

Enyhe kognitív zavar automatikus felismerése szekvenciális autoenkóder használatával

Vetráb Mercedes¹, José Vicente Egas-López¹,
Balogh Réka², Imre Nóra², Hoffmann Ildikó^{3,4}, Tóth László¹,
Pákási Magdolna², Kálmán János², Gosztolya Gábor^{1,5}

¹ Szegedi Tudományegyetem, Informatikai Intézet

² Szegedi Tudományegyetem, Pszichiátriai Klinika

³ ELKH Nyelvtudományi Kutatóközpont, Budapest

⁴ Szegedi Tudományegyetem, Magyar Nyelvészeti Tanszék

⁵ ELKH-SZTE Mesterséges Intelligencia Kutatócsoport, Szeged
{ vetrabm, ggabor } @ inf.u-szeged.hu

Kivonat Az enyhe kognitív zavar (EKZ) heterogén klinikai szindróma. Főbb tünetei közé tartozik a memória, a gondolkodás, az érvelés és a nyelvi képességek romlása, amely azonban nem okoz jelentős zavart a páciensek mindennapi életvitelében. A hanyatlás enyhe foka és a lappangó tünetek miatt azonban az EKZ diagnosztizálása nagyon gyakran ütközik nehézségekbe. Ebben a tanulmányban szekvenciális autoenkódert használunk a jellemzőknyeréshez, hogy robusztus és hatékony attribútumokat extraktálhassunk. A felhasznált adattbázis 25 EKZ-s alany és 25 egészséges kontrollszemély hanganyagait tartalmazza. Eredményeink alapján ez a megközelítés versenyképes teljesítményt nyújt: egy nagyobb adattbázison tanított x-vektor hálóval szemben is képes jobb eredményeket nyújtani. További kísérleteinkben enyhe Alzheimer-kórban (eAK) szenvedő alanyokat is megpróbáltuk elkülöníteni.

Kulcsszavak: paralingvisztika, demencia, szekvenciális autoenkóder

1. Bevezetés

Az enyhe kognitív zavar (EKZ) heterogén tünetegyüttes, melyet főként a memória, a gondolkodás, az érvelés és a nyelvi képességek hanyatlása jellemez. Az EKZ-t gyakran tekintik átmeneti állapotnak a normál öregedés és a demencia között; a kognitív tünetek ebben a stádiumban ugyanakkor még nem befolyásolják jelentősen a páciensek mindennapi életvitelét (Petersen és mtsai, 2014; Alzheimer’s Association, 2020). Az EKZ megjelenése akár 15 évvel is megelőzheti a demencia klinikai manifesztálódását (Laske és mtsai, 2015), amely periódus lehetőséget nyújthat az EKZ korai felismerésére és ezáltal a kognitív hanyatlás progressziójának mérséklésére is (Hahn és Andel, 2011). Ennek ellenére nem csak az EKZ, hanem a demencia is globálisan aluldiagnosztizált kórkép (Lang és mtsai, 2017). Az alacsony felismerési arány rávilágít új, hatékony módszerek szükségességére, amelyek segíthetik a betegség korai szűrését.

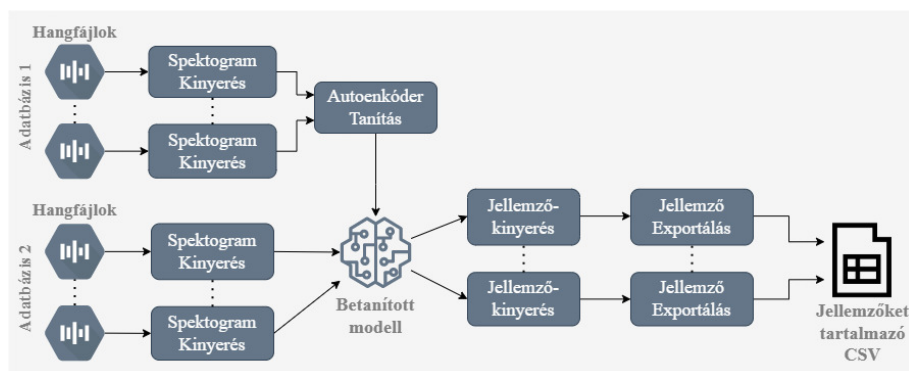
A nyelvi képességek változása az EKZ korai indikátoraként szolgálhat, mivel ezek a változások jóval az egyéb, jellegzetes kognitív tünetek megjelenése előtt is megmutatkozhatnak (McCullough és mtsai, 2019). Bizonyított, hogy a nyelvi teljesítményben bekövetkező változások összefüggést mutatnak a memória szubklinikai hanyatlásával, így például a spontán beszéd fluenssége romlik az EKZ-ban szenvedő páciensek esetében (Mueller és mtsai, 2018). Beszédük a betegség előrehaladtával egyre több szünetet és diszfluenciát tartalmaz (López-de Ipiña és mtsai, 2018), ami valószínűleg a betegek szótalálási nehézségeinek tulajdonítható (Szatlóczki és mtsai, 2015). Ezen beszédjellemzők elemzése ezért lehetővé teszi az EKZ közvetett vizsgálatát. Az automatikus beszédelemzés költség-hatékony, könnyen alkalmazható, non-invazív, és akár telemedicinás eszközként is használható lehet az EKZ tüneteinek kimutatására. Ugyan a közelmúltban számos tanulmány jelent meg az EKZ és a demencia egyéb formáinak detektálásáról (Haulcy és Glass, 2020; Pérez-Toro és mtsai, 2021), továbbra is vizsgálatok tárgyát képezi, hogy pontosan mely jellemzőtípusokat érdemes kinyerni az alanyok beszédéből. Egy lehetséges választás az általános (azaz a tényleges betegségre nem specifikus) jellemzőreprezentációs technikák alkalmazása, mint például az *i*-vektorok (Dehak és mtsai, 2009) és az *x*-vektorok (Snyder és mtsai, 2018). (Ezt a két technikát első sorban beszélőfelismeréshez dolgozták ki, de később más feladatokban is hasznosnak bizonyultak a jellemzők kinyerése során (Grzybowska és Kacprzak, 2016; Huckvale és mtsai, 2020)). Ezeknek a módszereknek a fő előnye, hogy a tanításhoz nincs szükség nagy méretű EKZ adatbázisokra, hanem általános célú, nagy beszédkorpuszokon előtaníthatók.

A mélytanulás-alapú technikák egyre növekvő népszerűségnek örvendenek. Ebben a tanulmányban szekvenciális autoenkódert használtunk a jellemzők kinyerésére. A célunk az volt, hogy megkülönböztessük az EKZ-s és az egészséges alanyok beszédét. Ezeket a technikákat korábban már sikeresen alkalmazták olyan feladatokra, mint például az akusztikus eseményfelismerés (Amiriparian és mtsai, 2017) vagy a főemlősök hangjainak kategorizálása (Schuller és mtsai, 2021). Meglátásunk szerint a végső jellemzőkinyerés során előnyt jelenthet, hogy közvetlenül a nyers hangfájlokat dolgozzuk fel, anélkül, hogy manuálisan konstruált jellemzőkészleteket használnánk (például MFCC-t).

Tanulmányunk újdonsága abban rejlik, hogy szekvenciális autoenkódereket használunk az enyhe kognitív zavar és az enyhe Alzheimer-kór kimutatására. Kísérleti eredményeinkben ez a technika felülmúlta az *x*-vektorok használatával kapott eredményeket, még akkor is, ha csak az adatok egy töredékén tanítottuk autoenkódereinket. Továbbá a peeking elkerülése és a robusztusság növelése érdekében, a kiértékelő halmaztól eltérő adatbázison tanítottuk a hálókat. Tudomásunk szerint ez az első olyan tanulmány amely keresztkorpusz-technikát alkalmaz szekvenciális autoenkódereknél.

2. Szekvenciális Autoenkóder

Az autoenkóder neurális hálók régóta jelen vannak a gépi tanulás területén (Hecht-Nielsen, 1995). A tanulási folyamatuk lényege, hogy a háló a bemenetét (ami nem



1. ábra: szekvenciális autoenkóder folyamatábrája.

feltétlenül hang) képes legyen rekonstruálni, miközben a bemenetből kinyert információt egy kis méretű rétegben tömöríti (*bottleneck layer*). Ez a tömörített reprezentáció felhasználható egy esetleges következő osztályozási lépés bemeneti jellemzőjeként. Hangfeldolgozás esetében a változó hosszú hangfájlok feldolgozására rekurrens neurális hálózatok vagy szekvenciális autoenkóderek is alkalmazhatók. Ezeket a technikákat korábban már sikeresen használták különböző feladatoknál (pl.: gépi fordításnál (Luong és mtsai, 2016), eseménydetektálásnál (Amiriparian és mtsai, 2017)).

A mélytanulási módszerek számára jellemzően jobbnak számítanak a nyersebb bemeneti jellemzők (pl. Mel-skála szerinti sávszűrők), mint a jobban előfeldolgozottak (pl. MFCC vagy PLP) (Mohamed és mtsai, 2011). Az autoenkóderek használatához első lépésben ki kell számítanunk a Mel-skála spektrogramokat a nyers hangokból. Amiriparian és munkatársai tanulmányát követve a Mel-spektrumokat a $[-1, 1]$ intervallumba normalizáljuk, ami jobban illeszkedik a neurális háló bemeneti tartományához (Amiriparian és mtsai, 2017). Ez a spektrum lesz a rekurrens neurális háló enkóderének bemenete. Az enkóder állhat akár Long Short-Term Memory (LSTM) vagy Gated Recurrent Unit (GRU) egységekből is.

Az enkóder háló utolsó rejtett rétegében lévő neuronok adják a bemenetből képzett tömörített (enkódolt) reprezentációt. Ezt a réteget a dekódoló háló követi, ami szintén LSTM vagy GRU neuronokból áll. A dekódoló feladata, hogy a tömörített reprezentációból rekonstruálja a bemenetet. A dekódoló szintén lehet egyirányú vagy kétirányú hálózat. A tanítás során a teljes hálót betanítjuk a bemenetek rekonstrukciójára. A háló súlyainak optimalizálása RMSE (Root Mean Square Error) hiba használatával történik, amelyet a bemenet és a kimenet eltéréséből kapunk. A betanítás után a háló középső, bottleneck rétege felhasználható jellemzőkinyeréshez. Ehhez elég megtartani az enkódoló hálót és annak megtanult súlyait.

Általában az autoenkóder hálót ugyanazon a korpuszon tanítjuk, mint amelyet majd a végső osztályozási feladathoz is használunk. Ennek egyik hátránya,

hogy a háló képes lehet megtanulni az adott adatbázisra jellemző tulajdonságokat. A másik hátrány, hogy az orvosi beszédfeldolgozás területén az adatbázisok mérete általában rendkívül kicsi, mivel az adott betegségben szenvedő alanyokat nehéz nagy létszámban elérni, valamint a pontos diagnosztizáláshoz szakszerűen képzett személyzetre (pl. orvosokra) van szükség. Ezeknek a hátrányoknak a kiküszöbölésére az enkóder hálót egy másik, általános célú adatbázison tanítottuk, hogy a jellemzőkinyerés minél robosztusabb maradjon. Az alkalmazott eljárás általános menetét az 1. ábrán szemléltettük.

3. Adatbáziskinyerés

A hangfelvételek a Szegedi Tudományegyetem Pszichiátriai Klinikáján kerültek rögzítésre. Ezekből a felvételekből összesen 50 beszélő anyagát választottuk ki. A felhasznált adatbázisrész 25 enyhe kognitív zavarral rendelkező beteg (EKZ) és 25 egészséges kontroll (EK) felvételeiből áll. A beszélőket az alapján választottuk ki, hogy a két vizsgálati csoport között ne legyen szignifikáns különbség a következők tekintetében: nem ($p = 0,734$), életkor ($p = 0,150$) és az oktatásban eltöltött évek száma ($p = 0,214$). Minden alany anyanyelvi magyar beszélő volt. Kizáró kritérium volt a kábítószer- vagy alkoholfogyasztás, a kognitív funkciókat befolyásoló gyógyszeres kezelésen való részvétel, a depresszió, a fejsérülés vagy pszichózis a kórtörténetben és a látás- vagy halláskárosodás. Az EKZ-betegeket orvosi diagnózist követően választották ki, amelyet neuropszichológiai vizsgálatokkal és CT vagy MRI vizsgálatokkal támasztottak alá. A spontán beszédre összpontosítottunk: a protokollunkban az alanyokat arra kértük meg, hogy meséljék el az előző napjuk eseményeit. A válaszokat diktafonnal vettük fel és csiptetős mikrofonnal rögzítettük. A válaszok időtartama 25...325 másodperc között volt, míg az átlagos időtartam 89,8 másodperc volt.

4. Kísérleti körülmények

4.1. Szekvenciális autoenkóderek

A kísérletek során az autoenkódert a BEA korpusz egy részhalmazán tanítottuk (Neuberger és mtsai, 2014), amely magyar spontán beszédet tartalmaz. A teljes BEA adatbázisból először leválogattuk azokat a felvételeket, amelyek a válaszadó alanyokhoz tartoztak. Ezután ezeket ezresével megkevertük a véletlenszerű sorrend érdekében és megtartottuk az első ezer felvételt. Ebben a részhalmazban 16 alany beszéde szerepel, amelynek teljes időtartama 3 óra 59 perc. Jellemzőkinyerésre és a neuronháló megvalósítására az AuDeep szoftvert használtuk (Freitag és mtsai, 2018), amely Python nyelven íródott. Előzetes tesztek eredményei alapján 128 log-skálájú Mel-spektrogram szűrőket alkalmaztunk 0,08ms széles ablakokkal és 0,04ms átfedéssel. Optimalizálóként az Adam algoritmust használtuk 0,001-es tanulási rátával, és 0,2 valószínűségű dropout-tal. A háló enkóder és dekóder része 2 – 2 rekurrens rétegből állt, amelyek mindegyike

128 GRU RNN neuront tartalmazott. A dekóder háló kétirányú volt. A modelleket 64-es mini-batch mérettel, valamint 32 epoch-kal tanítottuk. Ahogy korábbi szakirodalmak javasolják (lásd pl. (Amiriparian és mtsai, 2017; Schuller és mtsai, 2021)), mi is próbálkoztunk a háttérzaj eltávolításával. Az AuDeep szoftver az elkészített spektogramokat minden esetben 0 dB-hez normalizálja. A normalizálás után a programba beépített kapcsoló segítségével $-30/ -45/ -60/ -75$ dB-es küszöbérték alatt levágtuk a spektogramokat. A végső osztályozást elvégeztük a küszöbérték használata nélkül, valamint a 4 küszöbértékkel, és a négy küszöbértékkel kinyert jellemzővektorok egyesítésével (konkatenálásával) is.

4.2. Előfeldolgozás

Bár a szekvenciális autoenkóderek elméletileg képesek bármilyen időtartamú felvételt kezelni, az Audeep Tensorflow alapú megvalósításának korlátai miatt a gyakorlatban csak korlátozott hosszúságú hangfelvételeket tudtunk feldolgozni. Emiatt a tanítás előtt a BEA korpusz összes felvételét felosztottuk 5 másodperces darabokra. Demencia-adatbázisunkat a jellemzőkinyerés előtt szintén 5 másodperces darabokra kellett vágni. Az 50 felvételből összesen 1371 darabot kaptunk. Mivel az alanyok válaszait tartalmazó felvételek hossza változó volt ezért a felvételenkénti darabok száma 5-től 60-ig terjedt, valamint egy felvételhez átlag 27, 42 darab készült.

4.3. Osztályozás

Lineáris SVM osztályozó módszert használtunk (libSVM implementáció (Chang és Lin, 2011)); a C komplexitás paramétert a $10^{-5}, 10^{-4}, \dots, 10^1$ tartományban vizsgáltuk. A tanítás során 25-szörös keresztvalidációt (CV) alkalmaztunk, úgy, hogy minden egyes csoport (fold) egy egészséges és egy EKZ alany adataiból állt. A teljesítményt osztályozási pontossággal (egyenlő hiba (Equal Error Rate, EER) mellett vett döntési küszöbérték esetén) és görbe alatti terület (AUC) számításával mértük. Tanítás előtt a jellemzőket minden esetben standardizáltuk.

Az osztályozáshoz minden, azonos felvételekhez tartozó darabot, valamilyen módon együtt kellett kezelnünk a végső felvétalcímke meghatározásához. Ehhez két módszert is vizsgáltunk.

Az elsőben az osztályozást az 5 másodperces darabok szintjén végeztük el. Ezután minden felvétel (alany) osztályozásához vettük az adott felvételhez tartozó összes darab egyenkénti predikcióinak (poszteriorvektorának) számtani átlagát. Így mindkét osztályunkra (EKZ és EK) előállt egy-egy predikciós érték, amik közül a nagyobb alapján választottuk ki a predikált beszélőkategóriát.

Egy másik osztályozási módszerként, ahhoz, hogy megkapjuk az egyes felvételek osztályát, vettük az adott felvételhez tartozó összes darab jellemzővektorát majd ezeket a jellemzőket átlagoltuk és ezen a végső jellemzővektoron futtattuk le az osztályozást. (Ez bevett eljárás más beágyazás-alapú jellemzők, például x -vektorok esetén is.)

1. táblázat. A pontosság (Acc.) és az AUC értékek különböző tesztek során. Jellemzőkinyerés szekvenciálisautoenkódnél vágott és vágás nélküli adatokkal, és a vágások utáni jellemzők összefűzésével. A felvételek hosszát használva jellemzőnek és az X-vektor jellemzőkinyerésével.

Jellemzőkinyerési technika		Különálló		Átlagolt	
		Acc.	AUC	Acc.	AUC
Szekvenciális autoenkóder	-30 dB	64%	0,694	44%	0,482
	-45 dB	60%	0,706	56%	0,523
	-60 dB	68%	0,734	52%	0,478
	-75 dB	72%	0,763	64%	0,650
	Összefűzött	68%	0,643	48%	0,440
	Nincs vágás	68%	0,715	64%	0,603
Felvétel hossza		—	—	60%	0,615
x-vektor		—	—	60%	0,680

Viszonyításként egy másik modellt is tanítottunk, amelynek a felvételek hossza alapján kellett osztályoznia a mintákat, ahol a felvételek hosszát egyszerűen az 5 másodperces darabok számával közelítettük.

5. Eredmények

Az eredményeket megvizsgálva (1. táblázat) láthatjuk, hogy a frekvenciák egy bizonyos dB küszöbérték alatti levágása egyértelműen befolyásolja az EKZ osztályozásának sikerességét.

A tesztjeink során, a legalacsonyabb küszöbérték (−75 dB) eredményezte a legjobb pontosságot és AUC pontszámokat a felvételdarabok jellemzővektorainak különálló kezelésekor (72% és 0,763) és a vektorok átlagolásának esetében is (64% és 0,650). Ezekről a −60 dB-es vágáshoz tartozó eredmények sem maradtak el sokkal. Meglepő módon a négy jellemzővariáció összefűzése nem javított, hanem rontott az osztályozás pontosságán minden esetben. Habár a különálló darabok osztályozásával kapott 68%-os pontosság nem sokkal kevesebb, mint a legjobb 72%-os érték, a 0,643-as AUC-érték mégis itt a legalacsonyabb. A darabok jellemzővektorainak átlagolásakor az összefűzés jelentősen rontotta az osztályozást. A küszöbérték szerinti vágást nem tartalmazó hangfájlokból való jellemzőkinyerés eredményei érdekes eredményt hoztak. Nézzük először azt az esetet, mikor nem volt vágás és a hangfájlok darabjait először osztályoztuk és annak eredményét átlagoltuk. Ekkor szinte megegyező eredményeket kaptunk a korábbi, vágások jellemzőit összefűző módszerhez (68% és 0,715). Másodjára mikor a darabok jellemzőit átlagoltuk, a vágás nélküli eredményünk szinte megegyezett az ottani legjobb −75 dB-es eredményekkel (64% és 0,603).

Baselineként tanítottunk egy x-vektoros jellemzőkinyerő neuronhálót is a korábban említett BEA adatbázis 60 óra 14 másodpercnyi anyagával, 165 beszélővel és 40 Mel-frekvenciájú szűrősort alkalmazva („FBANK”). A szekvenciális

autoenkóderek legjobb pontosságai egyértelműen jobbnak adódtak, mintha az osztályozást csak a felvételek hossza alapján, vagy x -vektorral végeztük volna. Meglepő módon mind a pontosság, mind az AUC érték jelentősen alacsonyabb a szekvenciális autoenkóderekhez képest. Ez jól mutatja, hogy a tanulmányunkban alkalmazott módszertan valóban versenyképes.

5.1. Enyhe Alzheimer-kór vizsgálata

A következő kísérletben azt vizsgáltuk, hogy lehet-e a korábban használt jellemzőket három beszélő kategória megkülönböztetésére használni. Ehhez a 25 EKZ és a 25 kontroll alany (EK) mellett 25 enyhe Alzheimer-kóros (eAK) beteg beszédfelvételeit is felhasználtuk. Természetesen, az új csoport is megfelelt a korábbi csoportok statisztikai jellemzőinek az életkor, a nemi hovatartozás és az iskolai végzettségi szint tekintetében.

Az eAK betegek felvételeiből történő jellemzőkinyerésre a korábban betanított szekvenciális autoenkódereket használtuk. Az SVM modelljeinket a három osztály megkülönböztetésére tanítottuk, a korábban is használt keresztvalidációs módszerrel, tehát minden foldba egy-egy felvétel tartozott a három beszélőcsoportból.

A 2. táblázatban szemléltettük eredményeinket. Az AUC értékek tekintetében a kétszélyes esethez hasonlóan a -75 dB-es küszöbértéket használva kaptuk a legjobb eredményt. Azt is megfigyelhetjük, hogy a legnehezebb feladat az eAK betegek megkülönböztetése volt a többi beszélőtől. (Olyannyira, hogy a -30 dB-es vágás esetén egyáltalán nem sikerült őket azonosítani.) Ez meglepő, mivel a markásabb tünetek miatt általában könnyebb feladatnak számít az eAK alanyok megkülönböztetése az egészséges kontrolloktól, mint az EKZ enyhe tüneteinek kimutatása. Itt is megvizsgáltuk azt az esetet, mikor a hangfájlokat nem vágjuk le semilyen küszöbérték alatt. Az EKZ alanyok megkülönböztetésében hasonlóan jól teljesített, mint a legjobb -75 dB-es értékünk (0,703 és 0,710). Az eAK betegek felismerésében pedig ez a módszer bizonyult a legjobbnak (0,629).

6. Összegzés

Kutatásunkban az enyhe kognitív zavar (EKZ) detektálása volt a cél, amely az alany spontán beszédét is befolyásolja. A hangfelvételekből történő jellemzőkinyeréshez mély tanulást, azon belül szekvenciális autoenkóderet használtunk. A hálót olyan általános adatbázison tanítottuk, amely spontán beszédet tartalmaz. Ezáltal a háló tanítása és kiértékelése során keresztkorpusz technikát alkalmaztunk. A tömörített jellemzőreprezentációt az autoenkóderben lévő enkóder háló utolsó rétege adta. Kísérleti eredményeink azt mutatják, hogy ezzel az eljárással még jobb eredményeket értünk el, mint x -vektorok alkalmazásával, ráadásul az x -vektorok tanításához képest most a tanítóadat töredékét használtuk csak fel. Ezáltal kijelenthetjük, hogy a keresztkorpuszon tanított autoenkóder versenyképes a EKZ detektálásának területén. Eredményeink hozzájárulhatnak egy olyan

2. táblázat. Az AUC értékek a 3 osztályos kiértékelésnél. Jellemzőkinyerés szekvenciális autoenkódernél vágott és vágás nélküli adatokkal, és a vágások utáni jellemzők összefűzésével. A felvételek hosszát használva jellemzőnek és az X -vektor jellemzőkinyerésével.

Jellemzőkinyerési technika		AUC		
		EK	EKZ	eAK
Szekvenciális autoenkóder	-30 dB	0,706	0,618	0,503
	-45 dB	0,714	0,633	0,569
	-60 dB	0,732	0,706	0,606
	-75 dB	0,771	0,710	0,589
	Összefűzött	0,701	0,622	0,598
	Nincs vágás	0,682	0,703	0,629
Felvétel hossza		0,637	0,641	0,417
x-vektor		0,753	0,546	0,606

automatikus, nem invazív és költséghatékony EKZ-szűrésre alkalmas rendszer kialakításához, amely még csak nem is igényel személyes találkozást az alannyal (ahol a kontaktnélküliség egy, a Covid-19 világjárványhelyzethez hasonló időszakban kifejezetten hasznos).

Köszönetnyilvánítás

A tanulmányt a Nemzeti Kutatási Hivatal Fejlesztési és Innovációs Hivatala az NKFIH-FK-124413 szerződés keretein belül támogatta. Továbbá a Magyar Innovációs és Technológiai Minisztérium az NKFIH-1279-2/2020 számú pályázat és az Innovációs és Technológiai Minisztérium NKFIH Iroda Mesterséges Intelligencia Nemzeti Laboratóriumi Program (MILAB) keretében nyújtott támogatás. Gosztolya G. támogatásban részesült továbbá a Magyar Tudományos Akadémia által kiadott Bolyai János -ösztöndíjban és az Innovációs és Technológiai Minisztérium Új Nemzeti Kutatási Kiválósági Program ÚNKP-21-5-SZTE keretén belül.

Hivatkozások

- Alzheimer's Association: 2020 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* 16(3), 391–460 (2020)
- Amiriparian, S., Freitag, M., Cummins, N., Schuller, B.: Sequence-to-sequence autoencoders for unsupervised representation learning from audio. In: *Proceedings of DCASE*. pp. 17–21 (2017)
- Chang, C.C., Lin, C.J.: LIBSVM: A library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology* 2, 1–27 (2011)
- Dehak, N., Kenny, P., Dehak, R., Glembek, O., Dumouchel, P., Burget, L., Hu-beika, V., Castaldo, F.: Support Vector Machines and Joint Factor Analysis for speaker verification. In: *Proceedings of ICASSP*. pp. 4237–4240 (2009)

- Freitag, M., Amiriparian, S., Pugachevskiy, S., Cummins, N., Schuller, B.: auDeep: Unsupervised learning of representations from audio with Deep Recurrent Neural Networks. *Journal of Machine Learning Research* 18(173), 1–5 (2018)
- Grzybowska, J., Kacprzak, S.: Speaker age classification and regression using i-vectors. In: *Proceedings of Interspeech*. pp. 1402–1406 (2016)
- Hahn, E., Andel, R.: Nonpharmacological therapies for behavioral and cognitive symptoms of mild cognitive impairment. *Journal of Aging and Health* 23(8), 1223–1245 (2011)
- Haulcy, R., Glass, J.: Classifying alzheimer’s disease using audio and text-based representations of speech. *Frontiers in Psychology* 11, 624137 (2020)
- Hecht-Nielsen, R.: Replicator Neural Networks for universal optimal source coding. *Science* 269, 1860–1863 (1995)
- Huckvale, M., Beke, A., Ikushima, M.: Prediction of sleepiness ratings from voice by man and machine. In: *Proceedings of Interspeech*. pp. 4571–4575 (2020)
- López-de Ipiña, K., Martínez-de Lizarduy, U., Calvo, P., Beitia, B., García-Melero, J., Fernández, E., Ecay-Torres, M., Faundez-Zanuy, M., Sanz, P.: On the analysis of speech and disfluencies for automatic detection of Mild Cognitive Impairment. *Neural Computing and Applications* pp. 1–9 (2018)
- Lang, L., Clifford, A., Wei, L., Zhang, D., Leung, D., Augustine, G., Danat, I.M., Zhou, W., Copeland, J.R., Anstey, K.J., Chen, R.: Prevalence and determinants of undetected dementia in the community: A systematic literature review and meta-analysis. *BMJ Open* 7(2), e011146 (2017)
- Laske, C., Sohrabi, H., Frost, S., López-de Ipiña, K., Garrard, P., Buscema, M., Dauwels, J., Soekadar, S., Mueller, S., Linnemann, C., Bridenbaugh, S., Kanagasingam, Y., Martins, R., O’Bryant, S.: Innovative diagnostic tools for early detection of Alzheimer’s disease. *Alzheimer’s & Dementia* 11(5), 561–578 (2015)
- Luong, M.T., Le, Q., Sutskever, I., Vinyals, O., Kaiser, L.: Multi-task sequence to sequence learning. In: *Proceedings of ICLR* (2016)
- McCullough, K., Bayles, K., Bouldin, E.: Language performance of individuals at risk for Mild Cognitive Impairment. *Journal of Speech, Language, and Hearing Research* 62(3), 706–722 (2019)
- Mohamed, A.R., Dahl, G., Hinton, G.: Acoustic modeling using Deep Belief Networks. *IEEE Transactions on Audio, Speech and Language Processing* 20(1), 14–22 (2011)
- Mueller, K., Kosciak, R., Hermann, B., Johnson, S., Turkstra, L.: Declines in connected language are associated with very early mild cognitive impairment: Results from the Wisconsin registry for Alzheimer’s prevention. *Frontiers in Aging Neuroscience* 9, 437 (2018)
- Neuberger, T., Gyarmathy, D., Gráci, T., Horváth, V., Gósy, M., Beke, A.: Development of a large spontaneous speech database of agglutinative Hungarian language. In: *Proceedings of TSD*. pp. 424–431 (2014)
- Pérez-Toro, P., Bayerl, S., Arias-Vergara, T., Vásquez-Correa, J., Klumpp, P., Schuster, M., Nöth, E., Orozco-Aroyave, J., Riedhammer, K.: Influence of the interviewer on the automatic assessment of Alzheimer’s disease in the context

- of the ADReSSo challenge. In: Proceedings of Interspeech. pp. 3785–3789 (2021)
- Petersen, R., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., Fratiglioni, L.: Mild Cognitive Impairment: A concept in evolution. *Journal of Internal Medicine* 275(3), 214–228 (2014)
- Schuller, B., Batliner, A., Bergler, C., Mascolo, C., Han, J., Lefter, I., Kaya, H., Amiriparian, S., Baird, A., Stappen, L., Otth, S., Gerczuk, M., Tzirakis, P., Brown, C., Chauhan, J., Grammenos, A., Hasthanasombat, A., Spathis, D., Xia, T., P.Cicuta, Rothkrantz, L., Zwerts, J., Treep, J., Kaandorp, C.: The INTERSPEECH 2021 computational paralinguistics challenge: COVID-19 cough, COVID-19 speech, escalation & primates. In: Proceedings of Interspeech. pp. 431–435 (2021)
- Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., Khudanpur, S.: X-vectors: Robust DNN embeddings for speaker verification. In: Proceedings of ICASSP. pp. 5329–5333 (2018)
- Szatlóczki, G., Hoffmann, I., Vincze, V., Kálmán, J., Pákási, M.: Speaking in Alzheimer’s disease, is that an early sign? Importance of changes in language abilities in Alzheimer’s disease. *Frontiers in Aging Neuroscience* 7, 195 (2015)