Depresszió detektálása korrelációs struktúrán alkalmazott konvolúciós hálók segítségével

Jenei Attila Zoltán[1], Kiss Gábor[2]

[1] Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
ja1504@hszk.bme.hu
[2] Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
kiss.gabor@tmit.bme.hu

Kivonat: Jelen kutatásban a depressziós állapot automatikus detektálásának lehetőségét vizsgáltuk a beszédjelből kinyert speciális korrelációs struktúrán alkalmazott konvolúciós neurális hálok segítségével. A depresszió korunk egyik legelterjedtebb gyógyítható pszichiátriai betegsége. A depressziótól szenvedő egyén életminőségét nagymértékben befolyásolja a depresszió súlyossága, ami extrém esetben öngyilkossághoz is vezethet. Ezek alapján kulcsfontosságú, hogy már korai stádiumában felismerhető legyen a betegség és az illető megfelelő kezelésben részesüljön, azonban a depresszió diagnosztizálása szakértelmet kíván, emiatt fontos a depresszió esetleges jelenlétének automatikus jelzése. Ebben a cikkben egy olyan eljárást mutatunk be, ami beszédjel feldolgozása alapján tisztán spektrális jellemzőkön keresztül képes felismerni a depressziót konvolúciós neurális hálók alkalmazásának segítségével. Bemutatjuk, hogyan változik a depresszió detektálásának pontossága különböző akusztikai-fonetikai jellemzők felhasználása alapján, illetve a korrelációs struktúrának változtatása következtében. A módszer alkalmazásával 84%-os pontossággal tudtuk elkülöníteni az egészséges és depressziós személyeket a beszédmintáik alapján.

1 Bevezetés

Számos betegség hatással bír a beszédkeltés folyamatára, és ez által hatással bír a kialakult beszédproduktumra is. Az egyes betegségek beszédre gyakorolt hatását lehetséges kimutatni a kialakult beszédproduktum akusztikai-fonetikai jellemzőinek megmérésével és ez alapján lehetséges az adott betegségek automatikus detektálása, illetve ezáltal szokás a beszédre mint egy objektív biomarkerre tekinteni (Sztahó és mtsai, 2019; Sztahó és mtsai, 2018; Liu és mtsai, 2017; Kiss és mtsai, 2017a; Tóth és mtsai, 2015; Orozco-Arroyave és mtsai, 2015; Cummins és mtsai, 2015).

A depresszió korunk egyik legelterjedtebb gyógyítható pszichiátriai betegsége (Friedrich, 2017), ami a WHO (World Health Organization) 2012-es felmérése alapján a harmadik leggyakoribb betegség világszerte (Marcus és mtsai, 2012).

A depresszió kialakulásának pontos okai még nem ismertek, azonban a depresszió élettani tünete leginkább a kortikális limbikus rendszer egyfajta diszfunkciójaként jelentkezik (Deckersbach és mtsai, 2006; Nestler és mtsai, 2002).

Depressziós állapot hatására az ettől szenvedő egyénnek a depresszió súlyosságának függvényében nehezére eshet elvégezni a napi teendőit, ami jelentős gazdasági károkat okozhat (Olesen és mtsai, 2012), ezen felül a depresszió súlyosbodásával megnövekedhet az öngyilkosság kockázata is (Hawton és mtsai, 2013). Azonban a depresszió diagnosztizálása szaktudást igényel, emiatt különösen fontos minden olyan megoldás ami segíthet a depresszió diagnosztizálásának támogatásában, illetve alkalmas lehet a depresszió veszélyének jelzésére.

A tény, hogy depresszió hatására megváltozik az emberi beszédproduktum már 1921-ben publikálta Kraeplin (Kraepelin, 1921), azonban a depressziós állapot és a beszéd kapcsolatának mélyebb vizsgálata, illetve a depresszió automatikus detektálása a megváltozott beszédproduktum alapján újszerű kutatási területnek számít, amit elsősorban az egyre nagyobb depressziós beszédadatbázisok megjelenése, illetve az informatika fejlődése tett lehetővé (Cummins és mtsai, 2015).

A korábbi kutatások esetében a depressziós és egészséges személyek automatikus elkülönítését, függetlenül a depressziós állapot súlyosságától, 50-86% közötti pontossággal voltak képesek megvalósítani beszédjel feldolgozás alapján (Kiss és Vicsi, 2017b; Kiss és Vicsi, 2017c; Alghowinem és mtsai, 2013; Ooi és mtsai, 2013; Cummins és mstai, 2013; Low és mstai, 2009). Természetesen a depresszió felismerésének a pontossága az egyes kutatások esetében nagyban függhet a kutatásban használt adatbázistól, illetve az adatbázis feldolgozottságától, az alkalmazott módszerektől. Annak ellenére, hogy több kutatás bizonyította már, hogy lehetséges a depresszió automatikus detektálása beszédjel feldolgozás alapján, több nyitott kérdés is van még, úgy, mint mely akusztikai-fonetikai jellemzők a legalkalmasabbak a depressziós állapot detektálásához, illetve milyen szintű feldolgozás szükséges a depressziós állapot detektálásához a minél nagyobb pontosság elérése érdekében.

Az utóbbi időben számos tanulmány a beszédakusztikai jellemzők széles skáláját alkalmazta annak érdekében, hogy (főleg bináris) osztályozást végezzen a depressziós és egészséges alanyok elkülönítésére (Kiss és Vicsi, 2017b; Vlasenko és mtsai, 2017; Cummins és mtsai, 2015; Valstar és mtsai, 2013). Azonban a megfelelő akusztikai-fonetikai jellemző halmaz kiválasztást nehezíti, hogy az eddig rendelkezésre álló depressziós beszédadatbázisok csupán 50-150 főtől tartalmaznak beszédmintákat (Cummins és mtsai, 2015; Kiss és Vicsi, 2017b), így az alkalmazott gépitanuló eljárások értelemszerűen csak limitált méretű jellemzővektorral képesek dolgozni a túltanulás elkerülése végett.

Jelen kutatásban Williamson és mtsai. 2013-ban publikált korrelációs mátrix alapú megoldását használjuk fel (Williamson és mtsai, 2013). Az adott publikációban az alacsony szintű akusztikai-fonetikai jellemzők auto- és keresztkorrelációs struktúrája alapján, nagy pontossággal képesek voltak a depressziós állapot súlyosságát becsülni. A magas pontosság mellett még figyelemre méltó volt a kutatásban, hogy mindösszesen a beszédjelből kinyert MFCC (Mel-frequency cepstral coefficients) és a formáns frekvenciák jellemzőkre támaszkodtak. Az eredményeket a német nyelvű depressziós beszédadatbázison érték el (Valstar és mtsai, 2013). Az eljárást már korábban sikeresen alkalmazták az agyi EEG (Electroencephalography) jeleken a kezdődő epilepszia jelzésére (Williamson és mtsai, 2011). Az auto- és keresztkorrelációs eljárás egy adott jellemzővektor halmazból kiindulva előállítja annak egy speciális korrelációs mátrix struktúrájú reprezentációját. A mátrix egyes celláiban a jellemzővektor halmazból vett két jellemzővektor (a két jellemzővektor lehet ugyanaz) korrelációs együttható értéke

található, meghatározott eltolások mellett. Az eljárás pontos ismertetését a 3.2-es fejezetben részletezzük.

Az előállított korrelációs mátrix az átlóra szimmetrikus, emiatt Williamson és mtsai. az előállított korrelációs mátrix sajátértékeit számíttatták ki, majd azoknak csupán egy részhalmazát használták fel a gépi tanuló eljárás bemeneteként, és becsülték ez alapján depresszió súlyosságát.

Az eljárás sikeressége feltehetőleg abban rejlik, hogy a beszéd nagy időablakban vett megváltozott struktúráját képes megfelelően reprezentálni. Korábbi kutatásunkban a korrelációs mátrix alapú eljárást mi is sikeresen alkalmaztuk egyidejűleg több betegség felismerésére (Sztahó és mtsai, 2018), ahol is 3 különböző betegség (Parkinson kór, depresszió és egyéb gégészeti elváltozások) és egészséges beszélőktől származó beszédminták automatikus elkülönítését végeztük el 78%-os pontossággal. Azonban Williamson és mtsai. által publikált eljárásnak van egy fő hátránya, ugyanis az előállított korrelációs mátrix sajátértékekkel vett reprezentációja nem feltétlenül optimális és még mindig redundáns, illetve túl nagy méretű. Emiatt szükséges a sajátértékek alapú reprezentáció dimenziójának további csökkentése, aminek megfelelő megválasztásától a gépi tanuló eljárás pontossága és általánosító képessége is nagyban függhet. A problémát tovább rontja, ha egyszerre sok akusztikai-fonetikai jellemzőt is fel szeretnénk használni a depresszió felismerésére.

Emiatt jelen kutatásban a korrelációs mátrixok sajátértékeinek használata helyett, a mátrixokat közvetlenül alkalmaztuk a gépi tanuló eljárás bemeneteként. Ehhez konvolúciós (CNN) neurális hálókat alkalmaztunk. Az eljárás előnye, hogy így a gépi tanuló eljárás feladata a korrelációs mátrix megfelelő feldolgozása is. A kutatásban még újszerű, hogy nemzetközi viszonylatban is nagy mintaszámúnak számító, közel 200 beszédmintán tudtuk tesztelni az eljárást.

A cikk a következő felépítést követi. A bevezetés után a második fejezetben bemutatjuk a felhasznált beszédadatbázist. A harmadik fejezetben bemutatjuk az alkalmazott alacsony szintű jellemzőket, a korrelációs mátrix kiszámításának módját és az azon alkalmazott konvolúciós neurális hálók felépítését, illetve a kiértékelési módszereket. A negyedik fejezetben bemutatjuk az eredményeket. Az ötödik fejezetben röviden összefoglaljuk a kutatás fő eredményeit és a további terveinket.

2 Magyar Depressziós Beszédadatbázis

A kutatás során a Magyar Depressziós Beszédadatbázis beszédmintáira támaszkodtunk. A beszédadatbázist folyamatosan bővítjük. Jelen kutatásban a beszédadatbázisban elérhető 91 egészséges és 91 depressziós személytől gyűjtött beszédmintákat használtuk fel, minden személytől pontosan egy beszédmintát. Az egészséges személyek esetén csak olyan személyek beszédmintáit használtuk fel, akik - saját bevallásuk alapján - nem szenvedtek semmilyen olyan betegségtől, ami hatással bírhat a beszédükre. A depresszióval diagnosztizált betegek esetén szintén csak az olyan személyektől származó beszédmintákat használtunk fel, akik nem voltak diagnosztizálva más olyan betegséggel, ami szintén hatással bírhat a beszédproduktumukra (pl.: Parkinson kór, ALS).

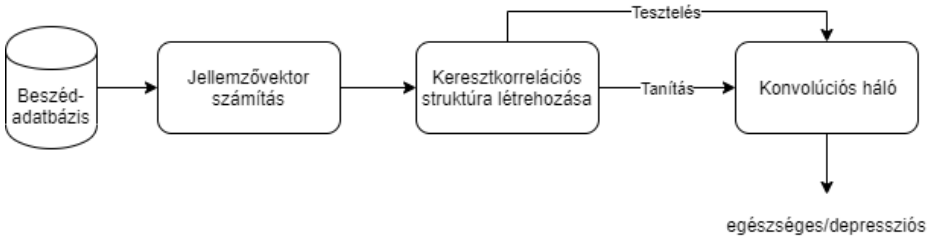
A beszédminták gyűjtését a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikájával együtt végezzük. A beszédminták gyűjtésénél törekedtünk arra, hogy a

beszélők lefedjék a depresszió súlyosságának különböző fokozatait, az egészséges állapottól az egészen súlyos depresszióig. A depressziós személyek esetében körülbelül egyenletes eloszlással szerepeltek alanyok a BDI-II (Beck Depression Inventory-II) (Beck és mtsai, 1996) által definiált depressziós súlyosság szerinti kategóriák között, úgy, mint az enyhe depresszió, közepes depresszió és súlyos depresszió.

A vizsgált személyeknek egy fonetikusan gazdag mesét ("Az északi szél és a Nap") kellett felolvasniuk. A felvételek csendes helyiségben kerültek rögzítésre 44,1 kHz mintavételi frekvenciával, csiptetős mikrofonnal.

3 Módszerek

A kutatásunkban alkalmazott, a depressziós állapot detektálására alkalmas módszer folyamatát az 1. ábra mutatja be. Az eljárás bemenetén a beszédminta áll, míg az eljárás kimenete a bemondó bináris osztályozása (egészséges/depressziós) a beszédmintája alapján. Az eljárás először az adott beszédmintából különböző akusztikai-fonetikai jellemzőenergia vektorokat nyer ki. Ezt követően a kiszámított jellemzővektorok részhalmazából előállítja azok auto- és keresztkorrelációs mátrixát. A létrehozott kétdimenziós korrelációs mátrix lesz a bemenete a 2D konvolúciós hálónak, ami tanítás esetén létrehozza a megfelelő modellt, majd tesztelés esetén a modell segítségével elvégzi a bemondó bináris osztályozását, ami az eljárás végső kimenete.



1. ábra. A kutatásban bemutatott depressziós állapot detektálására alkalmas módszer folyamat ábrája.

3.1 Felhasznált akusztikai-fonetikai jellemzők

Az akusztikai-fonetikai jellemzők számítása előtt a beszédminta minden esetben csúcsértékre lett normalizálva, ezzel kiküszöbölve a felvételek rögzítése esetén esetlegesen felmerülő eltérő erősítésbeli különbségeket. A következő alacsony szintű jellemzőket használtuk:

Mel-sávos energiaértékek: Az emberi hallás frekvenciabeli felbontásához hasonló sávokban adja meg az energiaértékeket. A sávokat 100-dik mel-től kezdve 100 melenkénti összegzéssel valósítottuk meg, összesen 27 mel-sávos energiaérték kiszámíttatásával, amivel körülbelül 60 Hz és 8 kHz között végeztük el a beszédjel frekvenciabeli felbontását.

MFCC együtthatók: Az MFCC együtthatók alkalmazása és azoknak fontossága a beszédjel feldolgozás területén bevett gyakorlatnak számít. Az MFCC együtthatókat a 27 mel-sávos energiaérték diszkrét koszinusz transzformáltjaként számíttattuk ki és összesen 14 együtthatót használtunk fel végül.

Formáns frekvenciák: Formáns frekvenciákon a beszédjel feldolgozás esetében, a rezonátorüregek által felerősített felhangnyalábok burkoló görbéinek maximum helyeit értjük. A kutatás során az első három formánsfrekvencia értékeket számíttattuk ki és használtunk fel, amikre a továbbiakban mint F1, F2 és F3 hivatkozunk.

Formáns frekvenciák sávszélessége: Az adott formánsfrekvencia sávszélessége alatt, a formánsfrekvencia 3 dB-es csökkenésénél mért sávszélességet értjük. A kutatás során az F1, F2 és F3 formáns frekvenciák sávszélességét számíttattuk ki és használtuk fel, amiket a továbbiakban B1, B2 és B3-al jelölünk.

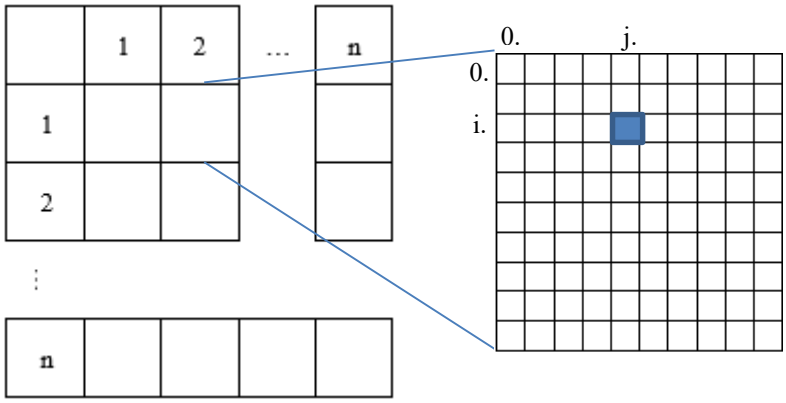
A kutatás során felhasznált akusztikai-fonetikai jellemzőket 10 ms-os lépésközzel, 50 ms-os ablakkal számíttattuk ki. A mel-sávos energiaértékeket és az MFCC együtthatókat a teljes beszédmintából, míg a formáns frekvenciákat és azok sávszélességeit pedig a zöngés szakaszokból számíttattuk ki. Így minden egyes akusztikai-fonetikai jellemző esetében egy jellemzővektort kaptunk az adott beszédmintára, ahol is a mel-sávos energiaértékeket és az MFCC együtthatókat tartalmazó vektorok hossza, illetve a formáns frekvenciák és azok sávszélességeinek hossza megegyezik.

A jellemzők kiszámításhoz a Praat szoftvert használtuk (Boersma, 2001).

3.2 Korrelációs struktúra

A jellemzővektorok adott halmazaiból azok auto- és keresztkorrelációs együtthatóit tartalmazó korrelációs mátrixait hoztuk létre. A korrelációs mátrix számítását tömören ismertetjük, bővebb leírása Williamson és mtsai 2013-as cikkében található (Williamson és mtsai, 2013)

A korrelációs mátrix felépítését a **2. ábra** szemlélteti, ahol *n* jelöli a korrelációs mátrix bemeneti jellemzővektorainak számát.



2. ábra. A korrelációs mátrix felépítése.

A korrelációs mátrix $n \cdot n$ darab almátrixból épül fel (2. ábra bal oldala). A főátló mentén található almátrixok az adott jellemzővektorok autokorrelációs együtthatóit, míg a többi almátrix két különböző jellemzővektor keresztkorrelációs együtthatóit tartalmazza *dt* darab eltolás mellett. Jelen kutatás során a $dt = 10$ értéket alkalmaztunk.

Minden almátrix összesen $dt \cdot dt$ darab korrelációs együtthatót tartalmaz. (Vagyis a teljes mátrxinak összesen $(n \cdot dt) \cdot (n \cdot dt)$ darab cellája van.) Az adott almátrix egyértelműen meghatározza, hogy mely két jellemzővektor korrelációs értékei találhatók benne a felhasznált jellemzővektor halmazból. Az almátrix első cellája (0. sor, 0. oszlop) a két jellemzővektor eltolás nélküli korrelációs együttható értékét tartalmazza. Az adott almátrix egy tetszőleges i. sorában és j. oszlopában található korrelációs együttható értéke pedig az első jellemzővektor *i* szer vett eltolása és a második jellemzővektor *j* szer vett eltolása melletti korrelációs együttható értékét tartalmazza (2. ábra jobb oldala). A korrelációs mátrix felépítéséből fakadóan a fő átló elemei csupa 1-et tartalmaznak, illetve a mátrix szimmetrikus a fő átlóra. Fontos megjegyezni, hogy értelemszerűen komolyabb módosítások nélkül, csak egyforma hosszúságú jellemzővektorokra működik az eljárás. Kutatás során az eltolás mértékére 3 különböző értéket is kipróbáltunk (1, 2 és 8), vagyis például ha 2 volt az eltolás mértéke és az *i* értéke éppen 3 volt, akkor az adott jellemzővektort 6 értékkel toltuk el.

Az eljárás alapján minden egyes beszédmintából pontosan egy korrelációs mátrixot számíttattunk ki egy adott jellemzővektor halmaz esetében.

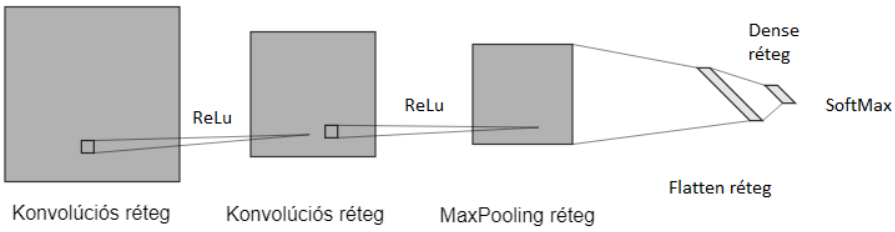
Összesen 5 különböző jellemzővektor halmazt alkalmaztunk a kutatás során:

- Mel-sávos energiaértékeket tartalmazó jellemzővektorok
- MFCC együtthatókat tartalmazó jellemzővektorok
- Formáns frekvenciákat tartalmazó jellemzővektorok
- Formáns frekvenciák sávszélességét tartalmazó jellemzővektorok
- Formáns frekvenciák és azok sávszélességeit tartalmazó jellemzővektorok

3.3 Konvolúciós háló

Az osztályozó algoritmusnak 2D konvolúciós neurális hálókat alkalmaztunk. A gépi tanuló eljárás bemenete az adott jellemzővektor halmazból kiszámított korrelációs mátrix volt.

Az algoritmus létrehozása Phyton kódban történt TensorFlow környezetben. Felépítését a **3. ábra** szemlélteti.



3. ábra. Az alkalmazott konvolúciós háló szerkezete.

A konvolúciós rétegek 32 filtert használtak, amik mérete $10 \cdot 10$-es az almátrix méretének megfelelően. A kernel lépésközének szintén 10 volt beállítva az eltolások száma (*dt*) alapján. A maxPooling kernel mérete $2 \cdot 2$-es volt és same paddinget alkalmaztunk. Az első három réteg után dropOut regulációt használtunk, ami véletlenszerűen a neuronok 25 %-át figyelmen kívül hagyta a tanítás során. A Flatten rétegbe már a 32 filter értéke került, így mérete $1 \cdot 32$-es volt, ami a bemenete egy fully connected neurális hálónak (a Dense rétegnek). Ennek kimenetén SoftMax függvényt alkalmaztunk bináris osztályozáshoz.

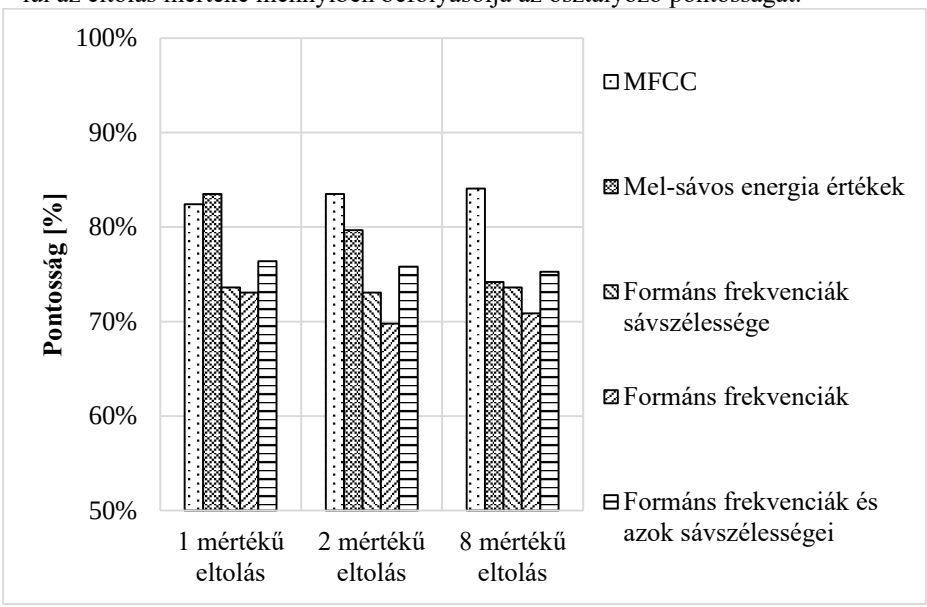
3.4. Kiértékelési módszerek

A viszonylag alacsony mintaszám miatt az ilyenkor szokásos teljes kereszt-validációs eljárást alkalmaztunk, a háló tanítás és tesztelése során. Vagyis minden egyes mintát egyszer, mint teszt halmaz a maradék mintákat pedig mint tanító halmaz alkalmaztuk. Fontos megjegyezni, hogy természetesen ez által a tesztelő és tanító halmazok minden esetben diszjunktak voltak.

Az osztályozási kísérletek során a minél nagyobb pontosság (helyesen osztályozott minták száma osztva az összes minta számával) elérését tűztük ki célul. Emellett vizsgáltuk még az orvosi diagnosztikában nagy fontosságú és emiatt gyakran alkalmazott specificitás és szenzitivitás értékeket is, hiszen a gyakorlatban nem feltétlenül számít ugyanakkora hibának, ha egy egészséges embert depressziósnak ítélünk, mint fordítva.

4. Eredmények

Összesen 15 különböző osztályozási kísérletet valósítottunk meg. 5 jellemzővektor halmazt vizsgáltunk és mindegyikből 3 különböző mértékű eltolás alkalmazásával állítottunk elő korrelációs mátrixokat (lásd 3.2-es fejezet). Elsősorban azt vizsgáltuk, hogy mely jellemzővektor halmazzal lehet elérni a legnagyobb pontosságot, illetve azon belül az eltolás mértéke mennyiben befolyásolja az osztályozó pontosságát.



4. ábra. Depresszió felismerésének pontossága különböző jellemzővektor halmazokból és eltolás mértékkel előállított korrelációs mátrixok alapján.

4. ábrán látható a 15 különböző kísérlet elvégzése során kapott pontosságok értéke. Amint látható MFCC együtthatók felhasználásával 8 mértékű eltolással előállított korrelációs mátrix esetén kaptuk a legnagyobb pontosságot, ebben az esetben 84%-os pontossággal tudtuk elkülöníteni a depressziós és egészséges beszélőket. Továbbá megfigyelhető, hogy az eltolás mértékének a növelésével a legtöbb esetben csökkenő pontosság értékeket kaptunk, kivétel az MFCC és a formáns frekvenciák esetében. Ezért további vizsgálatokat végeztünk 16 és 32 mértékű eltolást alkalmazva, ahol jelentősebb csökkenést tapasztaltunk a pontosságban.

Az 1. táblázatban látható minden egyes kísérlet esetében az elért pontosság, specificitás, és szenzitivitás értékek. Félkövérrel kiemeltük az egyes metrikák szerinti legnagyobb elért értékeket. Megfigyelhető, hogy minden esetben ezeket a maximális értékeket az MFCC jellemzővektor halmaz használata esetében kaptuk (1. táblázat). Továbbá megfigyelhető, hogy a formáns frekvenciák és azok sávszélességei együttes felhasználása javította az osztályozás pontosságát 73,6%-ról 76,4%-ra, azonban az így elért pontosság elmarad az MFCC-vel (82,4% - 84,1%) és a mel-sávos energiaértékekkel (74,2% - 83,5%) elért pontosságoktól.

A legjobb eredményünket (84,1%) összehasonlítva hasonló kutatások eredményeivel (50-86%) kijelenthető, hogy viszonylag nagy pontosságot voltunk képesek elérni, de természetesen, ahogy arra már a bevezetőben utaltunk, az eredmények nehezen összehasonlíthatók. Legpontosabb összehasonlítást a Magyar Depressziós Beszédadatbázison általunk publikált korábbi eredményekkel lehetséges megtenni, ahol is eltérő módszereket alkalmazva 83%-os (Kiss és Vicsi, 2017c) illetve 86%-os (Kiss és Vicsi, 2017b) pontosságot tudtunk elérni.

1. táblázat: Depresszió felismerésének leíró jellemzői az eltérő korrelációs mátrixok alapján

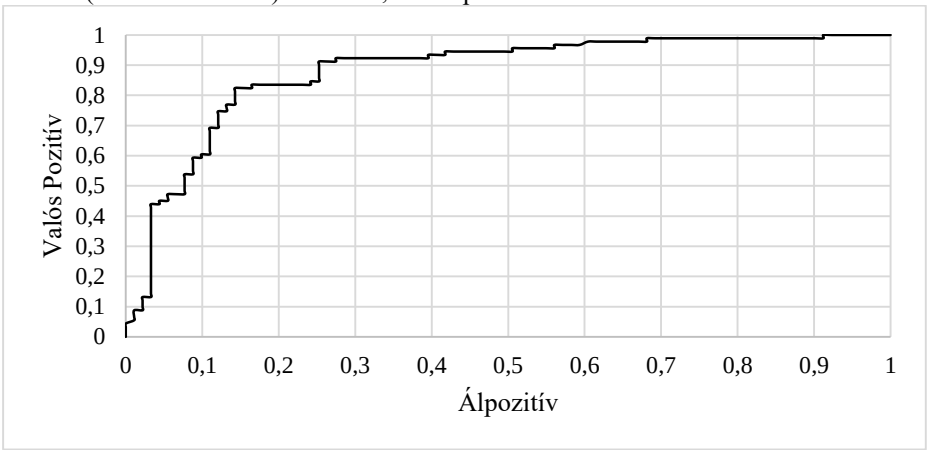
Jellemzővektor halmaz neve	Eltolás mértéke	Specificitás	Szenzitivitás	Pontosság
MFCC	1	81,3%	**83,5%**	82,4%
	2	83,5%	**83,5%**	83,5%
	8	**87,9%**	80,2%	**84,1%**
Mel-sávos energia értékek	1	84,6%	82,4%	83,5%
	2	80,2%	79,1%	79,7%
	8	73,6%	74,7%	74,2%
Formáns frekvenciák	1	76,9%	69,2%	73,1%
	2	73,6%	65,9%	69,8%
	8	70,3%	71,4%	70,9%
Formáns frekvenciák sávszélessége	1	78,0%	69,2%	73,6%
	2	82,4%	63,7%	73,1%
	8	78,0%	69,2%	73,6%
Formáns frekvenciák és azok sávszélességei	1	76,9%	75,8%	76,4%
	2	75,8%	75,8%	75,8%
	8	75,8%	74,7%	75,3%

Ez alapján megállapítható, hogy körülbelül ugyanakkora pontossággal voltunk képesek detektálni a depressziót mint korábban. Azonban fontos megjegyezni, hogy az általunk most bemutatott eljárás független a beszélő nemétől, és nem igényli a beszédminták bonyolult, beszédhang szintű előfeldolgozását, szemben a korábbi munkáinkkal. Viszont a korábbi eredményeink a Magyar Depressziós Beszédadatbázis egy régebbi állapotán készültek, ahol összesen csak körülbelül 130 beszélőtől állt rendelkezésünkre beszédminta. Összeségében kijelenthető, hogy az eredmények bizakodásra adnak okot és a bemutatott módszer további vizsgálata mindenképpen fontos.

Egy valós rendszer esetében a fő cél, hogy ha valaki depressziós azt a rendszer helyesen depressziósnak jelezze (valós pozitív arány), míg az kisebb hiba, hogyha valakit egészségesként dönt depressziósnak (álnegatív). Emiatt a legjobb pontosságot elérő beállítások mellett megvizsgáltuk, hogy egy adott valós pozitív arány mellett mekkora

lenne az álnegatívok aránya, amit a ROC (reciver operating characteristic) görbe megadásával szemléltetünk (5. ábra).

Ennek megvalósítására a neurális háló közvetlen kimenete adott lehetőséget, hiszen valójában 0 és 1 közötti számot adott vissza, ahol 0,5 értéknél kisebbek jelentették az egészséges osztályozást, míg az ennél nagyobbak a depressziós osztályozást. Így a komparátor értékét változtatva 0 és 1 között megfigyelhető, hogy adott valós pozitív arány mellett, mekkora lenne az álnegatív arány. A ROC-görbe integrálása alapján az AUC (area under curve) értékre 0,79-t kaptunk.



5. ábra. A depresszió detektálásának ROC görbéje MFCC jellemzővektor halmazból 8 mértékű eltolással számított korrelációs mátrixok alapján.

A ROC-görbe alapján megállapítható (5. ábra), hogy például 90%-os valós pozitív arányt elvárva az álnegatívok aránya már 25%. A gyakorlatban egy önálló diagnosztikát támogató rendszernek valószínűleg ennél nagyobb pozitív arány mellett kisebb álnegatív arányt kellene biztosítani, ahhoz hogy igazán jól alkalmazható lehessen, emiatt kívánatos lenne a módszer további fejlesztése.

5. Összefoglalás

Jelen kutatásban a depressziós állapot automatikus detektálásának lehetőségét mutattuk be beszédjel feldolgozás alapján. A kutatás eredménye hozzájárulhat egy a depresszió diagnosztizálását támogató minél pontosabb rendszer megvalósításához. Egy ilyen esetleges rendszer megvalósítása nagyban segíthetné a depresszió meglétének automatikus felismerését. A figyelmeztetés alapján az esetlegesen depressziótól szenvedő alany minél hamarabb megfelelő szakemberhez fordulhatna segítségért, ami megnövelheti a gyógyulás esélyét, illetve csökkentheti a kezelés időtartamát is. Ezek pedig csökkentenék a depresszió által okozott gazdasági károkat, illetve az öngyilkosságok számát.

A kutatásban a depresszió detektálásának lehetőségét a mel-sávos energiaértékek, az MFCC együtthatók, a formáns frekvenciák és azok sávszélességei, mint alacsony szintű akusztikai-fonetikai jellemzőkre támaszkodva ismertettük. A bemutatott alacsony

szintű akusztikai-fonetikai jellemzők adott részhalmazaiból képeztük azoknak egy speciális korrelációs struktúráját (mátrixát), amit mint bemenet kapott meg egy konvolúciós neurális hálót megvalósító gépi tanuló eljárás. Megvizsgáltuk, hogy mely akusztikai-fonetikai jellemzőhalmazra támaszkodva, milyen korrelációs struktúra esetén mekkora pontossággal detektálható a depressziós állapot. A vizsgálatok alapján legjobb eredményt az MFCC együtthatókra támaszkodva értünk el, 8 mértékű eltolást alkalmazva a korrelációs mátrix kialakítása során (84%-os pontosság). Az eredményt összehasonlítva más kutatások hasonló eredményeivel (50% - 86%-os pontosság), kijelenthető, hogy magas pontossággal voltunk képesek a depresszió automatikus felismerésére beszédjel feldolgozás alapján.

Az általunk bemutatott módszernek számos előnye van. A fő előnyei közt említhető, hogy független a vizsgált személy nemétől, nem szükséges hozzá bonyolult előfeldolgozása a beszédmintának (például beszédhangszintű szegmentálása) és az általunk bemutatott eredményt képesek voltunk csupán a beszéd MFCC együtthatóira támaszkodva elérni. Fontos továbbá azt is megjegyezni, hogy az általunk alkalmazott módszerekkel minimális volt a túltanulás veszélye.

Jelen kutatást mindenképpen folytatni tervezzük. A jövőben több vizsgálatot is tervezünk megvalósítani. Az eljárást tesztelni fogjuk a tovább bővített Magyar Depreszszziós Beszédadatbázison (200-200 egészséges és depressziós beszélőtől gyűjtött mintaszám a cél). A konvolúciós háló struktúrájának módosításával lehetővé tenni, hogy az eljárás egyszerre több és újabb akusztikai-fonetikai jellemző halmazokból elállított korrelációs mátrixokat is képes legyen a bemenetén fogadni. Egyéb olyan jellemzők felhasználásának megvizsgálása (pl: prozódiai jellemzők), amiket bár nem lehetséges vagy érdemes felhasználni a korrelációs mátrix(ok) előállítása során, azonban értékük bizonyítottan megváltozik a depressziós állapot hatására, így hasznosak lehetnek a depressziós állapot detektálásában. Nemek szerint eltérő modellek alkalmazása esetében megvizsgálnánk, hogy az vajon javít-e az általunk bemutatott módszer pontosságán.

Köszönetnyilvánítás

Project no. K128568 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K_18 funding scheme.

Hivatkozások

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

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

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

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

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T. F.: A review of depression and suicide risk assessment using speech analysis. Speech Communication 71, pp. 10–49, (2015)

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

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

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

Kiss, G., Simin, L., Vicsi, K.: Estimation of the severity of depression based on speech processing on Hungarian langugae (original title: Depresszió súlyosságának becslése beszédjel alapján magyar nyelven). In XIII. Magyar Számítógépes Nyelvészeti Konferencia,(MSZNY2017). Conference (pp. 125-135). (2017a)

Kiss, G., Vicsi, K.: Mono-and multi-lingual depression prediction based on speech processing. International Journal of Speech Technology, 20(4), 919-935. (2017b)

Kiss, G., Vicsi, K.: Comparison of read and spontaneous speech in case of automatic detection of depression. In 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000213-000218). IEEE. (2017c)

Kraepelin, E.: Manic depressive insanity and paranoia. J. Nerv. Ment. Dis. 53, 350. (1921)

Liu, Y. , Lee, T.., Ching, P. C., Law, T. K. T., Lee., K. Y. S.: Acoustic assessment of disordered voice with continuous speech based on utterance-level ASR posterior features" in INTERSPEECH 2017 pp. 2680–2684. (2017)

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

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

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

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

Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J., Skodda, S., Rusz J., Nöth, E.: Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson's disease. in INTERSPEECH 2015 pp. 95-99, Dresden, Germany, (2015)

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

Sztahó, D., Kiss, G., Tulics, M. G., Vicsi, K.: Automatic Separation of Various Disease Types by Correlation Structure of Time Shifted Speech Features. In 2018 41st International Conference on Telecommunications and Signal Processing (TSP) (pp. 1-4). IEEE. (2018)

Sztahó, D.; Kiss, G.; Tulics, M. G.; Dér-Hajduska, B.; Vicsi, K.: Automatic discrimination of several types of speech pathologies. In: 10th Conference on Speech Technology and Human-Computer Dialogue (SpeD 2019) Paper: 119 , 2 p.(2019)

Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., Zsura, F., Pákáski, M., Kálmán, J.: Automatic detection of mild cognitive impairment from spontaneous speech using ASR. In Proceedings of INTERSPEECH 2015 (pp. 2694-2698) (2015)

Valstar, M. F., Schuller, B. W., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., Pantic, M.: AVEC 2013: the continuous audio/visual emotion and depression recognition

challenge. in: 3rd ACM International Workshop on Audio/Visual Emotion Challenge, ACM. pp. 3–10, (2013)

Vlasenko, B., Sagha, H., Cummins, N., Schuller, B.: Implementing gender-dependent vowel-level analysis for boosting speech-based depression recognition. INTERSPEECH 2017, pp. 3266–3270, Stockholm, Sweden, (2017)

Williamson, J. R., Bliss, D. W., Browne, D. W.: Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG. In Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on (pp. 665-668). IEEE. (2011)

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