the remaining samples. The accuracy of the SVR depends largely on the appropriate choice of the input feature vector and hyperparameters (cost and gamma). Thus, the training and optimization of the models were also implemented using the LOOCV. This double cross-validation is called as nested loop (10, 35), in our case nested LOOCV. The advantage of the method is that the sets of train, development and test are completely independent of

each other, thus minimizing the chance of overfitting and so gives a realistic overview of the accuracy and generalizing ability of the given method. In addition, we also give the results with fivefold nested loops at the beginning to compare the nested LOOCV with the classic 80/20 train/test separation as well.

A feature selection process was applied to reduce the feature set size by selecting the 20 features correlating the most with depression severity (highest Pearson correlation) based on the train samples and then applied a fast forward selection (selecting max 20 features) method (9) to reduce redundancy and find the best performing feature set. Finally, the optimization of the hyperparameters (cost and gamma) was

TABLE 2 The used descriptive features calculated from LLDs.

	Calculated from	Total feature number
Fundamental frequency	Voiced parts	6
Intensity	Vowels, whole speech	12
Jitter	E, O, vowels	18
Shimmer	E, O, vowels	18
First and second formant frequencies and their bandwidths	E, O, vowels	72
13 MFCC	Vowels, whole speech	156
Articulation rate	Whole speech	1
Pause ratio	Whole sample	1
Ratio of transients	Whole speech, whole sample	2
Total:		286

TABLE 3 Accuracy of depression prediction, when the BDI scores were used for training.

	Target	RMSE	MAE	Pearson	Spearman
Set I	BDI (=H2B)	10.1 (m: 7.7; f: 11.2)	7.7 (m: 6.2; f: 8.5)	0.63 (m: 0.76; f: 0.56)	0.59 (m: 0.73; f: 0.52)
Set II	BDI	11.6 (m: 9.5; f: 12.4)	8.7 (m: 7.5; f: 9.2)	0.19 (m: 0.32; f: 0.14)	0.24 (m: 0.11; f: 0.15)
	H2B	9.3 (m: 6.0; f: 10.3)	7.1 (m: 4.7; f: 8.0)	0.21 (m: 0.56; f: 0.15)	0.27 (m: 0.55; f: 0.25)
All	BDI	10.4 (m: 8.0; f: 11.5)	7.9 (m: 6.4; f: 8.7)	0.61 (m: 0.73; f: 0.56)	0.61 (m: 0.71; f: 0.56)
	H2B	9.9 (m: 7.4; f: 11.0)	7.6 (m: 5.9; f: 8.4)	0.61 (m: 0.76; f: 0.54)	0.61 (m: 0.74; f: 0.54)

realized by grid search method, where the powers of 2 were tried out between -7 and $+7$.

Evaluation

Beside Root Mean Square Error (RMSE) as primary evaluation metric, mean absolute error (MAE), Spearman and Pearson correlation coefficients between the original and predicted target scores were also calculated.

Based on the predicted values of the regression method, classification experiments were also performed, where the classification accuracy, sensitivity and specificity metrics were examined. In addition, the receiver operating characteristic (ROC) curve and its area under the curve (AUC) score of the classification models are given.

It is important to note that due to the application of the nested loop, test samples were completely independent of the training and optimization of the models, so the results obtained reflect the real generalizing ability of our method.

Results

Depression severity estimation based on Beck Depression Inventory

In the first experimental setup, we performed model training and optimization based on the BDI scores. The best results

were obtained with five features in the feature vectors for both males and females using nested LOOCV (RMSE: 11.5 in case of females and 8.0 in case of males). In contrast, for fivefold nested loops, the best results were slightly worse (RMSE: 12.0 in case of females, 8.3 in case of males). Of course, this is not surprising, as in this case the size of the training set is smaller and the model optimization is also more easily overfit, resulting in worse scores during testing. For this reason, only the nested LOOCV evaluation was used in the following.

Since the model optimization and testing was implemented with a nested loop, it is conceivable that a different model was selected for each test sample. For this reason, we provide those features that were selected in at least 90% of the all cases, however, some of the features are highly correlated with each other, so they are easily interchangeable. For female models, these features were ratio of transients on the whole speech parts, the standard deviation of mfcc6 on vowels, and the mean of mfcc1 over the entire speech parts. For male models, these features were ratio of transients on the whole speech parts and the mean of mfcc4 on the vowels. The models were tested for both BDI and H2B target variables, results are shown in Table 3. Results for male and female sets are marked with “m” and “f,” respectively. Scatter plot of the predicted and the original H2B scores is shown in Figure 2, where the boundary of depression is indicated by a dashed line ($H2B = 14$), and the line $x = y$ (perfect decision) was also plotted.

Paired samples *T*-test was used to examine whether the difference in mean absolute error between testing on BDI and H2B was significant in case of Set II. We got $p = 0.129$, based

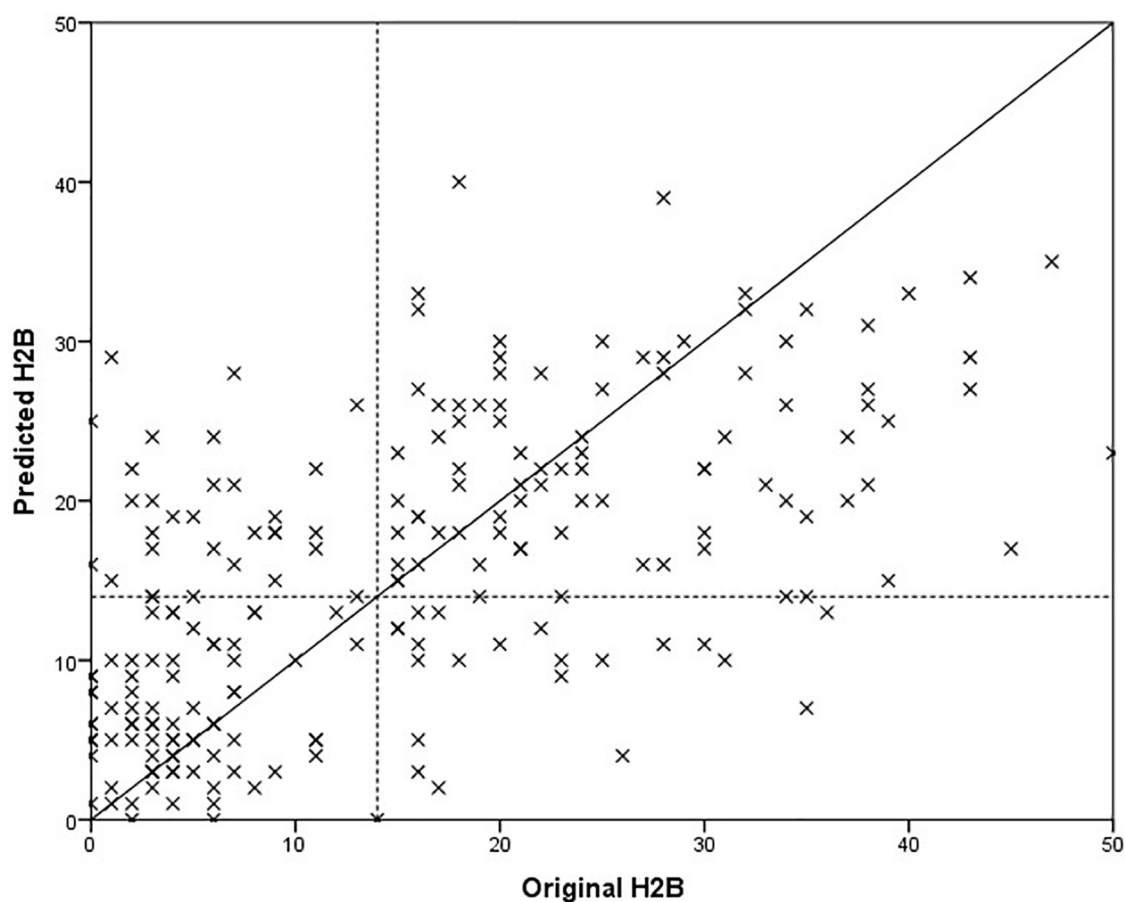


FIGURE 2

Comparison of predicted and original depression severity scores, when the BDI score was used for training.

on which it is not possible to speak of a significant difference. In case of males, it can be observed that the model was able to estimate the severity of depression with a smaller error ($p = 0.007$, using independent samples T -test). One possible explanation is that the mean and the standard deviation of the original BDI scores seems smaller for males (males: 15.0 ± 11.7 ; females 16.7 ± 13.7), and the accuracy of the model was worse for a more severe depression, although no significant difference can be detected in the values of the original BDI scores with an independent sample t -test ($p = 0.381$). At first glance, the results obtained on *Set I* and *Set II* testing on BDI do not appear to differ significantly in the case of the RMSE error value (*Set I*: 10.7; *Set II*: 11.6, $p = 0.370$ using independent samples T -test); however, a significant deterioration can be observed for the correlation values for *Set II* ($p < 0.001$ using Fisher's r to z transformation), from this it can be concluded that the performance of the model is different for *Set I* and *Set II*. It is important to note that while *Set I* contained samples from completely healthy to severely depressed, *Set II* contains mainly speech patterns of depressed subjects (Figure 1). From these it can be concluded that the

method is primarily able to properly separate the whole scale, but in the case of a narrower range it differentiates less reliably.

Although we could not detect a significant difference in the case of *Set II* (where both BDI and HAMD scores were available), when the predicted values were compared to the BDI or H2B. However, it is possible that this was due to the small number of samples (43), which is why we performed further descriptive analyses. In the case of the MAE error value, we obtained an error value of 8.7 on BDI outcomes opposed to 7.1 for H2B prediction (18% relative improvement). Generally, this means an estimate closer to the H2B score by 1.7 on average. To investigate this in more depth, we looked at the difference in absolute errors for each sample when using the BDI or H2B target variable. The histogram of the differences is shown in Figure 3. Negative difference means a closer result to the BDI, while the positive difference means a closer result to the H2B.

Figure 3 shows that in 19% of the cases, the predicted score shows nearly identical differences from the BDI and H2B scores. However, in 26% of cases it approached H2B with a value of 2–6, in 24% with a value of 6–13, while in a further 4% with a

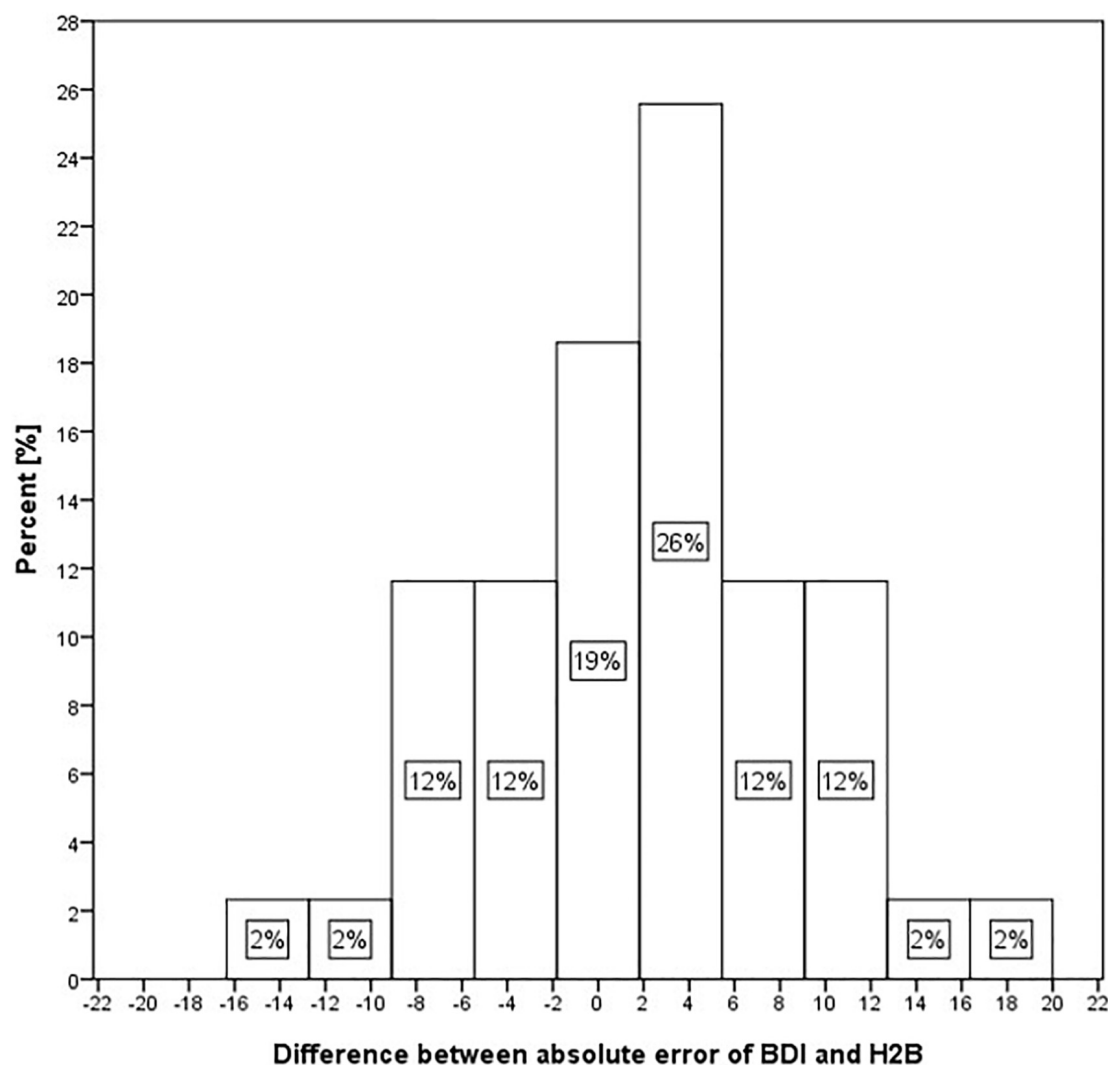


FIGURE 3
Histogram of differences of absolute error in the case of BDI and H2B for Set II.

higher value, in contrast only in 12% of cases it approached the BDI with a value of 2–6, in 14% with a value of 6–13, while in a further 2% with a higher value. Based on these, it can be stated that the estimated scores tend to be closer to the H2B score.

Depression severity estimation based on H2B

In the previous chapter, it was shown that models trained with BDI tended to estimate H2B more accurately than BDI itself. For this reason, we examined how the accuracy of the method changes if we replace the BDI with H2B for *Set II* (the HAMD was available only for this set, resulting different BDI and H2B scores). Best results were obtained with feature vector of size 7 for males and size 9 for females.

As in the previous section, we present the features that were selected in at least 90% of the cases. For female models, these features were both of ratio of transients, standard deviation of intensity on the whole speech parts, mean of mfcc1 on the vowels and mfcc2 on the whole speech parts, standard deviation of mfcc6 on the vowels. For male models, these features were ratio of transients on the whole speech parts, mean of mfcc1 on the whole speech parts and mfcc4 on the vowels. The results are shown in [Table 4](#). A comparison of the predicted scores and the original H2B scores is shown in the [Figure 4](#), where the boundary of depression is indicated by a dashed line ($H2B = 14$), and the line $x = y$ (perfect decision) was also plotted.

In the case of set *All*, we obtained an error value of 8.2 RMSE, which is a 17% relative improvement compared to training and optimizing the models purely with BDI (9.9 RMSE). Paired

TABLE 4 The accuracy of the depression prediction, when the H2B scores were used for training.

	Target	RMSE	MAE	Pearson Coef.	Spearman Coef.
Set I	H2B	8.2 (m: 5.7; f: 9.3)	6.3 (m: 4.4; f: 7.4)	0.77 (m: 0.88; f: 0.72)	0.72 (m: 0.85; f: 0.65)
Set II	H2B	8.2 (m: 3.8; f: 9.4)	6.1 (m: 3.1; f: 7.3)	0.42 (m: 0.87; f: 0.25)	0.46 (m: 0.79; f: 0.32)
All	H2B	8.2 (m: 5.4; f: 9.3)	6.3 (m: 4.2; f: 7.4)	0.75 (m: 0.88; f: 0.69)	0.73 (m: 0.87; f: 0.67)

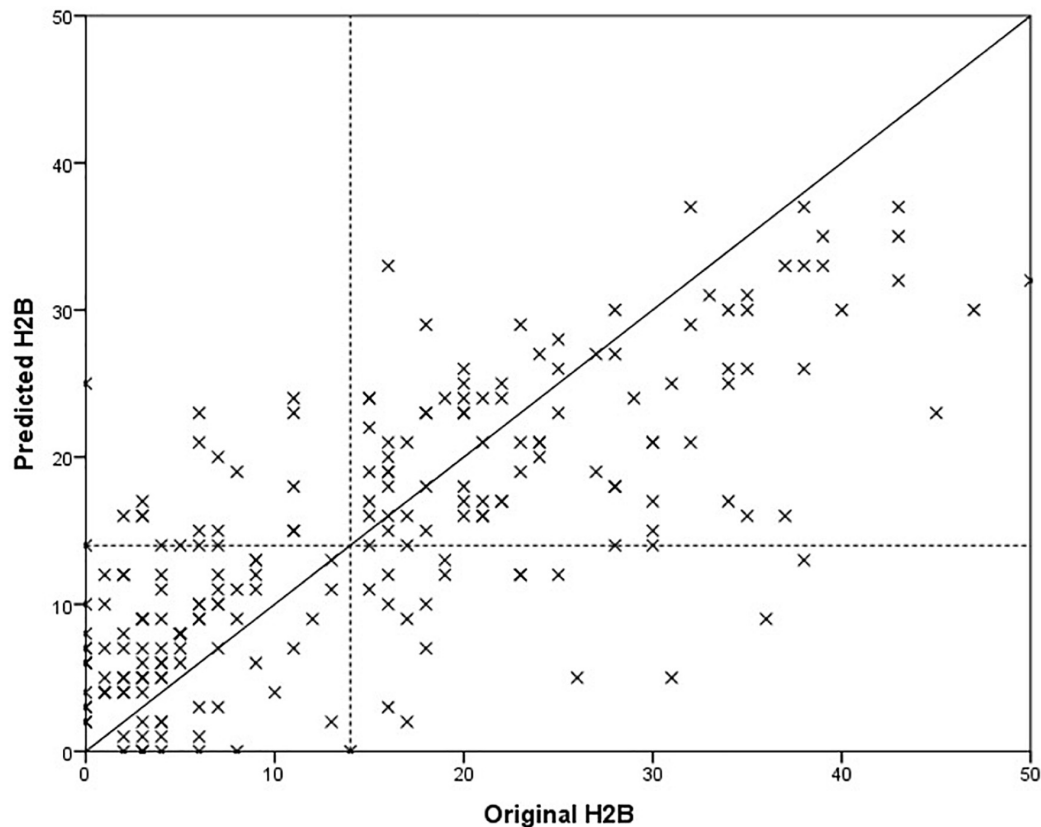


FIGURE 4

Comparison of predicted and original depression severity scores, when the H2B scores were used for training.

samples *T*-test was used to examine whether this difference was significant or not. We got $p < 0.001$, which means that the significant difference can be claimed at a significance level of 99.9%. There was a similar improvement for *Set I* (19%, $p < 0.001$) and *Set II* (12%, $p < 0.001$). It is important to note that the results improved not only for *Set II* but also for *Set I*, which were still trained based on BDI (for *Set I*, BDI = H2B). Based on this, it can be stated that the automatic model created based on the speech signal is better able to learn H2B than BDI.

The degree of improvement can also be observed by comparing Figures 3, 4, as in the second case the estimated scores are closer to the $x = y$ line (perfect prediction), and fewer samples were placed in the upper left and lower right rectangles, which represent erroneous recognitions (upper left: false positive, lower right: false negative).

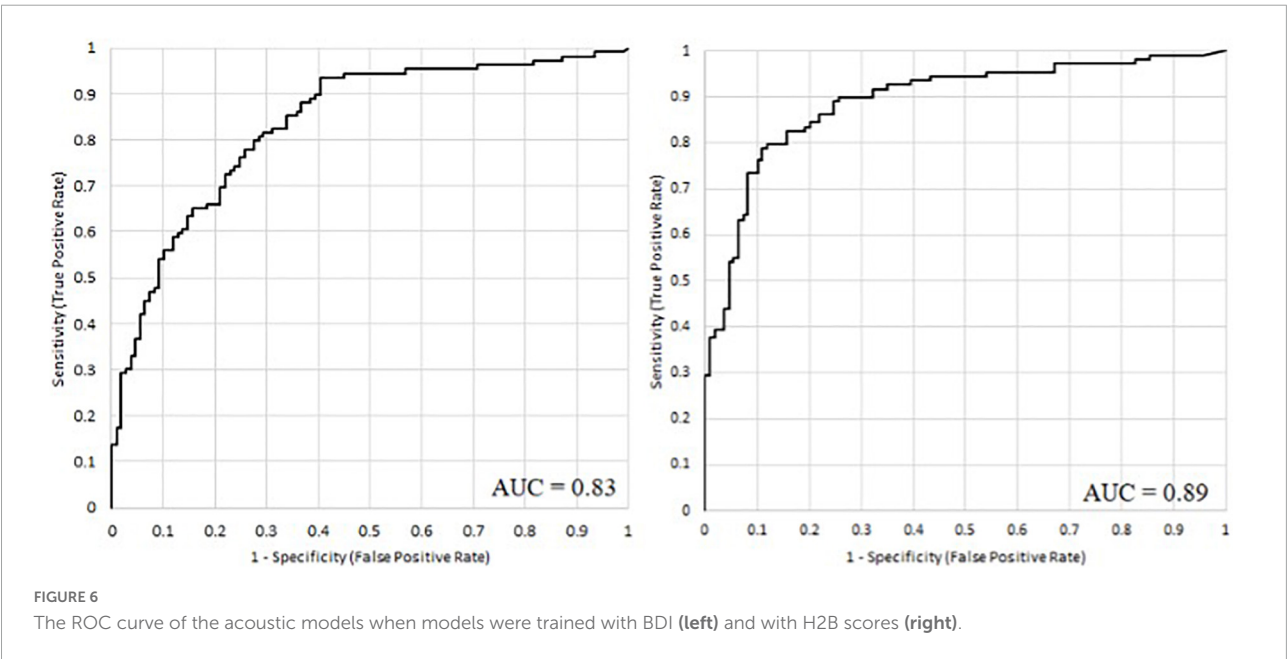
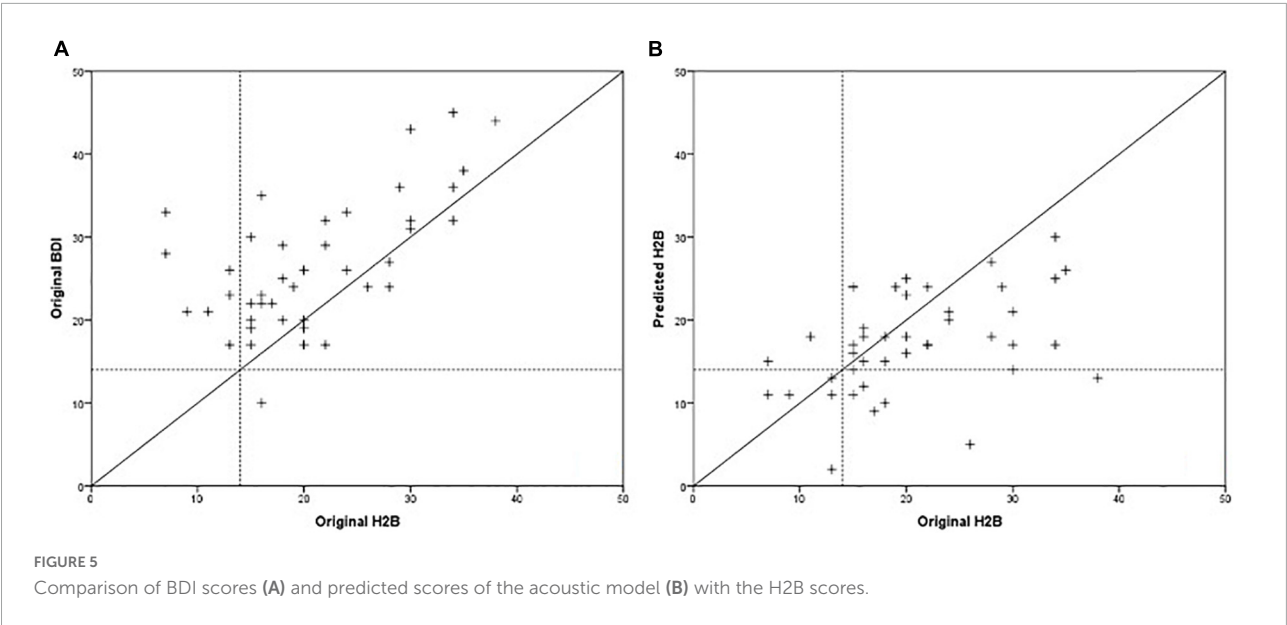
Comparing the precision of acoustic model and Beck Depression Inventory questionnaire

Since both the BDI and HAMD scores were available for *Set II*, this set allowed us to compare the accuracy of the BDI questionnaire and the developed acoustic model. Since the set includes mainly depressed patients, only limited conclusions can be drawn from this comparison, and it is not a good descriptor of the accuracy of the BDI questionnaire in general. However, it may be interesting to compare the two methods, as the long-term goal is to develop a diagnostic support tool that may even surpass the usability of the BDI questionnaire.

Based on Table 5 and Figure 5, it is not possible to clearly determine which method is more suitable for predicting the

TABLE 5 Accuracy of the acoustic model and BDI questionnaire.

	RMSE	MAE	Mean Error	Pearson Coef.	Spearman Coef.
BDI Questionnaire	8.8	6.9	5.7	0.63	0.55
Acoustic model	8.2	6.1	-3.3	0.42	0.46



severity of depression. While the acoustic model has slightly lower RMSE and MAE error values ($p = 0.513$, paired samples T -test), the BDI questionnaire has a better correlation with the H2B (converted HAMD) score ($p = 0.094$, using Fisher's r to z transformation), so it has a better ability to differentiate in this range of depression severity. Comparing the mean errors,

it can be said that while the BDI questionnaire overestimates the severity of depression, the acoustic model underestimates it.

It can also be observed that in the case of the acoustic model, the majority of significant errors occur in the case of severe depression ($H2B > 28$), and the maximum predicted score was around 30. This may be due to the insufficient number of severe

TABLE 6 The accuracy of the classification of depressed and healthy subjects, when H2B or BDI scores were used for training.

	Training variable	Classification accuracy	Sensitivity	Specificity
At maximum classification accuracy	BDI	76%	80%	72%
	H2B	84%	79%	89%
At 90% sensitivity	BDI	75%	90%	62%
	H2B	81%	90%	73%

depressed patients in the database. However, in the 7 cases where subjects received moderate to severe depression scores based on the BDI questionnaire, the speech-based method correctly predicted them to be as non-depressed or mildly depressed.

Distinguishment between depressed and healthy subjects

It has been shown that it is possible to accurately predict the severity of depression based on speech signal processing. However, from the magnitude of the RMSE and MAE error values, it is difficult to infer the applicability of the method directly, and since the long-term goal is to support the diagnosis of depression, we examined how accurately the model can distinguish between depressed and healthy subjects. With the help of the regression method, classification can also be implemented, since a comparator value can be given, above which the given person is classified as depressed, while below it he is classified as non-depressed. Based on this, we examined how accurately it is possible to distinguish between depressed and non-depressed subjects using models trained with BDI and H2B. [Figure 6](#) shows the ROC curves and the AUC values of the two models, and [Table 6](#) shows their classification sensitivity, specificity as well as at their maximum classification accuracy, and when the comparator limit is set to achieve sensitivity value of at least 90%.

Using the Wilcoxon signed ranks test, we examined whether the difference is significant in the case of maximum accuracy, and we obtained a value of $p = 0.022$, which means that the difference is significant. Based on the examination, it can be stated that the use of H2B in training also improves the classification, as we achieved higher AUC (area under curve) value (0.89), higher maximum classification accuracy (84%), higher specificity (73%) and classification accuracy (81%) at 90% sensitivity.

Discussion

As the prevalence of depression is increasing worldwide, burdening the health care systems and the global

economy, screening systems for early diagnostics are much more needed. Analyzing speech samples can be an easy, cheap and widely available tool in screening process of depression, throughout general medical practices (10). In our study we try to achieve a more accurate, more sensitive model to determine depression by speech samples.

We primarily sought to answer how a limited target variable can be used to estimate the actual severity of depression or to differentiate between depressed and non-depressed subjects. BDI score may be a good descriptor of the severity of depression; however, since it can be determined using a self-reported questionnaire, due to exaggeration and concealment of symptoms it cannot be considered as a fully objective descriptor of the severity of depression. However, determining the BDI score is faster and less resource intensive, making it easier to create a large speech database based, which is a prerequisite for accurate machine learning models.

Hamilton Rating Scale for Depression score, as a clinician-administered scale, assess externally determinable symptoms of depression (like somatic and behavior symptoms), and a more sensitive measure of symptom change, as it is determined by a psychiatrist, also depends on the clinician's training and subjective judgment.

During the research, we had access to a large speech database containing speech samples from depressed and healthy subjects. The database contains the BDI score of each subject and the HAMD score of 20% of the subjects. This fact provided us with an opportunity to compare the usability of the BDI and the HAMD for the training of the models of the automatic recognition of depression based on speech signal processing. For comparison, HAMD scores were converted to the BDI scale by fitting the endpoints of the two scale categories, and linear scaling was performed within each category to obtain the converted HAMD score: H2B.

We have shown that when models were trained with BDI scores, the predicted scores are on average 1.7 points closer to the converted HAMD scores than to the BDI scores, despite the fact that the BDI scores themselves approximate the HAMD scores quite accurately (MAE 6.9). Of course, while interpreting the results we had to take it into account that we had speech sample from 43 subjects available to perform the experiment, of

which 36 subjects were depressed. Thus, the study may be worth performing on a larger set of samples in the future that includes a balanced number of depressed and healthy subjects.

We have shown that changing the BDI scores to HAMD scores at the training improves the automatic prediction and recognition of depression. We examined this by using the converted HAMD scores (H2B) instead of the BDI scores of the 43 samples available to us. Thus, for regression, the RMSE error decreased from 9.9 to 8.2, which is a 17% relative improvement, while the MAE error decreased from 7.6 to 6.3, which is a 17% relative improvement also, which is significant at 99.9% significance level. In the case of classification, accuracy increased from 81 to 84%, which is a 10% relative improvement in the number of erroneously decided subjects.

We also examined whether the acoustic model, or the BDI scale, provides a more accurate prediction of HAMD scores. In the experiment, we found that while the acoustic model estimates HAMD scores (i.e., severity of depression) with smaller RMSE and MAE values than the BDI scale, BDI scores show a higher correlation with HAMD scores than the predicted ones. Thus, it is not possible to clearly determine which is more suitable for a more accurate estimate of the severity of depression; however, it can be concluded that the usability of the acoustic model is comparable to the BDI scale. It is worth noting that data augmentation would likely to further slightly improve the results, for example, if the recordings are splitted into smaller segments (36) and then nested speaker LOOCV evaluation is applied.

The results of our study show that screening by acoustic biomarkers, the diagnosis of depression can be recognized earlier, and with a machine learning based automatic decision system, it can be used widely as a diagnostic tool in the general medical practices. A machine learning based system using acoustic biomarkers of depression can have public applicability, with the possibility of screening the general population for depression by easily accessible, cheap methods, e.g., smartphone applications, webpages. The early recognition of depression by speech analysis can help to refer the patients to psychiatrists earlier to start the proper treatment. The early recognition of depression can improve the quality of life, decrease the mortality and the average length of stay in hospitals, in this way can reduce the economic cost and ease the health care systems. Another application of speech analysis in the diagnosis of depression is to follow the patient's recovery under treatment and to compare effectivity of variable therapies.

In the near future, we would like to test the method we have developed in practice in primary healthcare, as well as further expand the database to create even more accurate models and better compare the usability of the acoustic model with the BDI scale.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Semmelweis University Regional and Institutional Committee of Science and Research Ethics. The patients/participants provided their written informed consent to participate in this study.

Author contributions

BH-D, GK, DS, KV, and LS contributed to the design and implementation of the research. BH-D, GK, and DS collected data on the patients. GK performed the computations and statistical analysis. BH-D and GK wrote the manuscript. DS, KV, and LS supervised the project. All authors provided critical comments and approved the final version of the manuscript.

Funding

This research was supported by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the Neurology thematic programme of Semmelweis University. Project no. K1285968 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K_18 funding scheme.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- World Health Organization [WHO]. *Depression fact sheet* Geneva, Switzerland. Geneva: WHO (2020).
- World Health Organization [WHO]. *The global burden of disease: 2004 update*. Geneva: WHO (2008).
- Stuart H. Reducing the stigma of mental illness. *Glob Ment Health (Camb)*. (2016) 3:e17. doi: 10.1017/gmh.2016.11
- Lin CH, Lu MJ, Wong J, Chen CC. Comparison of physician-rating and self-rating scales for patients with major depressive disorder. *J Clin Psychopharmacol*. (2014) 34:716–21. doi: 10.1097/JCP.0000000000000229
- Lakhan SE, Vieira K, Hamlat E. Biomarkers in psychiatry: Drawbacks and potential for misuse. *Int Arch Med*. (2010) 3:1. doi: 10.1186/1755-7682-3-1
- Schmidt HD, Shelton RC, Duman RS. Functional biomarkers of depression: Diagnosis, treatment, and pathophysiology. *Neuropsychopharmacology*. (2011) 36:2375–94. doi: 10.1038/npp.2011.151
- de Aguiar Neto FS, Rosa JLG. Depression biomarkers using non-invasive EEG: A review. *Neurosci Biobehav Rev*. (2019) 105:83–93. doi: 10.1016/j.neubiorev.2019.07.021
- Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. *Speech Commun*. (2015) 71:10–49. doi: 10.1016/j.specom.2015.03.004
- Kiss G, Vicsi K. Mono-and multi-lingual depression prediction based on speech processing. *Int J Speech Technol*. (2017) 20:919–35. doi: 10.1007/s10772-017-9455-8
- Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investig Otolaryngol*. (2020) 5:96–116. doi: 10.1002/lio2.354
- Low L-SA, Maddage NC, Lech M, Sheeber LB, Allen NB. Detection of clinical depression in adolescents' speech during family interactions. *IEEE Trans Biomed Eng*. (2010) 58:574–86. doi: 10.1109/TBME.2010.2091640
- Alghowinem S, Goecke R, Wagner M, Epps J, Breakspear M, Parker G editors. From joyous to clinically depressed: Mood detection using spontaneous speech. *Proceedings of the twenty-fifth International Florida artificial intelligence research society conference. Association for the Advancement of Artificial Intelligence (AAAI) Citeseer*. Menlo Park, CA: AAAI (2012). p. 141–6.
- Pampouchidou A, Simantiraki O, Vazakopoulou C-M, Chatzaki C, Pedititis M, Maridaki A, et al. editors. Facial geometry and speech analysis for depression detection. *Proceeding of the 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. Piscataway, NJ: IEEE (2017). doi: 10.1109/EMBC.2017.8037103
- Huang Z, Epps J, Joachim D, Stasak B, Williamson JR, Quatieri TF. Domain adaptation for enhancing Speech-based depression detection in natural environmental conditions using dilated CNNs. *Interspeech*. (2020) 2020:4561–5. doi: 10.21437/Interspeech.2020-3135
- Kiss G, Vicsi K editors. Comparison of read and spontaneous speech in case of automatic detection of depression. *Proceeding of the 2017 8th IEEE international conference on cognitive infocommunications (CogInfoCom)*. Piscataway, NJ: IEEE (2017). doi: 10.1109/CogInfoCom.2017.8268245
- Cummins N, Sethu V, Epps J, Williamson JR, Quatieri TF, Krajewski J. Generalized two-stage rank regression framework for depression score prediction from speech. *IEEE Trans Affect Comput*. (2017) 11:272–83. doi: 10.1109/TAFFC.2017.2766145
- Rejaibi E, Komaty A, Meriaudeau F, Agrebi S, Othmani A. Mfcc-based recurrent neural network for automatic clinical depression recognition and assessment from speech. *arXiv [Preprint]*. (2019):190907208. doi: 10.48550/arXiv.1909.07208
- Williamson JR, Young D, Nierenberg AA, Niemi J, Helfer BS, Quatieri TF. Tracking depression severity from audio and video based on speech articulatory coordination. *Comput Speech Lang*. (2019) 55:40–56. doi: 10.1016/j.csl.2018.08.004
- Kiss G, Jenei AZ editors. Investigation of the accuracy of depression prediction based on speech processing. *Proceeding of the 2020 43rd international conference on telecommunications and signal processing (TSP)*. Piscataway, NJ: IEEE (2020). doi: 10.1016/j.cmpb.2021.106433
- Lopez-Otero P, Docio-Fernandez L. Analysis of gender and identity issues in depression detection on de-identified speech. *Comput Speech Lang*. (2021) 65:101118. doi: 10.1016/j.csl.2020.101118
- Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, et al. The montgomery asberg and the Hamilton ratings of depression: A comparison of measures. *Eur Neuropsychopharmacol*. (2006) 16:601–11. doi: 10.1016/j.euroneuro.2006.04.008
- Richter P, Werner J, Heerlein A, Kraus A, Sauer H. On the validity of the Beck depression inventory. A review. *Psychopathology*. (1998) 31:160–8. doi: 10.1159/000066239
- Bukumiric Z, Starcevic V, Stanisavljevic D, Marinkovic J, Milic N, Djukic-Dejanovic S, et al. Meta-analysis of the changes in correlations between depression instruments used in longitudinal studies. *J Affect Disord*. (2016) 190:733–43. doi: 10.1016/j.jad.2015.10.054
- Lambert MJ, Hatch DR, Kingston MD, Edwards BC. Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison. *J Consult Clin Psychol*. (1986) 54:54–9. doi: 10.1037//0022-006x.54.1.54
- Edwards BC, Lambert MJ, Moran PW, McCully T, Smith KC, Ellingson AG. A meta-analytic comparison of the Beck Depression Inventory and the Hamilton Rating Scale for Depression as measures of treatment outcome. *Br J Clin Psychol*. (1984) 23(Pt 2):93–9. doi: 10.1111/j.2044-8260.1984.tb00632.x
- de Boer JN, Voppel AE, Brederoo SG, Wijnen FNK, Sommer IEC. Language disturbances in schizophrenia: The relation with antipsychotic medication. *NPJ Schizophr*. (2020) 6:24. doi: 10.1038/s41537-020-00114-3
- Kiss G, Sztahó D, Vicsi K editors. Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. *Proceeding of the 2013 IEEE 4th international conference on cognitive infocommunications (CogInfoCom)*. Piscataway, NJ: IEEE (2013). doi: 10.1109/CogInfoCom.2013.6719169
- Cummins N, Vlasenko B, Sagha H, Schuller B editors. Enhancing speech-based depression detection through gender dependent vowel-level formant features. *Proceeding of the Conference on artificial intelligence in medicine in Europe*. Berlin: Springer (2017). doi: 10.1007/978-3-319-59758-4_23
- Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. *Adv Neural Inform Process Syst*. (1997) 9:155–61.
- Cortes C, Vapnik V. Support vector machine. *Machine Learning*. (1995) 20:273–97. doi: 10.1007/BF00994018
- Suykens JA, Vandewalle J. Least squares support vector machine classifiers. *Neural Process Lett*. (1999) 9:293–300.
- Kiss G, Tulics M, Sztahó D, Esposito A, Vicsi K, FaundezZanuy M, et al. Language independent detection possibilities of depression by speech. *Recent Adv Nonlinear Speech Process*. (2016) 48:103–14. doi: 10.1007/978-3-319-28109-4_11
- He L, Cao C. Automated depression analysis using convolutional neural networks from speech. *J Biomed Inform*. (2018) 83:103–11. doi: 10.1016/j.jbi.2018.05.007
- Chang C-C, Lin C-J. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol (TIST)*. (2011) 2:1–27. doi: 10.1145/1961189.1961199
- Gosztolya G, Vincze V, Tóth L, Pákási M, Kálmán J, Hoffmann I. Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Comput Speech Lang*. (2019) 53:181–97. doi: 10.1016/j.csl.2018.07.007
- Rozmán P, Sztahó D, Kiss G, Jenei AZ. Automatic recognition of depression and Parkinson's disease by LSTM networks using sample chunking. *Proceeding of the 2021 IEEE 12th international conference on cognitive infocommunications (CogInfoCom)*. Piscataway, NJ: IEEE (2021).

**ÖSSZEFOGLALÓ
KÖZLEMÉNY****REVIEW ARTICLE**

Automatizált hangelemzés és depresszió: Szisztematikus umbrella review

HAJDUSKA-DÉR Bálint^{ID}, SIMON Lajos^{ID}, RÉTHELYI János^{ID},
HALUSKA-VASS Edit^{ID}

Semmelweis Egyetem, Pszichiátriai és Pszichoterápiás Klinika, Budapest

 | Hungarian | <https://doi.org/10.18071/isz.77.0087> | www.elitmed.hu**Levelező szerző**

(correspondent):
HALUSKA-VASS Edit, PhD,
Semmelweis Egyetem,
Pszichiátriai és Pszicho-
terápiás Klinika;
1083 Budapest, Balassa u. 6.
Telefon +36 30 360 9164,
e-mail:
vass.edit@semmelweis.hu
<https://orcid.org/0000-0002-8858-7277>

Érkezett:

2024. szeptember 27.

Elfogadva:

2024. december 14.

Bevezetés – A depresszió korai és pontos diagnóza elengedhetetlen a megfelelő kezelés mielőbbi megkezdéséhez, azonban a hagyományos diagnosztikai módszerek használata gyakran szubjektív, ember- és időigényes. Az automatizált hangelemzés ígéretes megoldást kínálhat e kihívások leküzdésére, lehetővé téve a hangalapú biometrikus jellemzők objektív mérését. A jelen tanulmány célja a gépi tanulóval támogatott hangelemzési módszerek kutatási gyakorlatának és eredményeinek áttekintése, a következetlenségek feltárása, valamint a jövőbeli kutatások számára konstruktív javaslatok megfogalmazása.

Módszerek – A jelen közlemény egy úgynevezett umbrella review eredményeit ismerteti, ami már publikált szisztematikus irodalmi áttekintések és metaanalízisek eredményeit integrálja. A közlemények azonosításához a PubMed, Scopus és ProQuest adatbázisokat használtuk, követve a PRISMA irányelveit. A keresési intervallum az elmúlt 5 évre terjedt ki. A keresés előre meghatározott keresőkulcs és szelekciós kritériumok segítségével történt, független áttekintők alkalmazásával. A részletes elemzést megelőzően a közlemények módszertani megfelelőségét az AMSTAR2 értékelőrendszerrel vizsgáltuk.

Eredmények – A szisztematikus irodalomkutatás során összesen 162 egyedi rekordot azonosítottunk. A bevonási és kizárási kritériumok alapján végül hat közlemény részletes elemzésére került sor. Az eredmények rámutatnak az eddig fejlesztett modellek korlátozott alkalmazhatóságának háttértényezőire, valamint hangsúlyozzák azon akusztikai jellegzetességek fontosságát, amelyek a feltárt

Automated audio analysis and depression: A systematic umbrella review

Hajduska-Dér B; Simon L, PhD; Réthelyi J, PhD; Haluska-Vass E, PhD

Background and purpose – The early and accurate diagnosis of depression is essential for the timely initiation of appropriate treatment. However, the use of traditional diagnostic methods is often subjective, labour-intensive, and time-consuming. Automated voice analysis offers promising solution to overcome these challenges by enabling the objective measurement of voice-based biometric features. The aim of this study is to review the research practice and findings of voice analysis methods supported by machine learning, to uncover inconsistencies, and to formulate constructive suggestions for future research.

Methods – This paper presents the results of a so-called umbrella review, which integrates the findings of already published systematic literature reviews and meta-analyses. For the identification of publications, we used the PubMed, Scopus, and ProQuest databases, following the PRISMA guidelines. The search interval covered the last 5 years. The search was conducted using predefined search terms and selection criteria, with the involvement of independent reviewers. Prior to detailed analysis, the methodological quality of the publications was assessed using the AMSTAR2 evaluation system.

Results – Through the systematic literature search, we identified a total of 162 unique records. Based on the inclusion and exclusion criteria, 6 publications were selected for detailed analysis. The results highlight the background factors limiting the applicability of the developed models and also emphasize the importance of acoustic characteristics that can be identified as biomarkers of

módszertani következetlenségek ellenére a depresszió biomarkereként azonosíthatók. A jelen áttekintés alátámasztja a gépi tanulás és a hangelemzés jelentőségét a depresszió diagnosztikájának fejlesztésében. Ahhoz azonban, hogy a kutatási eredményeket a gyakorlatba át tudjuk ültetni, standardizált módszerek alkalmazásán túl többek között különböző összetételű tesztcsoportokon történő validálás szükséges.

Következtetés – A gépi tanulás alkalmazása a depresszió detektálásában számos előnyt ígér, ilyen az objektív diagnosztika vagy a korai felismerés. Ez a technológia hosszú távon költséghatékony megoldást kínálhat, miközben nagyobb hozzáférhetőséget biztosít a mentális egészségügyi szolgáltatásokhoz. Mindazonáltal, a terület még fejlődőben van, és további kutatásokra van szükség a módszerek megbízhatóságának növelése érdekében.

Kulcsszavak: depresszió, gépi tanulás, automatizált hangelemzés

depression despite methodological inconsistencies. This review supports the importance of machine learning and voice analysis in advancing the diagnostics of depression. However, to translate research outcomes into practice, beyond the application of standardized methods, validation across diverse test groups is necessary, among other things.

Conclusion – The application of machine learning in depression detection promises numerous advantages, such as objective diagnosis or early detection. This technology could offer cost-effective solution in the long run while providing greater access to mental health services. Nevertheless, the field is still evolving, and further research is needed to enhance the reliability of these methods.

Keywords: depression, machine learning, automated voice analysis

A depresszió korunk egyik leggyakoribb mentális betegsége, ami világszerte közel 280 millió embert érint, és jelentős hatással van az egyének életminőségére. A tünetek korai felismerése és a hatékony kezelés kulcsfontosságú a depresszióval küzdők támogatásában; azonban az alacsony és közepes jövedelmű országokban az emberek több mint 75%-a nem részesül megfelelő kezelésben¹. Az évtizedek óta használt pszichodiagnosztikai tesztek és klinikai interjúk hatékonyan kiszűrjük a depressziós állapotokat, de ezek gyakran ember- és időigényesek, ami növeli a gazdasági ráfordítást, emellett gyakran szubjektív természetük is problémát jelenthet.

Az utóbbi években egyre nagyobb figyelmet kapott a gépitanulás- és a hangalapú biomarkerek alkalmazása a depressziós állapotok diagnosztizálásában, mivel ezek a technológiák potenciálisan objektív és gyors módszereket kínálnak a mentális állapotok értékelésére¹. A beszédelemzés biztató eredményt mutat a depresszió potenciális biomarkereként, ahogy azt többek között azok a kutatások is alátámasztják, amelyek depresszív állapotok esetén a beszéd csökkent sebességét, megnövekedett szünethosszokat és a hangmagasság kisebb variabilitását azonosították². A hatékonyságra irányuló biztató eredmények mellett a beszéd könnyen és invazív eszközök nélkül begyűjthető adattípus, ami számos, a depresszió tüneteivel kapcsolódó

információt hordoz³. A gépi tanuló eljárások fejlődésével a beszéd akusztikai elemzése felgyorsítható és automatizálható, lehetővé téve olyan szűrőeszközök létrehozását, amelyek emberi beavatkozás nélkül is használhatók. Jelenleg a különböző kutatások által elért diagnosztikai pontosság 70–80% közötti, azonban az eltérő vizsgálati tervek, a validációs vizsgálatok hiánya és a kis mintanagyságok korlátozzák az általánosítható modellek kidolgozását⁴.

Az említett hiányosságok és a gyakran ellentmondásos eredmények feltérképezése érdekében úgynevezett umbrella review alkalmazásával foglaltuk össze a területen született eddigi eredményeket. Ez a megközelítés lehetővé teszi, hogy a már publikált szisztematikus irodalmi áttekintések és metaanalízisek alapján integrált és átfogó képet nyújtsunk a kutatási eredményekről, az esetleges hiányosságokról és a feltárható ellentmondások természetéről, ezzel elősegítve a jövőbeli kutatások irányvonalának meghatározását az általánosítható modellek hatékony kidolgozása érdekében.

Fontos megjegyezni, hogy a vizsgálatba bevont tanulmányok a fókusz nem az egyértelműen diagnosztizálható depressziós epizódok detektálására helyezik, hanem a depressziós tünetekkel jellemezhető állapotok felismerésére töreksenek. Így a továbbiakban depresszió szűrése alatt a depressziós állapot azonosítását értjük.

Módszer

A jelen vizsgálat már publikált szisztematikus irodalmi áttekintések és metaanalízisek vizsgálatára épülő, úgynevezett umbrella review, a vizsgált tudományterületen született eddigi eredmények összefoglalása és integrálása érdekében.

Szisztematikus irodalomkeresés

A vizsgált téma szempontjából releváns közlemények azonosításához a PubMed, Scopus és ProQuest adatbázisokra támaszkodtunk. A közlemények azonosításának folyamata a PRISMA irányelv alapján, szisztematikus irodalomkeresés révén történt. A keresést 2024. szeptember 18-án bonyolítottuk le, a keresési intervallum a tudományterület dinamikus fejlődésére tekintettel az elmúlt 5 évre terjedt ki.

A keresés során az alábbi keresőkulcsot alkalmaztuk: „acoustic”* OR „speech” OR „vocal” OR „audio” AND „depress”* AND „Machine learning” OR „Deep learning” OR „Neural network” OR „automated detection”.

A tanulmányok azonosításakor az adatbázisokban ismerten előforduló rekordok (duplumok) kizárását követően cím, majd absztrakt alapján határoztuk meg a releváns közleményeket. A fennmaradó közlemények teljes szövegének áttekintésére sor került. A szűrési folyamat minden lépését az első és utolsó szerző egymással párhuzamosan, egymástól függetlenül végezte. A rekordok kizárásának okát a szerzők a szűrést követően egyeztetették és rögzítették. Amennyiben a közlemények bevonhatóságát tekintve döntésük nem egyezett, a folyamatba a második és harmadik szerző mint független értékelő bevonására is sor került.

A közlemények relevanciájának megítélésére az alábbi bevonási és kizárási kritériumok mentén került sor:

Bevonási kritériumok: 1. A közlemény angol nyelven íródott; 2. az elmúlt 5 évben publikálták; 3. típusát tekintve PRISMA irányelv alapján készült irodalmi áttekintés (annak további típusától függetlenül); 4. legalább részben automatikus, géptanulás-alapú hangelemzésre épülő detekciós rendszerek vizsgálatára irányul depressziós betegpopulációban (de nem feltétlenül csak depresszióban).

Kizárási kritériumok: 1. A vizsgálat nem fogalmaz meg következtetéseket a depresszió automatikus hangelemzéssel történő detektálásáról, ideértve hatékonyságvizsgálati eredményeket vagy módszertani megfontolásokat. 2. A közlemény teljes szövege nem elérhető.

Tekintettel arra, hogy a bevont tanulmányok közül csak egy volt metaanalízis, a bevont közlemények kvalitatív elemzésére került sor. A kvalitatív elemzést megelőzően az elemzésben szereplő szisztematikus áttekintések és metaanalízis minőségét AMSTAR 2 értékelőrendszerrel mértük és értékeltük, hogy láthatóvá tegyük azokat a hi-

bákat, melyek az áttekintő tanulmányok megbízhatóságát és érvényességét befolyásolhatják⁵.

Eredmények

Szisztematikus irodalomelemzés

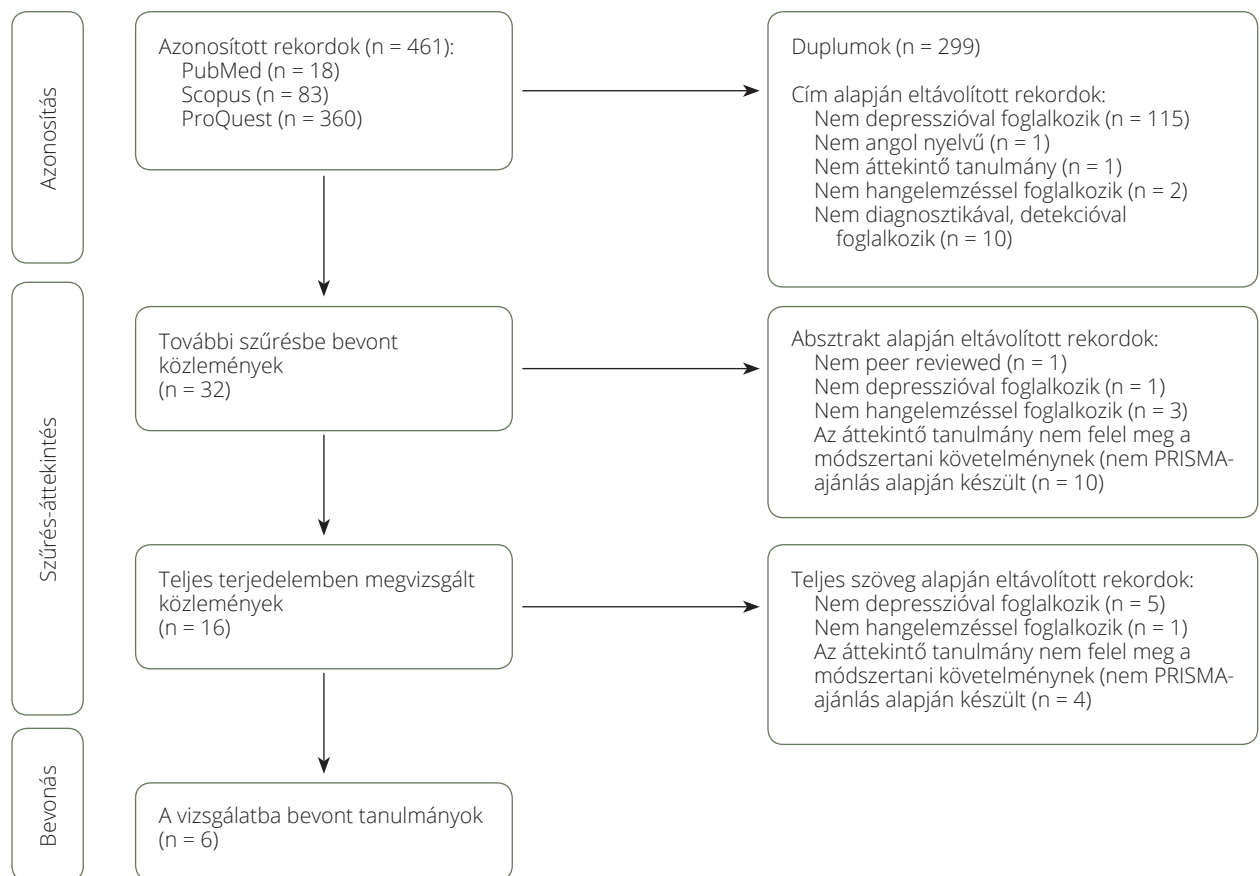
A szisztematikus irodalomkutatás során összesen 162 egyedi, nem duplikált rekordot azonosítottunk. A bevonási és kizárási kritériumok alapján a megvizsgált tanulmányok közül 16 esetben kerülhetett sor azok tartalmának teljes terjedelemben történő átvizsgálására. Az utóbbi munkaszakasz eredményeként további 10 közlemény kizárására került sor. A közlemények kiválasztásának folyamatát az **1. ábra** szemlélteti.

A bevont közleményeket a részletes kvalitatív elemzést megelőzően AMSTAR 2 értékelőrendszer segítségével vizsgáltuk, melynek eredményét az **1. táblázat** rögzíti. Az AMSTAR 2-elemzés alapján megállapítható, hogy a vizsgálatba bevont közlemények többsége közepes vagy magas minőségű, míg egy tanulmány alacsony minősítést kapott. A minősítés során a közlemények többsége megfelelt az eredmények ismertetésére és interpretálására irányuló követelményeknek. A tanulmányok értékelésének alapját így főként az elemzés módszertani megfelelősége képezte. Ennek alapján három tanulmány egyetlen adatbázisban végzett keresést, illetve mindössze egy alkalmazott független értékelőket az irodalomkutatás folyamatában. Emellett mindössze egy közlemény nyilatkozott a protokoll előzetes regisztrációjáról (**1. táblázat**). Egyetlen tanulmány végezte el a bevont közlemények irodalomjegyzékének áttekintését.

A hat vizsgálatba bevont összefoglaló tanulmány főbb jellemzőit a **2. táblázat** összesíti. A bevont közlemények átlagosan 90 tanulmány szisztematikus áttekintését végezték el (minimum: 36, maximum: 264); a depresszió géptanulás-alapú hangelemzésre épülő detekciójával átlagosan 23 tanulmány foglalkozott. Ennek oka, hogy mindössze két olyan áttekintés született, ami a bevonási kritériumoknak megfelel, és depresszión kívül más pszichiátriai zavar detekciójáról nem nyilatkozik. Egy tanulmány nem határozta meg a bevonási időszak alsó határát, a fennmaradó tanulmányok áttekintési időszaka 1 és 10 év között mozgott. Az általunk bevont áttekintő közlemények többnyire olyan tanulmányokat vizsgáltak, amelyek mintaelemszáma a géptanulás-alapú predikciós modellek tesztelése szempontjából szuboptimális, azaz a predikált kimeneteli osztályokhoz asszociált létszám 50 fő alatti, és a nők általában felülreprezentáltak.

A hangalapú elemzés formái és módszerei

Ahogy áttekintő tanulmányunk is szemlélteti, a mesterségesintelligencia-alapú detekciós modellek vizsgálata általában nem alkalmaz kontrollcsoportot, mivel ezek a



1. ábra. Az irodalomkutatás folyamatának bemutatása a PRISMA útmutatás alapján javasolt folyamatábrában

vizsgálatok meglévő adatok elemzésére és a modellek tanulmányozására összpontosítanak. Elrendezésüket tekintve tipikusan retrospektív elemzések.

Noha a nyelvi paraméterek és a depresszió közötti összefüggések vizsgálatára alkalmazott technikák alkalmassak lehetnek az egyénen belüli változások mérésére, ezen lehetőségek vizsgálata a bevont közleményekben nem kerül kellőképpen elemzésre. Ez részben annak köszönhető, hogy a mesterségesintelligencia-alapú modellek vizsgálata adatéhes folyamat, ahol az egyénen belüli eltérések nyomon követése a longitudinális elrendezés szükségessége miatt jelentős mértékben megnövelheti a szükséges adatmennyiséget.

A vizsgálatba bevont tanulmányok alapvetően két megközelítést vázolnak a depresszió gépi tanulásra és hangelemzésre együttesen épülő detekciójának lehetőségeivel kapcsolatban. Az egyik, tradicionális megközelítést azok a tanulmányok képviselik, amelyek a beteg által előállított hanganyag mesterségesintelligencia-alapú algoritmusok segítségével történő elemzését végzik. Ezen tanulmányok célja, hogy meghatározzák azokat az akusztikai jellemzőket, amelyek átlagtól való eltérése a depresszió fennállására utal. Emellett azt is vizsgálják, hogy az említett jellemzőkre épülő predikciós modellek a tradicionális diagnosztikai

folyamatokkal összevetve mennyire pontosak^{3, 6-9}. Egy másik, talán szokatlanabb megközelítést képviselnek azok a tanulmányok, amelyek a mesterségesintelligencia-alapú algoritmusokat elsősorban a betegekkel kommunikálni képes, úgynevezett beszélgető ügynökök használata során alkalmazzák. Ezek az ügynökök virtuális asszisztensek, amelyek képesek emberi beszélgetést szimulálni, segítve a páciensekkel folytatott interakciók fenntartását. Többségük meglepő módon a mesterséges intelligencia nyújtotta lehetőségeket nem a hangelemzés, hanem a beszélgetés fenntarthatóságának biztosítása során igyekszik kihasználni, míg a betegek állapotfelmérésében továbbra is standardizált tesztek alkalmazását. A vonatkozó áttekintő tanulmány szerzői mindössze öt olyan tanulmányt azonosítottak, amelyben az ügynök nem valamely teszt, hanem a beteg beszédének elemzése alapján vont le következtetéseket a depresszió fennállására vonatkozóan¹⁰. A területen született meglepően magas számú tanulmány ellenére azonban ezen a területen csak egyetlen szisztematikus irodalmi áttekintést tudtunk azonosítani. Ez a munka sem nyilatkozik ugyanakkor az alkalmazott mesterségesintelligencia-kliensek hatékonyságáról, elsősorban azok technológiai leírására szorítkozik¹⁰. Ezért a továbbiakban a korábban ismertetett tradicionális megközelítés módszereit ismertetjük.

1. táblázat. A bevont közlemények elemzése AMSTAR 2 értékelőrendszerrel⁵

Közlemény	PICO	Protokoll	RCT/NRSI	Keresési stratégia	Duplikált szelekció	Duplikált beválasztás	Kizárt lista	Találatok leírása	RoB felmérés	Funding elemzés	Metaanalízis-statisztika	Metaanalízis-RoB felmérés	RoB interpretálás	Heterogenitás magyarázata	Kvantitatív-publikációs torzítás	Érdeklentét	Átfogó értékelés
6	I	RI	I	RI	N	N	N	I	N	N	NA	NA	N	I	NA	I	Közepes
7	I	I	I	RI	I	I	N	I	I	N	I	I	I	I	I	I	Magas
3	N	RI	I	N	N	N	N	I	N	N	NA	NA	N	I	NA	I	Közepes
8	I	RI	I	N	N	N	N	I	N	N	NA	NA	I	I	NA	I	Közepes
9	I	RI	I	N	N	N	N	I	N	N	NA	NA	N	I	NA	I	Közepes
10	N	RI	I	RI	N	N	N	I	N	N	NA	NA	N	I	NA	I	Alacsony

I: igen, N: nem, RI: részleges igen, NA: nem alkalmazható, PICO: Population, Intervention, Comparison, Outcome, RCT: randomized controlled trial/randomizált kontrollált vizsgálat, NRSI: non-randomized study interventions/nem randomizált elrendezés, RoB: risk of bias/elfogultsági kockázat

A géptanulás-alapú hangelemzés legfontosabb lépése maga az adatgyűjtés. E tekintetben számos kutatás támaszkodik nyilvánosan elérhető adatbázisokra, mint például az AVEC Audio/Visual Emotion Challenge Workshop vagy a DAIC-WOZ (Analysis Interview Corpus Wizard of Oz) adatbázis. Az áttekintő tanulmányok által vizsgált kutatások közül a gépi tanuló algoritmusokkal dolgozó munkák 60-70%-a, míg a mélytanuló architektúrák körülbelül 40%-a saját adatbázissal dolgozik, amelyek jelentősen eltérnek méretüket, a résztvevők demográfiai adatait, a depresszió értékeléséhez használt eszközöket, vagy a beszédrögzítés során alkalmazott feladatokat illetően. Noha az összefoglaló tanulmányok által megvizsgált közlemények az említett paraméterek mentén mintáikat részletesen jellemzik, sajnos szinte egyetlen esetben sem történik meg a minta komorbiditás mentén történő jellemzése, ami a minták jellemzőinek változékonysága mellett tovább nehezíti az eredmények összehasonlíthatóságát^{3, 6-9}.

A minta azonosításánál fontos kérdés, hogy a kutatók milyen eszközökkel méri fel a depressziós pácienseket. E tekintetben aranystandardnak a klinikai pszichológusok és pszichiáterek értékelésére épülő diagnózisalkotási folyamatot tekintik. Mivel azonban ez gyakran költséges folyamat, a kutatások többsége úgynevezett önkitöltős kérdőívek eredményei alapján határozza meg a betegcsoportokat. Ezek leggyakrabban a PHQ-9 (Patient Health Questionnaire-9) és a Beck Depression Inventory (Beck depresszió kérdőív) kérdőívek (mélytanuló architektúrák esetén ez az arány 90% fölötti, míg a géptanulás-alapú

modelleknél a modellek alacsonyabb információéhsége miatt alacsonyabb). Az áttekintő tanulmányok mindössze néhány olyan esetet azonosítottak, ahol a klinikus értékelését és az önkitöltős kérdőíveket egymással párhuzamosan alkalmazták^{3, 7-10}.

Hasonlóan nagy változékonyság jellemzi az audioadatok rögzítésénél használt feladatokat is. Ez lehet többek között strukturált interjú, szabad beszéd, szerepjáték, érzelmeket kiváltó történetek elmondása vagy felolvasási feladat. Az általunk bevont áttekintő tanulmányok alapján leggyakrabban szabályozott beszédkörnyezet alkalmazására, azon belül felolvasási feladat alkalmazására került sor, de meglehetősen gyakran használtak szabad szövegelemzést is, melynek során a spontán, kötetlen beszédből származó nyelvi mintákat és hangparamétereket vizsgálták. Az alkalmazott feladatok változékonysága mellett fontos megjegyezni, hogy azok felvétele minden alkalommal ellenőrzött körülmények között történt, jó minőségű eszközök használatával. Ez a megközelítés bár biztosítja az adatok megfelelő minőségét, erőforrásigénye miatt korlátozhatja a gyűjtött adatok mennyiségét. E tekintetben bár több szerző jutott arra a következtetésre, hogy az úgynevezett „off site”, azaz távoli eszközök, például mobiltelefonok, okosórák használatával történő, kevésbé ellenőrizhető adatgyűjtés fontos alternatívát kínálhat az adatmennyiség növelése érdekében, egyetlen általunk bevont áttekintő tanulmány sem azonosított „off site” eszközökkel végzett vizsgálatot. Ennek hátterében az utóbbi megközelítés sajátos kockázatai feltételezhetőek (például az adatminőség egyenetlensége)^{3, 6, 9}. Bár az áttekintő ta-

2. táblázat. *A bevont tanulmányok főbb jellemzői*

Hivatkozás	Áttekintő tanulmány típusa	Regisztráció	Vizsgált intervallum	Bevont tanulmányok száma	Ebből releváns	Vizsgált adatbázis	Vizsgálat fókusza	Vizsgált paraméterek	Főbb eredmények
⁶	Szisztematikus áttekintés	Nem regisztrált	2015–2021	43	5	Web of Science, Pubmed, IEEE Xplore	Átfogó áttekintés a neurológiai és pszichiátriai zavarok hangalapú felismerésére vonatkozó tanulmányokról.	Hangelemzés során használt feladatok típusa, alkalmazott algoritmusok, depresszióval kapcsolatba hozható akusztikai jellemzők	Leggyakrabban alkalmazott feladatok: felolvasás, spontán beszéd, leggyakoribb algoritmus: SVM, DT, RF, differenciálást segítő akusztikai jellemzők: shimmer, jitter, HNR, összszünethossz, f0 variabilitás.
⁷	Szisztematikus áttekintés metaanalízis	PROSPERO	2021–2024	25	25	Web of Science, Pubmed, IEEE Xplore, Medline, Embase, PsycINFO, Scopus	Deep-learning alapú beszédelemzésre épülő diagnosztikai rendszerek pontosságának felmerése depresszióban.	A deep-learning modellek becslési pontossága (szenzitivitás, specificitás, pontosság)	Összesített pontosság: 0,87; Összesített szenzitivitás: 0,87 Legmagasabb pontosság CNN-transformer modell esetén: 0,94.
³	Szisztematikus áttekintés	Nem regisztrált	2010–2020	127	55	Google Scholar	Automatizált beszédelemzésre épülő pszichiátriai diagnosztikai rendszerek vizsgálata, elsősorban a beszéd megkülönböztető jegyeinek azonosítása céljával.	Depresszióval kapcsolatba hozható akusztikai jellemzők	Differenciálást segítő akusztikai jellemzők: shimmer, jitter, HNR, f0 variabilitás, összszünethossz. A modellek becslési pontossága a mérések komplexitásából adódó overfitting miatt nem általánosítható.
⁸	Konceptcionális és kritikai áttekintés	Nem regisztrált	2017–2021	45	2	PubMed	A hangulatzavarok digitális fenotipizálására irányuló törekvések elemzése, ideértve a beszédelemzésre épülő diagnosztikát is.	Machine learning modellek becslési pontossága (szenzitivitás, specificitás, pontosság), depresszióval kapcsolatba hozható akusztikai jellemzők	Pontosság: 0,65–0,82; szenzitivitás: 0,65%; specificitás: 0,65%; differenciálást segítő akusztikai jellemzők: megnövekedett válaszadási latencia, szünetek gyakoriságának és hosszának növekedése.

Folytatás az előző oldalról

Hivatkozás	Áttekintő tanulmány típusa	Regisztráció	Vizsgált intervallum	Bevont tanulmányok száma	Ebből releváns	Vizsgált adatbázis	Vizsgálat fókusz	Vizsgált paraméterek	Főbb eredmények
⁹	Szisztematikus áttekintés	Nem regisztrált	2022–2023	264	18	Google Scholar	Géptanulás-alapú diagnosztikai rendszerek pontosságának felmérése depresszióban.	Depresszióval kapcsolatba hozható akusztikai jellemzők	Differenciálást segítő akusztikai jellemzők: pozitív korreláció: jitter, shimmer, HNR. Negatív korreláció: f0 átlag, f0 variabilitás, intenzitásátlag.
¹⁰	Szisztematikus áttekintés	Nem regisztrált	–2023	36	36	PubMed, IEEE Xplore, Psych-INFO, Scopus	Hangalapú interakciókra épülő beszélgető kliensek (például avatarok) detekcióban játszott szerepének felmérése depresszióban.	A beszélgető kliensek technikai jellemzőinek vizsgálata	A bevont tanulmányok 13,9%-a használt kizárólag hangalapú dialógusokat az elemzések során.

Jitter: a zöngé periódusainak időbeli ingadozását méri, jelezve a hangadás stabilitását vagy instabilitását; shimmer: a zöngé periódusainak amplitúdóbeli ingadozását méri, jelezve a hangadás stabilitását vagy instabilitását, SVM: támogató vektorgép (support vector machine), DT: döntési fa (decision tree), RF: véletlen erdő (random forest), HNR: harmonikus-zaj arány (Harmonics-to-Noise Ratio), f0: alaphangmagasság; CNN: konvolúciós neurális hálózat (Convolutional Neural Network)

nulmányok nagy gondot fordítanak az alkalmazott feladatok leírására, arról, hogy miként hathatnak az alkalmazott feladatok az algoritmusok predikciós pontosságára csak egy áttekintő tanulmány nyilatkozott, ami a vizsgált kutatások alapján a spontán beszédre épülő elemzést általában jobbnak találta⁹.

Az audioadatok elemzésének egyik első lépése a mintavételt követően az előfeldolgozás, ami magába foglalja a csend és zaj eltávolítását, a jel normalizálását, a zajszűrést és az esetleges torzítások korrigálását. Ez a folyamat többek között digitális jelfeldolgozó technikák, szoftveres zajszűrő algoritmusok segítségével zajlik. Az általunk vizsgált összefoglaló tanulmányok azonban az általuk vizsgált kutatásokat e tekintetben részletesen nem elemzik. Nagyobb hangsúly helyeződik ugyanakkor a releváns akusztikai jellemzők kinyerésére (feature extraction). Az összefoglaló tanulmányok alapján ehhez leggyakrabban nyílt forráskódú platformokat (openSMILE, COVAREP és openEAR), illetve Python-alapú könyvtárat alkalmaznak (pyAudioAnalysis) a kutatók. A leggyakrabban kinyert/vizsgált akusztikai jellemzőket, azok rövid meghatározásával a **3. táblázat** részletezi^{3, 6, 9}.

Az akusztikai jellemzők kinyerését követően a következő lépés a jellemzők elemzése és az algoritmusok tanítása, ami magába foglalja a géptanulás-alapú modellek kiválasztását és az adatok előkészítését a modellek számára.

E tekintetben a depresszió kutatások során leggyakrabban géptanulás-architektúrák, azon belül is SVM (Support vector machine) használatára kerül sor. Viszonylag ritkán alkalmaznak mélytanulási architektúrákat. Ennek oka valószínűleg az, hogy a mélytanulási modellek adatigénye nagyobb, míg például az SVM kis adathalmazon is képes osztályok elkülönítésére. Az SVM mellett a kutatók más gépi tanulási modelleket is előszeretettel alkalmaznak, ilyen a DT (decision trees), ami átlátható és interpretálható, klinikusok és kutatók számára könnyen érthető modelleket kínál, vagy az RF (random forest), ami a magas zajarányjal jellemezhető adatokat is jól tolerálja, emellett a DT-ekkel összehasonlítva csökkenti az overfitting kockázatát^{3, 6, 7, 9}. A mélytanuló architektúrák közül az úgynevezett CNN (Convolution Neural Network) alkalmazása volt a leggyakoribb. Ennek oka valószínűleg az, hogy a CNN a nyers audioadatokból komplex mintázatok (például spektrális és időbeli jellemzők) kinyerésére képes anélkül, hogy kézzel definiált jellemzőket kellene használni; emellett rétegzetten lehet felépíteni, így komplexebb összefüggések megtanulására és modellezésére is képes (például intonációs mintázat)⁷. Az egyes architektúrák közötti különbségek nem azt jelentik, hogy a mélytanuló algoritmus minden esetben jobb választás. A modellek kiválasztása ugyanis gyakran attól függ, milyen típusú audio-jellemzőkkel dolgoznak a kutatók.

3. táblázat. *A depresszióval összefüggésben vizsgált akusztikai jellemzők listája és azok rövid meghatározása (a táblázat nem terjed ki valamennyi ismert akusztikai jellemző ismertetésére, a leggyakrabban vizsgált paramétereket kiemeltük)*

Jitter	A zöngé periódusainak időbeli ingadozását méri, jelezve a hangadás stabilitását vagy instabilitását.
Shimmer	A zöngé periódusainak amplitúdóbeli ingadozását méri, jelezve a hangadás stabilitását vagy instabilitását.
Tremor	A hangszalagok rezgésének frekvenciabeli ingadozását méri, amely a hang stabilitásának periodikus változásait jellemzi.
Harmonics-to-noise ratio (HNR)	Az alaphfrekvencia (f_0) és a zajenergia arányát méri a hangban, és a hang tisztaságát, valamint a vokális zaj jelenlétét jellemzi.
Quasi-open quotient (QOQ)	A hang minőségét és természetességét értékeli különböző paraméterek, például a hangszín és a tisztaság alapján.
Normalized Amplitude Quotient (NAQ)	A beszéd érthetősége és a háttérzaj arányát méri, ezzel segítve a beszéd minőségének értékelését.
Peak slope	A hanghullám csúcsának emelkedési sebességét méri, amely a hang intenzitásának gyors változását jelzi.
F1 átlag	Első formáns frekvencia átlaga.
F2 átlag	Második formáns frekvencia átlaga.
F1 variabilitás	Az első formáns (F1) frekvenciájának szórása.
F2 variabilitás	A második formáns (F2) frekvenciájának szórása.
F1 tartomány	Az első formáns frekvenciáinak intervalluma.
Vowel space	A vowel space (magánhangzó-tér) egy akusztikai tér, amely a magánhangzók formáns-frekvenciáinak (F1 és F2) alapján mutatja a nyelv helyzetét a beszéd során, és vizuálisan ábrázolja a különböző magánhangzók képzését.
Mel Frequency Cepstral Coefficients (MFCC)	A beszéd Mel-frekvenciák szerinti sávokra bontott intenzitás értékeinek a diszkrét koszinusz transzformációjával nyert együtthatói.
f_0 átlag	A beszédhang alaphangjának átlagos értéke.
f_0 tartomány	Beszédhang alaphangjának legkisebb és legnagyobb értékei közötti intervallum.
f_0 variabilitás	Beszédhang alaphangjának szórása.
Intenzitásátlag	A hangenergia (hangnyomás) időbeli átlagolt mértéke.
Intenzitásvariabilitás	A hang intenzitásának időbeli változékonyságát jelenti.
Energiasebesség	A beszédjel intenzitásának időbeli változását méri, vagyis azt mutatja meg, hogy a hangenergia milyen gyorsan ingadozik a beszéd során.
Energiaváltozékonyság	A hanghullámok által közvetített energia változékonyságát jelenti.
Beszédsebesség	Másodpercenkénti beszédhangok száma egy adott beszédszakaszon, a szüneteket is figyelembe véve.
Artikulációs sebesség	Másodpercenkénti beszédhangok száma egy adott beszédszakaszon a szünetek nélkül.
Átlagos szünethossz	A beszédben előforduló szünetek átlagos időtartama.
Szünetvariabilitás	A beszédben előforduló szünetek időtartamának szórása.
Szünetgyakoriság	A beszédben előforduló szünetek és nem szünetek időbeli aránya.
Össz-szünethossz	A beszédben előforduló szünetek összhosszát jelenti.

A modellek bemeneteként az adatkészlet gazdagítása érdekében általában javasolt különböző hosszúságú beszédrészeket használni, mivel a beszédrészek hossza befolyásolja, hogy milyen részleteket tud a modell megra-

gadni. Arról azonban, hogy a bevont tanulmányok e tekintetben milyen gyakorlatot követnek, csak egy tanulmány nyilatkozott⁷. Ezen áttekintés alapján a legjobb teljesítményt (pontosság: 0,94, szenzitivitás: 0,92) 9 másodperces

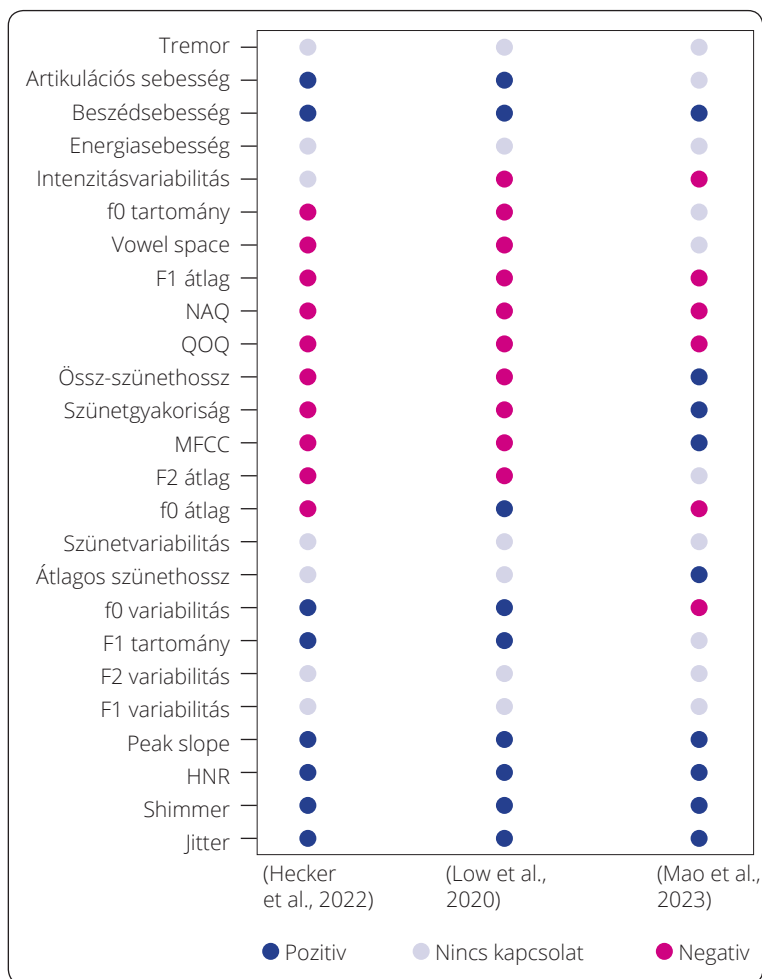
szegmensekkel érték el, általánosságban pedig az 5 másodpercnél hosszabb szegmensek alkalmazásának előnye ábrázolódik⁷.

A modelltréníngezés előkészítésének fontos, de gyakran nem kellő mértékben hangsúlyozott része a hiperparametrizálás. Ez tulajdonképpen az alkalmazott algoritmusok előre meghatározott beállítását jelenti. Ezeket a változókat optimális esetben nem az adatokból tanulják meg a modell tréningezése során, hanem előre meghatározzák. Ebben a folyamatban gyakran iteratív tréningezési eljárást alkalmaznak, ahol az optimális teljesítmény megtalálása érdekében különböző beállításokat próbálnak ki. Ezek a beállítások nagy mértékben befolyásolják, hogy a beszédmintázatból az adott modell milyen gyorsan és hatékonyan képes tanulni. Ilyen hiperparaméter lehet például az úgynevezett epochszám (hány teljes ciklust végez a modell az adatokon a tréningezése során) vagy a CNN-modell esetén a konvolúciós rétegek száma. Sajnos, bár több összefoglaló tanulmány kiter a hiperparametrizálás fontosságára, azok mentén mintájukat kevésbé elemzik^{3, 6, 7, 9}. Ez alól kivételt képez Liu és munkatársai⁷ tanulmánya, ami rámutat, hogy a finomhangolás fontossága ellenére sok tanulmányból hiányzik ezen paraméterek pontos leírása.

A kifejlesztett modellek teljesítményének értékelése során a kutatások változatos mutatókat alkalmaznak. Ezek között leggyakrabban (több mint az esetek 80%-a) a pontosság, érzékenység, specificitás és az F1 érték használata. Valamivel ritkábban jellemzik a teljesítményt AUC (görbe alatti terület), CCC (concordance correlation coefficient), RMSE (root mean square error/átlagos négyzetes hiba négyzetgyöke) és MAE (mean square error/átlagos négyzetes hiba) segítségével^{6, 7, 9}. E tekintetben az áttekintő tanulmányok többsége igyekszik az összehasonlíthatóság kedvéért egyidejűleg több mutató mentén jellemezni a modellek teljesítményét. A modellek pontosságának jellemzésével később részletesebben is foglalkozunk.

A depresszióval összefüggésbe hozható akusztikai jellemzők

Az általunk bevont közlemények összegzése alapján számos akusztikai jellemző prediktív értéke megerősítést nyert. Ezek a jellemzők lehetőséget nyújtanak a depresszió megfigyelhető, hallható, gyakran csupán szubjektíven észlelhető jegyeinek objektivizálására. A depresszió



2. ábra. A leggyakrabban vizsgált akusztikai jellemzők és a depresszió összefüggéseinek grafikus összefoglalása

'Pozitív kapcsolat': Az egyik változó növekedése a másik változó növekedésével jár együtt. 'Negatív kapcsolat': Az egyik változó növekedése a másik változó csökkenésével jár együtt. 'Nincs kapcsolat': Az egyik változó változása nem befolyásolja a másik változót. '-' : Az összefüggést a bevont tanulmányok nem vizsgálták

esetében általában és leggyakrabban vizsgált jellemzőket a 2. táblázat mutatja be.

A 2. ábra jól szemlélteti, hogy a depresszió és az akusztikai jellemzők között kapcsolatot kereső összefoglaló tanulmányok következtetései nagy mértékben összhangban állnak egymással. Például az f0 tartomány (meghatározása a 2. táblázatban található) csökkenését depresszió esetén következetesen megfigyelték, ami jól tükrözi, és mérhetővé teszi a depresszióban gyakran szubjektíven is észlelhető monoton beszédet. Hasonlóan, a vowel space, az F1-átlag, a NAQ és QOQ változókkal való negatív kapcsolat a betegeknél tapasztalható meglás-szubbodott gondolkodást, csökkent reakcióidőt tükrözheti. Ezzel párhuzamosan megfigyelték, hogy más akusztikai jellemzők, mint például a jitter, a shimmer, az úgynevezett Peak slope, vagy a HNR értéke általában a depresszió

súlyosságával együtt emelkedik, a gondolkodás, a mozgás és a reakcióidő lassulása mellett a beszédminőség romlására, valamint a hangzásbeli variabilitás csökkenésére utalva^{3, 6, 9}.

A többnyire egy irányba mutató eredmények ellenére néhány esetben azok ellentmondása látható (például MFCC, átlagos szünethossz, szünetgyakoriság). Ennek kapcsán fontos megjegyezni, hogy a depresszió egy spektrumot ölel fel, amennyiben a depressziós tünetek és megnyilvánulások széles skálán mozoghatnak, függően többek között a betegség súlyosságától vagy a komorbid állapotoktól. Így egyes akusztikai jellemzők a depressziós állapotokkal komplex interakciót mutathatnak. Ennek megfelelően a különböző akusztikai minták és a hangzásbeli variabilitás a személyek aktuális érzelmi állapotától, a beszédkörnyezetüktől és a kommunikációs szándéktól függően is változhat. Erre lehet példa az MFCC-jellemzők és a depresszió kapcsolatában talált ellentmondás. E tekintetben pozitív és negatív összefüggést egyaránt tükröző eredmények születtek. Ez a mutató ugyanakkor érzékeny lehet a beszélők érzelmi állapotára és a kommunikációs kontextusra, ami így ellentmondásos eredményekhez vezethet. Az össz-szünethossz és a szünetgyakoriság esetén hasonló ellentmondás tapasztalható, ahol a depressziós tünetek súlyosbodásával párhuzamosan mind a szünethossz és szünetgyakoriság növekedésére, mind csökkenésére utaló eredményeket publikáltak. Az eredmények azonban nem feltétlenül zárják ki egymást. Bár a hosszabb szünethossz és a gyakoribb szünetek általánosságban utalhatnak arra, hogy a depressziós személyek nehezebben formálják a gondolatokat, arra is felhívják a figyelmet, hogy a depressziós személyek feszültebbek lehetnek, így hajlamosabbá válhatnak gyorsabban beszélni, vagy a szünetek elkerülésével kompenzálni a gondolkodási nehézségeiket.

Ezzel együtt a különböző tanulmányok eltérő módszertani megközelítése is tükröződhet az eredményekben. Általában elmondható, hogy a kisebb mintákra épülő következtetések instabilabb eredményeket hozhatnak. Esetünkben azonban ez az összefüggés csak az f_0 átlag eltéréseinek magyarázatánál jöhet szóba, mivel a további változók kapcsán a legkisebb és legmagasabb mintaelemszámmal dolgozó összefoglaló közlemények következtetései konzisztensek. Az eltérések döntően a közepes mintaelemszámmal dolgozó összefoglaló tanulmány esetén megfigyelhetők. A hangfelvételi és elemzési technikák változékonysága az összefoglaló közleményekbe bevont tanulmányok esetén újabb magyarázatot kínálhat az eltérések értelmezésére. Azonban az elemzésünkbe bevont közlemények közül mindössze egy nyilatkozik a bevont kutatásokban alkalmazott mérési módszerekről¹⁶. Végül az áttekintett tanulmányokban vizsgált klinikai csoportok jellemzői (például életkor, nem, beszélt nyelv, alkalmazott gyógyszeres kezelés, egyéb komorbiditások) is hatással lehetnek a megfigyelhető akusztikai mintákra^{3, 6, 9, 10}. Ezen paraméterek részletes elemzése azonban kívül esik az általunk bevont közlemények

hatáskörén. *Hecker* és munkatársai⁶ ugyanakkor megjegyzik, hogy sajnos sok esetben ezen releváns confounding faktorok kontrollálása elkerüli a szerzők figyelmét. Összességében ugyanakkor az ellentmondások háttérében feltételezhető okok sokfélesége fontos iránymutatásként szolgálhat jövőbeli kutatások számára. E tekintetben a standardizált módszertani keretek kidolgozása, valamint a különböző depressziós megnyilvánulások és akusztikai jellemzők kapcsolatának mélyebb és árnyaltabb feltérképezése egyaránt fontos szempont lehet a mesterségesintelligencia-alapú rendszerek jövőbeli vizsgálatakor^{3, 6, 9}.

A prediktív modellek teljesítménye

A géptanulás-alapú prediktív modellek teljesítménye a vizsgált közlemények alapján meglehetősen széles tartományban mozog. Géptanulás-modelleknél az átlagos abszolút hiba (MAE): 5,77–7,08; pontosság: 0,75–0,86; mélytanuló architektúráknál pontosság: 0,57–0,99; duplikált adatkészletek kizárását követően pontosság: 0,75–0,9)^{3,7,8}. Ennek megértéséhez érdemes megemlíteni, hogy a gépi tanulási modellek validálását alapvetően két megközelítés határozza meg, az úgynevezett „held-out” (azon belül például hold-out validation) és a „cross-validation” (azon belül például K-fold cross-validation vagy nested cross-validation) megközelítések. Előbbi esetben jellemzően a mintát két részre osztják: egy tréning- (tanulási) halmazra és egy teszt- (ellenőrzési) halmazra, ahol a teszt-halmazt csak egyszer használják a modell teljesítményének értékeléséhez. A cross-validation során ezzel szemben a mintát több részre osztják, és a mérést többször ismétlik. A modellt két alkalommal tanítják és tesztelik, és minden alkalommal más-más halmazt tartanak fenn a tesztelésre, míg a fennmaradó adathalmazt a tréninghez használják. A két megközelítés eltéréseiből adódóan a held-out tesztelési módszer előnye, hogy egyszerűbb és gyorsabb, ám az egyszeri felosztás és a véletlenszerűség miatt a teljesítménybecslés – különösen kis adathalmazon – ingadozhat. A cross-validation megközelítés ezzel szemben részletesebb és megbízhatóbb értékelést nyújt kis adathalmazon is. Ez a megközelítés jól alkalmazható az úgynevezett overfitting elkerüléséhez is (amikor egy modell túlságosan illeszkedik a tréningadatokra, ideértve azok zajait és speciális mintázatait, ezért kevésbé teljesít jól ismeretlen adathalmazon). Fontos azonban megjegyezni, hogy annak elkerülését nem garantálja: ha a modell túlságosan bonyolult, vagy ha a tréningadatokban magas a zajarány, az overfitting veszélye ugyanúgy fennállhat^{3, 7, 9}.

Low és munkatársai³ megjegyzik, hogy a depresszió esetén alkalmazott modellek meglehetősen kis arányban alkalmaznak held-out megközelítést. A szerzők azt is megfigyelték, hogy a jellemzően cross-validation megközelítéssel dolgozó tanulmányok hajlamosak ugyanazt a nyílt audioadatbázist (például AVEC) alkalmazni a modelljük tesztelésékor, mely gyakorlat segíthet a prediktív mo-

dellek teljesítményének különböző tanulmányok közötti összehasonlításában. A validációs folyamat eredménye ezen tanulmányok esetén a prediktív modellek teljesítményének csökkenését mutatták tesztadathalmazon történő alkalmazás esetén³. A szerzők azonban az egyes tanulmányok eredményeinek tételes összehasonlítása alapján az alkalmazott modellek komplexitását nem találták a teljesítményeltérés elégséges magyarázatának. A bevont tanulmányok tételes elemzése alapján a teljesítmény szempontjából a mintanagyság, a mintavétel körülményei és eszköze, a zavaró tényezők kontrollálása (ez sajnos a legtöbb tanulmány esetében elmarad), a minta további jellemzői (nem, beszélt nyelv, életkor, kulturális háttér), valamint a minta előfeldolgozása egyaránt és együttesen bizonyult meghatározónak. Bár az említett tényezők egyedi hozzájárulása nehezen megítélhető, bizonyos eljárások előnye kirajzolódik a bevont összefoglaló tanulmányok alapján. Ilyen például az akusztikai jellemzők kinyerése során alkalmazott eljárások közül az LLD (low-level descriptor), vagy az MFCC (Mel-Frequency Cepstral Coefficients), melyek alkalmazása mélytanuló architektúrákkal együtt következetesen 80% fölötti predikciós pontosságot eredményezett⁷. Emellett, bár csak néhány tanulmány vizsgálta, mégis egyértelműnek tűnik a multimodális jellemzők használatának teljesítményjavító hatása, illetve összességében a mélytanuló architektúrák fölénye géptanulás-alapú társaikkal szemben^{3, 7-9}.

Megbeszélés

Ismerteink szerint ez az első, a depresszió diagnosztikájában alkalmazott automatizált hangelemzés területén született, úgynevezett szisztematikus umbrella review. Természetéből adódóan ez a review típus szélesebb perspektívát nyújthat a kutatási eredmények értékelésére és az adott területen trendek azonosítására.

A bevont közlemények következtetései alapján számos akusztikai jellemző, mint például az f0 tartomány, a jitter és a shimmer prediktív értéke megerősítést nyert a depresszió objektívizálásában. Ezek a jellemzők lehetővé teszik, hogy a depresszió szubjektíven észlelhető, de a szakmai gyakorlatban, különösen az alapellátásban gyakran figyelmen kívül hagyott aspektusai, mint a monoton beszéd vagy a lassuló reakcióidő gyorsan és automatikusan mérhetővé váljanak. A depresszióval asszociált akusztikai jellemzők vonatkozásában a legtöbb tanulmány eredményei összhangban állnak egymással, a prediktív modellek teljesítménye azonban jelentősen eltér, ami azt mutatja, hogy a jellemzően a tréningmintára jól illeszkedő modellek generalizálhatósága megkérdőjelezhető.

Ideális esetben a géptanulás-alapú predikciós modelleknek képesnek kell lenni felismerni egy adott rendellenességet, olyan személyeknél is, akik nem voltak tagjai a tréningmintának. A teljesítménynek annak érdekében, hogy a modellek különböző környezetben is

megbízhatóan működjenek, optimális esetben függetlennek kell lennie a személy környezetétől, életkorától, akcentusától, a háttérzajtól, a felvételi eszközöktől vagy a komorbiditástól. Jelenleg azonban a rendelkezésre álló modellek esetén ez a generalizáció csak korlátozott mértékben valósítható meg⁷.

Ezért a jelen összefoglaló tanulmány nemcsak a korábbi kutatási eredmények áttekintésére törekszik, hanem azok következtetlenségeinek feltárására is. Célja, hogy az ellentmondások mögött húzódó lehetséges tényezőket azonosítva a jövőbeli kutatások számára konstruktív javaslatokat fogalmazzon meg. Ezek egy részére már történt utalás az eredmények ismertetésében, amennyiben az eredmények eseti inkonzisztenciája rámutatott a standardizált módszertani keretek kidolgozásának fontosságára.

Kirajzolódott az is, hogy a hagyományos klinikai vizsgálatokban a hangfelvételek rögzítésére leggyakrabban ellenőrzött környezetben került sor. Ez a megközelítés ugyanakkor korlátozhatja a mintanagyságot, és kérdéseket vethet fel az eredmények általánosíthatóságával kapcsolatban. Ere mintegy válaszként az utóbbi időben az úgynevezett távoli, off-site adatgyűjtésre is történtek kísérletek, melyek többek között és elsősorban okostelefonokat, okosórákat használnak. Ezek az eszközök lehetővé teszik a természetes környezetben történő adatgyűjtést, ezzel együtt jelentősen megnövelik a begyűjtött adatok mennyiségét. Ezek az adatgyűjtési módszerek a depresszió alakulásának hosszú távú követését is elősegítik, mivel akár többszöri, ismételt mintavétel lehetőségét is felkínálják, a felhasználók érdemi terhelése nélkül. A mindennapi életben történő adatgyűjtés azonban pont a standardizált módszertani keretek kidolgozásának igényével megy szembe, számos kihívás elé állítva a kutatókat. Ere példa, hogy kontrollált körülmények között a kutatók kiváló minőségű eszközökkel gyűjtenek adatokat. Off site adatgyűjtés miatt azonban az okostelefonok, okosórák sokfélesége többek között eltérő minőségű mikrofonok alkalmazását jelentheti, ami jelentős hatást gyakorolhat az adatok minőségére. A lehetséges nehézségek ellenére a mesterséges intelligencia egészségügyi alkalmazását vizsgáló kutatók jelentős része egyetért abban, hogy a jövőbeli kutatási trendekben a természetes környezetből származó adatok elemzését érdemes fókuszba helyezni, noha ez a trend a depresszió esetén még a közelmúltbeli kutatásokban sem érvényesült^{3, 6, 11}.

Ismert emellett, hogy a géptanulás-alapú prediktív modellek becslési pontosságát javíthatja, ha azok többféle modalitásból származó adatokat használnak. Depresszió esetén például az audioadatok elemzését videofelvételek elemzése egészítheti ki. Ez a megközelítés már régóta ismert, a depresszió területén azonban csak ritkán találkozunk olyan vizsgálattal, ami ezt a trendet követi. Ennek valószínűleg az adatelemzés összetettebbé válása áll a háttérben. A különböző adatforrásból származó információk ugyanis eltérő zajszintet, minőséget és rep-

rezentációt hordozhatnak, ami megnehezítheti az adatok egységes elemzését és értelmezését¹². Az utóbbi folyamat a különböző modalitások közötti eltéréseket kezelni képes algoritmusok és technikák alkalmazását is megköveteli, ami további szakmai kihívást jelenthet^{3, 6-8}.

A pszichiátriai kutatások a mesterséges intelligencia alkalmazásában az egyik legnagyobb előnyként a pszichiátriai zavarok objektív újradefiniálásának lehetőségét látják. Ennek ellenére azonban a kutatásokban döntően továbbra is a pszichiátriai zavarok meghatározásának kategoriális szemlélete érvényesül, amennyiben a géptanulás-alapú algoritmusokat nem egy-egy tünet vagy probléma észlelésére, hanem meghatározott szindrómák detekciójára tréningezik. Kérdés azonban, hogy nem lenne-e kívánatosabb az akusztikai jellemzőket olyan specifikus tünetekkel vagy problémákkal asszociálni, melyek több pszichiátriai zavarban megjelenhetnek³.

A módszertani szempontok átgondolásának szükségesén túl a bevont közlemények arra is rámutattak, hogy a jövőbeli kutatások során érdemes alaposan megfontolni az adatelemzési eszközök megválasztását. E tekintetben a feldolgozott közlemények alapján többek között az alábbi szempontok merülnek fel:

- Az overfitting elkerülése érdekében javasolt a tanulás és tesztkészleteken az előfeldolgozást külön elvégezni. Ennek hiányában olyan információ juthat a tesztalmazba, ami már a tréninghalmazból származik, ez pedig mesterségesen növelheti a modell teljesítményét, mivel így a két tesztkészlet már nem teljesen független³.

- Javasolt emellett az akusztikai jellemzők kinyerésének (feature extraction) standardizálása az eredmények tanulmányok közötti összehasonlíthatósága érdekében.

- Javasolt a hiperparaméterek pontos megadása a reprodukálhatóság és az átláthatóság érdekében^{3, 7}.

- A modellek hatékonyabb finomhangolását segítheti a kódok megosztása⁷.

- Kis mintaméret esetén javasolt a „held out” megközelítés kerülése. Ehelyett érdemes a bootstrapping módszer alkalmazása. Összetettebb modellek, mint neurális hálók

és nagyobb mintaelemszám esetén „cross-validation” megközelítés javasolt^{3, 6, 7, 9}.

- Az összehasonlíthatóság érdekében összességében hasznos a becslési pontosságot több mutató mentén jelezni, valamint az interpretálás folyamatában a feltárt akusztikai jellemzők eltéréseit pszichiátriai tünetekkel asszociálni^{3, 6, 9}.

Adatvédelmi szempontok

A géptanulás-alapú elemzésekkel dolgozó egészségügyi alkalmazások esetén kiemelt figyelmet kell fordítani az adatvédelmi előírásokra^{13, 14}. Az off site adatgyűjtés során számos személyes és érzékeny adat gyűjtése valósul meg, ami komoly adatvédelmi problémákat vet fel. Ezért, bár első pillantásra ígéretesnek tűnhet az olyan hangasszisztensek, mint Siri vagy Alexa alkalmazása olyan pszichiátriai zavarok, mint a depresszió gyors azonosítására, egészségügyi kontextusban az adatvédelmi szigor jelentős, ezért a kutatásoknak a hatályos adatvédelmi jogszabályokkal összhangban, világos adatkezelési irányelvek kidolgozása mellett, lehetőség szerint dedikált platformokon keresztül (weboldalak, engedélyezett alkalmazások) kell történnie^{3, 6, 13, 14}.

Összességében a depresszió géptanulás-alapú detekciója máig alulkutatott területnek tekinthető, amit a kevés, változó minőségű összefoglaló tanulmány is jól mutat. Ettől függetlenül a gépi tanulás alkalmazása a depresszió diagnosztizálásában számos előnnyel kecsegtet, hosszú távon a zavar költséghatékony, objektív, gyors és korai felismerését kínálva. A fenti szempontok alapján ugyanakkor fontosnak tartjuk további kutatások végzését, valamint a keresztmetszeti kutatások hosszmetzeti megközelítésekkel való kiegészítését. Az utóbbiak révén az akusztikai jellemzők és a depressziós tünetek mélyebb összefüggésrendszere válhat megismerhetővé. A hosszmetzeti kutatások nemcsak a tünetek időbeli változásait követhetik nyomon, hanem lehetőséget adnak különböző kezelési formák hatásának hosszú távú mérésére is.

References

1. WHO. Depressive disorder (depression). <https://www.who.int/news-room/fact-sheets/detail/depression>.
2. Koops S, Brederoo SG, de Boer JN, Nadema FG, Voppel AE, Sommer IE. Speech as a biomarker for depression. *CNS Neurol Disord Drug Targets* 2023;22(2):152-60. <https://doi.org/10.2174/1871527320666211213125847>
3. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Invest Otolaryngol* 2020;5(1):96-116. <https://doi.org/10.1002/lio2.354>
4. Almaghrabi SA, Clark SR, Baumert M. Bio-acoustic features of depression: A review. *Biomed Signal Process Control* 2023;85:105020. <https://doi.org/10.1016/j.bspc.2023.105020>
5. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;j4008. <https://doi.org/10.1136/bmj.j4008>
6. Hecker P, Steckhan N, Eyben F, Schuller BW, Arnrich B. Voice

- analysis for neurological disorder recognition—a systematic review and perspective on emerging trends. *Front Digit Health* 2022;4. <https://doi.org/10.3389/fdgth.2022.842301>
7. Liu L, Liu L, Wafa HA, Tydeman F, Xie W, Wang Y. Diagnostic accuracy of deep learning using speech samples in depression: a systematic review and meta-analysis. *Journal of the American Medical Informatics Association* 2024;31(10):2394–404. <https://doi.org/10.1093/jamia/ocae189>
 8. Maatoug R, Oudin A, Adrien V, et al. Digital phenotype of mood disorders: A conceptual and critical review. *Front Psychiatry* 2022;13. <https://doi.org/10.3389/fpsyt.2022.895860>
 9. Mao K, Wu Y, Chen J. A systematic review on automated clinical depression diagnosis. *npj Mental Health Research* 2023;2(1):20. <https://doi.org/10.1038/s44184-023-00040-z>
 10. Otero-González I, Pacheco-Lorenzo MR, Fernández-Iglesias MJ, Anido-Rifón LE. Conversational agents for depression screening: A systematic review. *Int J Med Inform* 2024;181:105272. <https://doi.org/10.1016/j.ijmedinf.2023.105272>
 11. Ermers NJ, Hagoort K, Scheepers FE. The predictive validity of machine learning models in the classification and treatment of major depressive disorder: State of the art and future directions. *Front Psychiatry* 2020;11. <https://doi.org/10.3389/fpsyt.2020.00472>
 12. Yang L, Jiang D, Xia X, Pei E, Oveneke MC, Sahli H. Multimodal measurement of depression using deep learning models. In: *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*. ACM 2017:53–9. <https://doi.org/10.1145/3133944.3133948>
 13. Karizat N, Vinson AH, Parthasarathy S, Andalibi N. Patent applications as glimpses into the sociotechnical imaginary: Ethical speculation on the imagined futures of emotion AI for mental health monitoring and detection. *Proc ACM Hum Comput Interact* 2024;8(CSCW1):1–43. <https://doi.org/10.1145/3637383>
 14. Vitale F, Carbonaro B, Cordasco G, et al. A privacy-oriented approach for depression signs detection based on speech analysis. *Electronics (Basel)* 2021;10(23):2986. <https://doi.org/10.3390/electronics10232986>