

Szeged, 2017. január 26–27.

125

Depresszió súlyosságának becslése beszédjel alapján magyar nyelven

Gábor Kiss¹, Lajos Simon², Klára Vicsi¹

¹ Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
{kiss.gabor, vicsi}@tmit.bme.hu

² Semmelweis Orvostudományi Egyetem Pszichiátriai és Pszichoterápiás Klinika,
simon.lajos@med.semmelweis-univ.hu

Kivonat: A depresszió korunk egyik legelterjedtebb, gyógyítható betegsége, ám diagnosztizálása szaktudást igényel, és így a kórkép felállítása a társadalom egy szűk rétegére hárul. A depresszió súlyossága nagyban befolyásolja az ebben szenvedő beteg életminőséget. Depresszió hatására megváltoznak az emberi beszédproduktum egyes jellemzői, amelyek számszerűsíthetőek és mérhetőek. Emiatt lehetőség nyílik a depresszió beszédjel alapú detektálásra, ami megkönnyítheti, illetve szélesebb körben lehetővé teheti a betegség diagnosztizálását. Ezen okok miatt fontos kutatási terület a depressziós állapot beszédjel alapú felismerése és súlyosságának becslése. Ebben a cikkben bemutatunk egy Szupport Vektor Regressziós számításon alapuló automatikus rendszert, ami képes a beszédjel alapján megbecsülni nemcsak a depresszió meglétét, hanem a beszélő állapotának súlyosságát is. Megvizsgáljuk, hogyan változik a rendszer pontossága, ha külön rendszert alkalmazunk a nők és a férfiak esetén, illetve ha felhasználjuk a beszéd fonéma szintű szegmentálását a beszédet leíró jellemzők előállításánál.

Kulcsszavak: depresszió, beszédjel alapú detektálás, beszédsegmentálás, regresszió, SVR

1 Bevezetés

Az emberi beszédproduktum sokrétű jelentést hordoz, így nem csupán a beszéd nyelvi tartalmát közvetíti, hanem számos, a kommunikációval kapcsolatos nonverbális üzenetet is hordoz, mint például a beszélő érzelmi töltete, és mindezek mellett a beszélőre jellemző fiziológiai állapotot is tükrözi. Pszichiátriai szakorvosok állítják, hogy a páciens beszéde alapján képesek annak pszichofiziológiai állapotát felmérni, így például a depressziót is. A depressziós betegek beszédét a szakorvosok a következő jellemzőkkel szokták leírni: fakó, monoton, élettelen. Természetesen ezek az érzeti jellemzők számszerűsíthetőek, és így kapcsolatba hozhatók a beszéd egyes akusztikai és fonetikai jellemzőivel, mint például alapfrekvencia, formáns frekvenciák, beszédtempó stb. Ezt a jelenséget már 1921-ben megfigyelte és publikálta Emil Kraepelin, a modern pszichiátria egyik megalapozója [10].

A WHO (World Health Organization) 350 millióra becsülte a depresszióban szenvedő betegek számát 2012-ben [11]. A WHO előrejelzései szerint 2030-ra a depresszió

a három legsúlyosabb betegség között lesz világviszonylatban a HIV/AIDS vírus és a szívbetegségek mellett [12]. Annak ellenére, hogy a betegségben szenvedők száma igen magas, a diagnózis felállítása egy kisszámú képzett szakorvosrétegre hárul. A depressziós betegek életminőségé a depresszió következtében és hatására erősen romlik, a tünetek súlyosságától függően akár képtelenek rendszeresen dolgozni, ami komoly gazdasági problémát jelent a társadalomnak. Ráadásul a súlyos depresszió megnöveli az öngyilkossági kockázatot is [3].

Ezek miatt hasznos lenne egy olyan objektív, robosztus diagnosztizáló rendszer kialakítása, amelyet akár nem szakképzett orvosok is használhatnának a felismerésére és követésére.

A depresszió és a beszéd kapcsolata már az 1980-as évektől kezdve fontos kutatási területnek számít, és több akusztikai illetve fonetikai paramétert kapcsolatba hoztak a depresszióval, mint például az átlagos alaphangfrekvencia értéket, az alaphangfrekvencia tartományát, beszédtempót [15]. Azonban a depresszió gépi detektálása új kutatási területnek számít, amit az informatika fejlődése tett lehetővé. Cummins és társai 2015-ben a *Speech Communication* folyóiratban közöltek egy átfogó tanulmányt a beszédjelalapú depresszió detektálásához kapcsolódó fontosabb kutatások legfrissebb eredményeiről [6].

Alapvetően kétféle gépi detektálási módszert alkalmaznak a kutatók: osztályozó eljárást, amely a beszélő depressziós állapotát detektálja, illetve regressziós eljárást, amely megbecsüli a depresszió súlyosságát. Ami közös bennük, hogy mindkét eljárás-hoz szükség van valamilyen orvosi besorolási rendszerre. A két legelterjedtebb besorolási rendszer a *Hamilton Rating Scale for Depression (HAMD)* [7] és a *Beck Depression Index (BDI)* [1]. Mi ebben a cikkben a BDI továbbfejlesztett változatát használjuk, a BDI-II skálát [1].

A depressziós állapot az agy motorikus működését befolyásolja, emiatt változik a depressziós ember beszédproduktuma. Elsősorban természetesen az a kérdés, hogy a depresszió hatására mely akusztikai és fonetikai jellemzők változnak meg. A nemzetközi irodalom több beszédparamétert is említ, ami a depresszió hatására megváltozik. Azonban abban még nincs általános egyezés, hogy az adott beszédparamétereket hogyan érdemes mérni. Természetesen a mérési lehetőség nagyban függ a beszédadatbázis feldolgozottságától, ami lehet csupán egyszerű beszéd/nembeszéd szerinti, de akár pontos fonéma szintű szegmentálása is az adatbázisnak. Az utóbbit értelemszerűen lényegesen költségesebb megvalósítani, emiatt a legtöbb eddigi kutatásban nem alkalmaztak fonéma szintű szegmentálást, viszont ebből kifolyólag egyes beszédparamétereket csak pontatlanul, nagyobb szórással tudtak megmérni, illetve egyes beszédparaméterek meg sem mérhetőek a beszéd szegmentálása nélkül. Talán pont ezen okokból kifolyólag depresszió esetén a kutatók az egyes beszédparamétereknél eltérő tendenciákat mértek. [13][14]. Egy másik alapvető eltérés a különböző kutatásokban, hogy a női és a férfi beszédmintákat hol egyben kezelik, hol pedig külön. Ez részben függhet attól is, hogy a jelenleg létező depressziós beszédadatbázisok viszonylag kisméretűek: 30-160 beszélőtől tartalmaznak beszédmintákat [6], így ha külön vizsgálják a nőket és a férfiakat, akkor a vizsgált halmaz mérete tovább csökken. Ugyanakkor a beszélő neme nagyban befolyásolhatja az egyes beszédjellemzők értékeit, így ha közös modellt használunk a nők és a férfiak esetében, az problémát okozhat a depressziós állapot felismerésében, ami mindenképpen megoldandó feladat. Másrésztől szakorvosok állítják, hogy a férfiak és a nők beszédében nem ugyanúgy realizálódik minden esetben a depressziós

állapot, így ez kifejezetten indokolhatja a beszélő neme szerinti eltérő eljárás kidolgozását.

Ebben a cikkben egy olyan gépi eljárást mutatunk be, ami képes a depresszió súlyosságának becslésére a vizsgált személy beszédjele alapján magyar nyelv esetén. Továbbiakban bemutatjuk, hogy mennyiben változik az eljárás pontossága, ha külön kezeljük a férfiakat és a nőket, illetve ha alkalmazunk automatikus beszédsegmentálót, így lehetőségünk nyílik egy másféle jellemzőhalmaz kinyerésére. A következő két hipotézist állítottuk fel a munka megkezdése előtt: A depresszió súlyossága pontosabban becsülhető, ha külön modellt használunk a férfiak és a nők esetén (H1), illetve a depresszió súlyossága pontosabban becsülhető, ha a jellemzők kiszámításhoz felhasználjuk a beszéd fonéma szintű szegmentálását (H2).

A cikk a következő felépítést követi. A bevezetés után a második fejezetben bemutatjuk a használt beszédatadbázist. A harmadik fejezetben a munka során használt módszereket. A negyedik fejezetben tárgyaljuk az elvégzett kísérleteket. Az ötödik fejezetben az eredmények értékelése történik meg. Majd a hatodik fejezetben összefoglaljuk a munka eredményeit.

2 Adatbázis

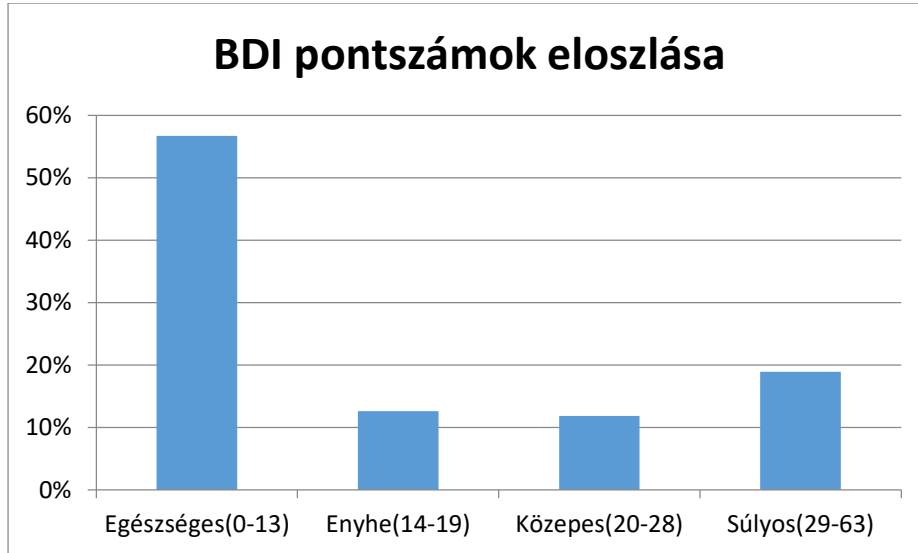
A beszédminták gyűjtését a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikájával együtt végeztük. A beszédminták gyűjtésénél törekedtünk arra, hogy a beszélők lefedjék a depresszió súlyosságának különböző fokozatait, az egészséges állapottól az egészen súlyos depresszióig. A vizsgált személyeknek egy fonetikusán kiegyensúlyozott mesét ("Az északi szél és a Nap") kellett felolvasniuk, ami széles körben elterjedt a miénkhez hasonló vizsgálatokban. A felvételek csendes helyiségben kerültek rögzítésre 44,1 kHz mintavételi frekvenciával. A cikk során erre a beszédatadbázisra „Magyar Depressziós Adatbázis”-ként hivatkozunk.

Az adatbázisba gyűjtött felvételekhez elkészítettük az egyes felvételekhez tartozó fonéma szintű szegmentálást, a labor által fejlesztett automatikus szegmentáló program segítségével [8]. Minden esetben rögzítésre került a BDI-II-es pontszám, amely az adott személy depressziójának súlyosságát írja le. A BDI-II skála 0-tól 63-ig terjed, ahol a 0 az egészséges állapotot jelöli, míg a 63 a legsúlyosabb depressziós állapotot. A BDI-II skála pontszámaihoz a következő besorolás adott: 0-13 egészséges, 14-19 enyhe depresszió, 20-28 közepes depresszió, 29-63 súlyos depresszió. A BDI pontszámok 0-tól 50-ig fordultak elő a gyűjtött mintákban. A vizsgált személyek átlagéletkora 42,2 év volt, (-+14,4; min.: 20; max.: 65). Az 1. táblázatban láthatóak az adatbázis főbb jellemzői.

1. Táblázat: A Magyar Depressziós Adatbázis főbb jellemzői.

Bemondók száma	BDI-értékek átlaga	BDI-értékek szórása
127 (nő:79/ffi:48)	14,2 (nő:14,7/ffi:13,3)	13,5 (nő:14/ffi:12,7)

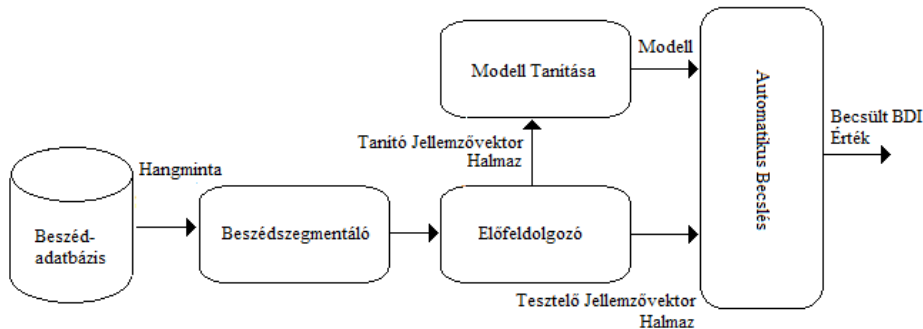
Az 1. ábra mutatja BDI pontszámok eloszlását az adatbázisban szereplő személyeknél, a depresszió súlyossága alapján.



1. ábra: A vizsgált személyek BDI-értékük szerinti eloszlása a Magyar Depressziós Adatbázisban.

3 Módszerek

A munka során használt automatikus gépi eljárás folyamatábráját a 2. ábrán lehet látni. Az eljárás először szétválasztja a beszédatadabázist tanító és tesztelő mintákra. Ezután a beszédmintákat szegmentálja és címkézi fonéma szinten. Majd az így felcímkézett hangmintákon elvégzi az akusztikai és fonetikai jellemzők kiszámítását. A kiszámított értékekből jellemzővektorokat generál, és az egyes paraméterek értékeit -1 és 1 közötti skálára normalja. Tanítás esetén ezekből a jellemzővektorokból készíti el a regressziós modellt adott gépi tanuló eljárással. A tesztelés során az eljárás az adott jellemzővektor és a regressziós modell alapján becsüli meg a vizsgált mintához tartozó beszélő depressziójának súlyosságát és rendel hozzá BDI pontszámot.



2. ábra: Az automatikus gépi eljárás folyamatábrája.

3.1 Előfeldolgozás

A beszédminták 16 kHz-en újra lettek mintavételezve és csúcsra normálva.

Korábbi tapasztalataink alapján és a nemzetközi irodalommal összhangban a következő jellemzőket használtuk a vizsgálatok során: alapfrekvencia, intenzitás, melsávós energiaértékek, jitter, shimmer, formánsértékek (első és második), formánsok sáv szélességei (első és második). Ezen jellemzők 10 ms-os lépésközzel Praat program segítségével kerültek kiszámításra [2]. Majd a számított értékekből a következő statisztikai függvények használatával – átlag, korrigált szórás, percentilis tartomány (a rendezett mintahalmaz alsó-felső 2,5%-nak elhagyása után képzett különbsége a maximum és minimum értéknek) – nyertük ki a hangmintákhoz rendelt jellemzőket. Ezt a jellemzőhalmazt még tovább bővítettük az artikulációs sebesség, a beszédtempó és a tranziensarány (rate of transients) jellemzőkkel [9].

Az akusztikai és fonetikai jellemzőket két osztályba lehet sorolni, a szegmentális és a prozódiai jellemzőkre. A szegmentális jellemzők számításának a helye nagyban befolyásolja a statisztikai jellemzők számított értékét. Emiatt kétféleképpen számítottuk ki ezeket, az egyik esetben úgy, hogy a zöngés szakaszon mért értékből képeztük a statisztikai jellemzőket (ez a számítás elvégezhető szegmentálás nélkül), illetve úgy is, hogy az adott bemondás összes „e” hangjának a közepén mért értékekből lettek képezve a statisztikai jellemzők. Azért az „e” hangot választottuk, mert ez a magánhangzó fordult elő leggyakrabban a felolvasott mesében.

3.2 Regressziós gépi tanuló eljárás

A kísérletek során gépi tanuló eljárásnak a Support Vector Regression (SVR) eljárást választottuk [16], ami a Support Vector Machine (SVM) regressziós feladatokra alkalmas változata [5]. Az SVM az általános lineáris osztályozók családjába tartozik, ám képes nemlineáris problémák megoldására is a kernel függvény megfelelő megválasztásával. Az SVR(SVM) egyedi tulajdonsága, hogy egyidejűleg minimalizálja a regressziós eljárás hibáját, és közben maximalizálja az eljárás általánosító képességét.

A kísérletek megvalósítása során a LibSVM 3.20 verzió számú függvény könyvtárát használtuk [4], Radial Basis Function (RBF) kernellel, a kernel által használt hiperparamétereket (cost és gamma) Grid Search kereséssel optimalizáltuk úgy, hogy a 2 hatványai lettek kipróbálva -10 és +10 között.

3.3 Jellemzővektor kiválasztás

Az SVR pontosságát nagyban befolyásolja a megfelelő jellemzővektorok kialakítása, vagyis a lényegtelen, zajszerű jellemzők elhagyása. Ez hatványozottan igaz a kis számú tanító mintahalmaz esetén, mint például a mi esetünkben is. Többféle jellemző kiválasztási algoritmus létezik a jobban teljesítő jellemzővektorok előállítására, mi a Fast Forward Selection (FFS) eljárást használtuk. Ennek az eljárásnak a lényege, hogy viszonylag gyorsan kiválaszt egy, az algoritmus által optimálisnak ítélt n elemű jellemzőhalmazt, ami az algoritmus által előállított eltérő számú, szuboptimális jellemzőhalmazok közül a legjobban teljesít. Az eljárás az üres jellemzőhalmazból indul ki. Az i -

dik lépésben rendelkezésére áll az algoritmus szerinti legjobb $i-1$ elemszámú jellemzőhalmaz, és ezt bővíti ki a legjobb i elemszámú jellemzőhalmazra úgy, hogy megvizsgálja, melyik eddig még fel nem használt jellemző hozzáadásával kapható a legjobb pontosan i elemszámú jellemzőhalmaz. Az előnye, hogy viszonylag gyors, a hátránya, hogy ha a k -dik lépésben beválaszt egy jellemzőt a jellemzőhalmazba, az onnantól kezdve minden halmazban benne lesz, ami k vagy annál nagyobb méretű.

3.4 A tesztelési eljárás

Az adatbázisban szereplő, viszonylag alacsony mintaszám miatt az ilyenkor szokásos, leave one out keresztvalidációs eljárást (LOOCV – leave one out cross validation) használtuk a tesztelések során minden esetben, így például az FFS alkalmazása és a hiperparaméterek optimalizálása során is. Az eljárás lényege, hogy a rendszer pontosságának a leírására mindegyik mintát pontosan egyszer felhasználja mint teszt-halmaz, míg a maradék mintákat mint tanító-halmazt, és így a tesztelőmintákon kapott eredmények írják le a rendszer teljes mintahalmazon számított pontosságát.

3.5 Az eljárás pontosságának mérése

Regressziós feladatok pontosságának jellemzésére többféle mérőszámot is lehet használni az adott módszer pontosságának leírására. Mi a következő három leíró jellemzőt választottuk, amelyek széles körben elterjedtek regressziós eljárások pontosságának leírására: az átlagos hibaértékét (MAE – mean absolute error), az átlagos négyzetes hibaértéknek a gyökét (RMSE – root mean square error), illetve az eredeti BDI pontszámoknak és az eljárás által becsült BDI pontszámoknak a Pearson-féle korrelációs együttható értékét.

4 Kísérletek

Összesen négy kísérletet hajtottunk végre a vizsgálat során. Minden kísérlet esetén külön alkalmaztuk az FFS eljárást, illetve optimalizáltuk a hiperparamétereket.

Az első kísérletnél együtt kezeltük a női és férfi mintákat, és csak olyan jellemzőket használtunk fel a jellemzővektorok kialakítása során, amelyeket a beszédjel szegmentálása nélkül is ki lehet számítani. Így a 3.1 alfejezetben tárgyalt jellemzők közül az artikulációstempó, a beszédtempó és az „e” hangokon számított szegmentális paraméterek statisztikai értékei nem kerültek bele ebbe a vizsgálatba. Erre a kísérletre a továbbiakban mint „baseline” kísérletre hivatkozunk. Azért neveztük el *baseline* kísérletnek, mivel a többi általunk elvégzett kísérlet ennek a „továbbfejlesztett” változata, ami speciálisabb előfeldolgozást illetve kialakítást igényelt.

A második kísérlet során ugyanazzal a jellemző halmazzal dolgoztunk, mint a *baseline* kísérlet esetében, de külön modellt hoztunk létre a női és férfi minták esetén. A gyakorlatban ez azt jelentette, hogy nemek szerint külön végeztük el a jellemzővektorok kialakítását (FFS), a hiperparaméter optimalizációt és a tesztelést. Erre a kísérletre

a továbbiakban, mint „gender” hivatkozunk. Ennek a lényege az volt, hogy megvizsgáljuk, hogyan módosul a *baseline* regressziós eljárás pontossága, ha nemek szerint eltérő modellt alkalmazunk, és ennek a kísérletnek a segítségével igazolhatjuk vagy cáfolhatjuk a H1 hipotézisünket.

A harmadik kísérletben hozzávettük azokat a jellemzőket is a vizsgálathoz, amelyek kiszámításához a beszédjel szegmentálása szükséges, vagyis a 3.1 alfejezetben felsorolt összes jellemzőből kerültek kialakításra a jellemzővektorok az FFS eljárás segítségével. De a női és a férfi mintákat együtt kezeltük, mint a *baseline* kísérlet esetében. Erre a kísérletre a továbbiakban mint „segmented” kísérlet hivatkozunk. Ennek a lényege az volt, hogy megvizsgáltuk, hogyan módosul a *baseline* regressziós eljárás pontossága, ha egyes jellemzőket nemcsak a beszéd egészén mérjük, hanem előre definiált pontos helyeken felhasználva a beszéd fonéma szintű szegmentálását, illetve a szegmentálás felhasználása által képesek voltunk artikulációstempó és beszédtempó mérésére is. Ennek a kísérletnek a segítségével igazolhatjuk vagy cáfolhatjuk a H2 hipotézisünket.

A negyedik kísérlet során a harmadik kísérlettel megegyező jellemzőhalmazzal dolgoztunk, de külön végeztük el a regressziós eljárást a férfiak és a nők esetében. Tehát egyszerre alkalmaztuk mindkét vizsgált eljárási módszert. Erre a kísérletre a továbbiakban, mint „gender+segmented” kísérletre hivatkozunk. Ennek lényege az volt, hogy megvizsgáltuk, hogyan módosul a *baseline* regressziós eljárás pontossága, ha egyszerre alkalmazunk az általunk javasolt két módszert, vagyis külön modellt a nők és a férfiak esetében, illetve plusz speciális jellemzőhalmaz használata, amelyek a beszéd fonéma szintű szegmentálásnak segítségével kerültek kiszámításra a 3.1 alfejezetben tárgyaltak szerint.

4 Eredmények

Az egyes kísérletek eredményeit a 2. táblázatban foglaltuk össze. A táblázat utolsó oszlopában (Rel. Vál.) a RMSE relatív változást adtuk meg a *baseline* kísérlet RMSE eredményének értékéhez képest. Azoknál a kísérleteknél (*gender* és *gender+segmented*), ahol külön modellt használtunk a nők és a férfiak esetén az eredmény alatt külön jelölve vannak a nemenként kapott eredmények is.

A 3. ábrán megadtuk az egyes kísérletek esetén az automatikus gépi eljárás által becsült BDI-értékeket az eredeti értékhez képest. Az ábrán négy kisebb ábra látható, mindegyik bal felső sarkában jeleztük, hogy melyik kísérlethez tartozik. Az ábrákon szereplő keresztek jelzik a Magyar Depressziós Adatbázisban szereplő hangmintákat, a vízszintes tengelyen leolvasható az eredeti BDI pontszámuk, míg a függőlegesen az adott kísérlet alapján becsült BDI pontszámuk. A könnyebb áttekintés érdekében mindegyik kísérlethez tartozó ábrán behúztuk az átlót, hiszen az ettől való távolság jelzi, hogy az adott minta esetében mennyit tévedett a gépi eljárás.

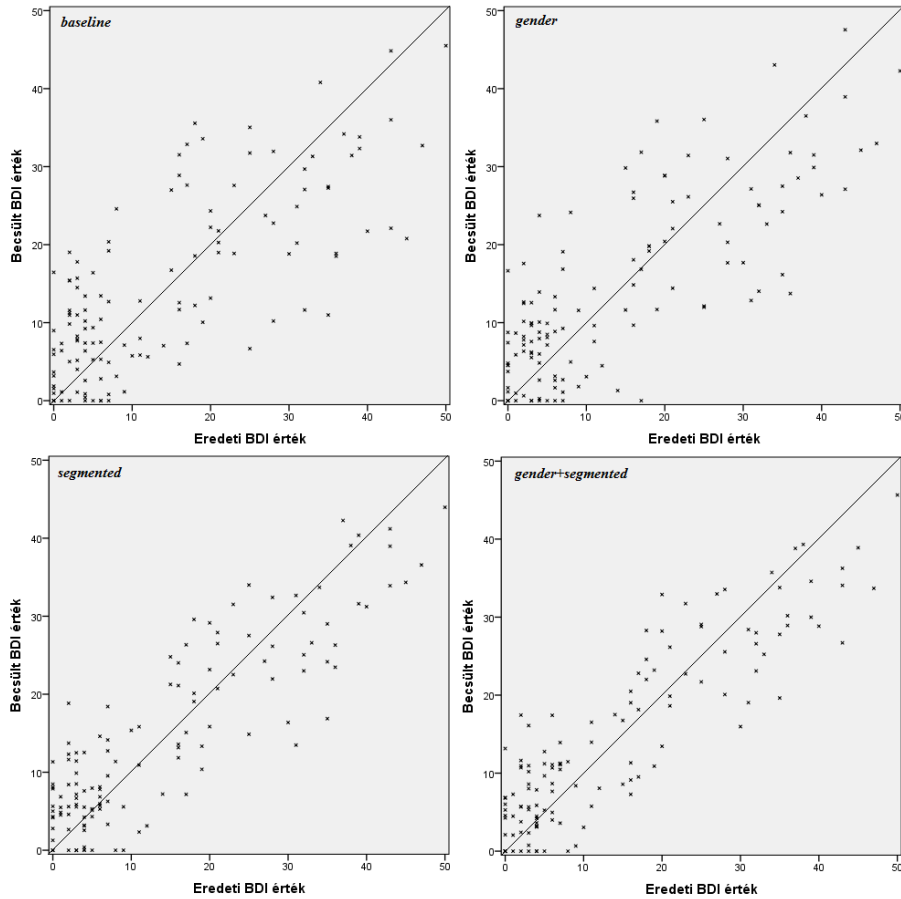
Ahhoz hogy a Magyar Depressziós Adatbázison elért regressziós eredményeink pontosságát össze lehessen hasonlítani más kutatók eredményével, beleraktuk a táblázatba az AVEC-2013 verseny győztese által használt regressziós eljárás leíró jellemzőit is (táblázatban szürke háttérrel van jelölve) [17]. Az AVEC-2013 versenyen regressziós eljárást kellett kidolgozni a megadott német nyelvű depressziós beszédatadtbázis hang-

felvételeit felhasználva, az azokat bemondók BDI-II skála szerinti pontszámuk becslésére. A német adatbázisban szereplő bemondók életkora (átlag 31,5 év és 12,3 szórás) és BDI-II szerinti BDI pontszámuk eloszlása (átlag 14,9 és 11,7 szórás) nagyon hasonló a Magyar Depressziós Adatbázishoz. Természetesen ettől még ez az összehasonlítás nem tökéletes, hiszen eltérő a két adatbázis nyelve, illetve a benne lévő személyek is, mégis úgy gondoljuk, hogy jó viszonyítási alapot ad.

2. Táblázat: Az automatikus becselő rendszer pontosságának leíró jellemzői az elévített kísérletek esetében, kiegészítve az AVEC 2013 verseny győztesének az eredményeivel.

	MAE	RMSE	Pearson corr.	Rel. Vál.
AVEC 2013 győztese	6,53	8,5	0,7	
Baseline	7,12	9,02	0,75	
Gender	6,76 Nők: 7,07 Férfiak: 6,26	8,37 Nők: 8,7 Férfiak: 7,81	0,79 Nők: 0,68 Férfiak: 0,79	-7%
Segmented	5,31	6,6	0,87	-27%
Gender+Segmented	5,1 Nők: 5,71 Férfiak: 4,1	6,28 Nők: 6,95 Férfiak: 4,99	0,89 Nők: 0,87 Férfiak: 0,92	-30%

Az eredmények alapján kijelenthetjük, hogy mind a H1, mind a H2 hipotézisünk beigazolódott, vagyis pontosabb becslést lehet adni, ha külön női és férfi modellt használunk, illetve beszédszegmentáló használatával ezáltal pontosabb és jobb – előállított jellemzőkkel. A javulás mértéke a beszédszegmentáló használata esetében volt jelentősebb, 27%-os relatív csökkenése az átlagos négyzetes hiba gyökének. Azonban a női-férfi elkülönítéssel sem elhanyagolható a javulás mértéke, az átlagos négyzetes hiba gyökének 7%-os relatív csökkenése.



3. ábra: Az automatikus becslések értékei a négy elvégzett kísérlet esetében, összehasonlítva a minták eredeti BDI-értékeivel.

4 Konklúzió

A cikkben bemutattunk egy automatikus rendszert, ami SVR regressziós eljárással képes a beszédjel alapján megállapítani a beszélő depressziós állapotának a súlyosságát. Több kísérletet is elvégeztünk, hogy megvizsgáljuk, hogyan változik a becslés pontossága, ha eltérő módon hozzuk létre a rendszert.

A kísérletek során a Magyar Depressziós Adatbázist használtuk. Megadtunk egy *baseline* eredményt, aminek a kialakításában a korábbi, ebben a témában nyert tapasztalatainkra támaszkodtunk összhangban a nemzetközi irodalomban található eredményekkel. A rendszer pontosságának a leírására RMSE, MAE és Pearson-féle korrelációs értékeket használtunk. Így a *baseline* eredményeknek, ahol együtt kezeltük a női és a férfi mintákat, illetve a jellemzővektorok kialakításához nem használtuk fel a beszéd fonéma szintű szegmentálását, a következő értékeket kaptuk *RMSE*: 9,02, *MAE*:

7,12, *Pearson corr.*: 0,75). Ezt az eredményt összehasonlítva az AVEC 2013 győztesének az eredményével (*RMSE*: 8,5, *MAE*: 6,53, *Pearson corr.*: 0,7) megállapíthatjuk, hogy a rendszerünk bár rosszabbul teljesített (*RMSE*-érték alapján), de a hiba növekményének a mértéke nem számottevő, természetesen ez az összehasonlítás nem tökéletes, mivel a két kísérlet eltérő adatbázist használ. Érdemes még összevetni a *baseline* *RMSE*-értéket az adatbázisban található személyek BDI-értékeinek a szórásával, ami 13,5, mivel ezt az értéket kapnánk, ha minden személyhez az átlagos BDI-értéket rendelnénk hozzá, vagyis a rendszer *RMSE*-ben kifejezve 4,48 hibapont értékkel teljesít jobban, mint az elvárt minimum.

A munka elején két hipotézist fogalmaztunk meg, miszerint a rendszer pontossága javul, ha külön kezeljük a női és férfi mintákat (H1), illetve ha automatikus beszéd-seggmentálót használva, a bemondások fonéma szintű szegmentálásnak segítségével, egy jobb, pontosabb, kibővített jellemzőhalmazt alkalmazunk (H2).

Mindkét hipotézisünket igazoltuk, ugyanis 7%-os relatív javulást értünk el az *RMSE*-értékben, ha külön kezeltük a férfi és női mintákat, illetve 27%-os relatív javulást értünk el, ha felhasználtuk a jellemzők előállításánál a beszéd fonéma szintű szegmentálását a *baseline* eredményekhez képest. Az eredmények alapján a szegmentálás alapú jellemző számítás tűnik fontosabbnak a pontosság szempontjából, ugyanakkor tény, hogy a nemek szétválasztása esetén csökkent a tanító minták száma az így kialakított két külön rendszerben, így ennek a tükrében a 7%-os relatív javulás mindenképpen jelentős.

Legvégül megvizsgáltuk, hogyan változik a pontosság, ha a két vizsgált módszert ötvözzük (*RMSE*: 6,28, *MAE*: 5,1, *Pearson corr.*: 0,89). *RMSE*-ben mérve a *baseline* eredményhez képest 30%-os relatív javulást értünk el, míg maga a hibaérték abszolút értelemben véve is kifejezetten alacsonynak mondható, így akár ez a módszer alkalmas lehet egy depressziós állapotot diagnosztizáló rendszer alapjának.

Továbbiakban tervezzük más kevésbé elterjedt jellemzők felhasználásával is megvizsgálni a rendszerünk pontosságának változását. Tervezzük, hogy módszerünket kipróbáljuk más adatbázisokon is. Illetve az adatbázisunkat folyamatosan bővítjük.

Köszönetnyilvánítás

A kutatást támogatta az ESA ügynökség COALA projekt: Psychological Status Monitoring by Computerised Analysis of Language phenomena (COALA) (AO-11-Concordia).

Bibliográfia

1. Beck, A.T., Steer, R.A., Ball, R., Ranieri, W.F., (1996). Comparison of beck depression inventories-ia and-ii in psychiatric outpatients. *J. Pers. Assess.* 67, 588–597.
2. Boersma, P., (2002). Praat, a system for doing phonetics by computer. *Glott international*, 5(9/10), pp.341-345.

3. Brendel, R.W., Wei, M., Lagomasino, I.T., Perlis, R.H., Stern, T.A., (2010). Care of the suicidal patient. *Massachusetts General Hospital Handbook of General Hospital Psychiatry*, 6th ed. W.B. Saunders, Saint Louis, pp. 541–554.
4. Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27.
5. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
6. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, 71, 10-49.
7. Hamilton, H., (1960). HAMD: a rating scale for depression. *Neurosurg. Psych.* 23, 56–62.
8. Kiss, G., Sztahó, D. and Vicsi, K., (2013), December. Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. In *Cognitive Infocommunications (CogInfoCom)*, 2013 IEEE 4th International Conference on (pp. 579-582). IEEE.
9. Kiss, G. and Vicsi, K., (2014). Physiological and cognitive status monitoring on the base of acoustic-phonetic speech parameters. In *International Conference on Statistical Language and Speech Processing* (pp. 120-131). Springer International Publishing.
10. Kraepelin, E., (1921). Manic depressive insanity and paranoia. *J. Nerv. Ment. Dis.* 53, 350.
11. Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D. and Saxena, S. (2012). Depression: A global public health concern. *WHO Department of Mental Health and Substance Abuse*, 1, 6-8.
12. Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *Plos med*, 3(11), e442.
13. Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of neurolinguistics*, 20(1), 50-64.
14. Mundt, J. C., Vogel, A. P., Feltner, D. E., & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological psychiatry*, 72(7), 580-587.
15. Nilsson, A., (1988). Speech characteristics as indicators of depressive illness. *Acta Psych. Scand.* 77, 253–263.
16. Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in neural information processing systems*, 9, 155-161.
17. Williamson, J. R., Quatieri, T. F., Helfer, B. S., Horwitz, R., Yu, B., & Mehta, D. D. (2013). Vocal biomarkers of depression based on motor incoordination. In *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge* (pp. 41-48). ACM.