



M Ű E G Y E T E M 1 7 8 2

Budapesti Műszaki és Gazdaságtudományi Egyetem  
Villamosmérnöki és Informatikai Kar  
Méréstechnika és Információs Rendszerek Tanszék

# Automatizált algoritmusok fejlesztése a kognitív hanyatlás korai diagnosztikájában

**TDK dolgozat**

Készítette:

Bolla Gergő Levente  
Második Szerző

Konzulens:

Dr. Hovráth András  
Dr. Antal Péter

2022

# Tartalomjegyzék

<b>Kivonat</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>1. Bevezetés</b>	<b>1</b>
<b>2. Szakirodalmi áttekintés</b>	<b>3</b>
2.1. Az Alzheimer-kór . . . . .	3
2.2. Enyhe kognitív zavar . . . . .	3
2.2.1. Képkalkotó vizsgálatok (MRI) . . . . .	4
2.2.2. Visuo-Spatialis rendszer . . . . .	4
2.3. funkcionális MRI és elemzés . . . . .	5
2.3.1. Előfeldolgozás (Preprocessing) . . . . .	6
2.3.2. Zajmentesítés (Denoising) . . . . .	6
2.3.3. Minőségellenőrzés . . . . .	7
2.3.4. <i>First Level Analysis</i> . . . . .	9
2.4. Strukturális MRI és elemzés . . . . .	10
2.4.1. Freesurfer . . . . .	10
2.5. Gépi tanulás . . . . .	10
2.5.1. feature selection . . . . .	11
2.5.2. Random Forest . . . . .	12
2.5.3. Support Vector Machine (SVM) . . . . .	13
2.5.4. Kereszt Validáció (KV) . . . . .	16
2.5.5. Hiperparaméter optimalizáció . . . . .	17
<b>3. Anyag és módszer</b>	<b>18</b>
3.1. Elemzések automatizációja . . . . .	18
3.2. Alanyok . . . . .	18
3.3. MR vizsgálat . . . . .	18
3.4. Felvételek elemzése . . . . .	19
3.4.1. Strukturális elemzés . . . . .	19
3.4.2. rs-fMRI felvételek elemzése . . . . .	19
3.5. Gépi tanulás . . . . .	20
<b>4. Eredmények</b>	<b>21</b>
4.1. Elemzések automatizálása . . . . .	21
4.2. Adattisztítás . . . . .	28
4.3. Gépi tanulás . . . . .	30
4.3.1. rs-fMRI . . . . .	30
4.4. Feature selection . . . . .	33
4.4.1. Strukturális MRI . . . . .	35

**5. Következtetés**

**37**

**Irodalomjegyzék**

**40**

## HALLGATÓI NYILATKOZAT

Alulírott *Bolla Gergő Levente*, szigorló hallgató kijelentem, hogy ezt a TDK dolgozatot meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző(k), cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy autentikált felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Budapest, 2022. november 1.



---

*Bolla Gergő Levente*  
hallgató

# Absztrakt

Az Alzheimer-kór (AK) az időskori demencia egyik leggyakoribb oka, mely a szellemi képességek romlásával jár. Az AK jelenleg világszerte jelentős problémát okoz, aminek legfőbb oka a késői diagnózis. Differenciál diagnosztikája jelenleg klinikai vizsgálatokon (neuropszichológiai tesztek, strukturális MRI (Magnetic Resonance Imaging) és liquor vizsgálatok) alapul. Egyik enyhébb, általában a tényleges betegséget megelőző formája – amellyel én is foglalkozom - az enyhe kognitív zavar (Mild Cognitive Impairment; MCI). Az MR felvételek elemzése során több probléma is felmerül: a strukturális felvételek elemzése csak UNIX rendszeren futtatható; a funkcionális felvételek elemzésénél különböző előfeldolgozási lépéseinek alkalmazásában nincs egységes konszenzus, továbbá minden kiértékeléshez alap programozói tudás szükséges.

A diplomamunka célja egy olyan felület létrehozása Python segítségével, amely lehetővé teszi mind a strukturális mind az fMRI felvételek elemzését és emellett lehetőség van az ezekből készült adatok automatikus adatbázisba való felvételére. További cél egy adatbázis létrehozása MCI-os betegekkel és nemben, valamint korban megegyező kontroll alanyokkal, majd az adatbázis felhasználása adatbányászathoz és gépi tanulás alapú tanításhoz, biomarkerek meghatározásához.

A diplomamunka során egy olyan GUI-t hoztam létre Pythonban ami automatizálja az MRI adatok elemzését. A CONN Toolboxot a MATLAB Python API segítségével a Freesurfer programot pedig a WSL-el integráltam. A programmal összesen 78 alanyt (32 MCI és 46 kontroll) elemeztem 2 kutató központból, majd az elemzett adatokat felhasználtam gépi tanulás alapú klasszifikációhoz és biomarker identifikációhoz. A gépi tanuláshoz három konnektivitási mértéket (IC, LCOR, fALLF) használtam fel az rs-fMRI mérésekből. Egy SFFS nevű algoritmust alkalmaztam a megfelelő változók kiválasztására, amiket a Random Forest és Support Vector Machine algoritmussal validáltam. A strukturális MRI adatokból pedig a releváns régiókat kiválogattam majd két mintás T próbát hajtottam végre a két csoport közötti különbség megítélésre.

Az általam készített GUI lehetővé teszi mind két fajta elemzést illetve az ebből létrejött adatok adatbázisba való felvételét. A gépi tanulás alapú algoritmusok során a legjobb eredmény 87% volt, amit a fALLF metrikával sikerült elérni. A strukturális adatok során kiderült, hogy az Entorhinális kéreg szignifikánsan kisebb ( $p < 0, 0001$ ) az MCI-os alanyokban, mint a kontroll csoportokban.

A GUI leegyszerűsíti és megkönnyíti az elemzéseket, illetve az ebből született adatok feldolgozását. A feature selection algoritmusok által kiválasztott változók, valamint a strukturális elemzésekből származó Entorhinális kéreg biomarkerként szolgálhatnak a betegség diagnosztizálásában. Az rs-fMRI metrikákat a gépi tanulás alapú algoritmusokkal pedig egyéb betegségek esetén is lehet alkalmazni.

# Abstract

Alzheimer's Disease (AD) is a serious neurological condition affecting elderly patients, and it's the most common cause of dementia. Differential diagnosis of AD consists of clinical tests (e.g., neuropsychological tests; MRI: Magnetic Resonance Imaging and cerebrospinal fluid analysis). A milder and prodromal condition in which AD can be prevented is MCI (Mild Cognitive Impairment). There are several problems associated with the analyses of MRI images: The structural files can only be analysed through UNIX based systems. Another concern regarding the analysis is that there isn't a common consensus about the pre-processing pipeline. Moreover, these assessments require basic programming knowledge.

The aim of this thesis was to create a Python framework which makes the analysis possible via Python, secondly it adds the created results to a pre-defined database and finally to create simplified functions which create opportunity to analyse both structural and functional MRI data without programming knowledge. Another goal is creating a database with MCI patients and controls and applying the database for a machine learning based classification and biomarker identification.

I created a GUI in Python that automates the analysis of MRI data. I integrated the CONN toolbox with the MATLAB Python API and I used WSL for the integration of Freesurfer. I analysed 78 patients (32 MCIs and 46 controls) from two research centres and used their data for classification with machine learning. Three connectivity measures from rs-fMRI measurements (IC, LCOR, fALLF) were used during the analysis. SFFS algorithm was applied to select the appropriate features. From the structural MRI data, I selected the relevant regions and used a two sample T test to assess the differences between the two groups.

The GUI enables the analyses of both the structural and functional MRI scans. Regarding to machine learning classification, 87% was the best accuracy with the fALLF metric. The structural data revealed that the Entorhinal cortex is significantly smaller ( $p < 0.0001$ ) in the MCI group compared to the healthy controls.

The GUI simplifies the analyses and data recording. The features selected by the SFFS algorithm and the Entorhinal cortex can serve as biomarkers for the diagnosis of the disease. Furthermore, the rs-fMRI metrics combined with the algorithms mentioned can be used on other conditions and datasets as well.

# 1 Bevezetés

Az Alzheimer-kór (AK) az időskori demencia egyik leggyakoribb oka, mely a szellemi képességek romlásával jár. Az AK jelenleg világszerte jelentős problémát okoz [1]. A legnagyobb probléma, hogy csak a súlyos tünetek megjelenésénél kerül sor vizsgálatra és a diagnózis felállítására. Az AK differenciál diagnosztikája jelenleg klinikai vizsgálatokon (neuropszichológiai tesztek; strukturális MRI (Mágneses Rezonancia Képképzés; Magnetic Resonance Imaging), liquor analízis) alapul. Az AK megelőző formája az enyhe kognitív zavar (Mild Cognitive Impairment; MCI). A betegség e fázisában még van esély megelőzni/megakadályozni a súlyos tünetek kialakulását [2]. Mind AK-ban, mind MCI-ban a jelenlegi kutatások arra törekednek, hogy olyan biomarkert találjanak, amely előre jelezheti az AK kialakulását vagy progresszióját. Ezen kutatások döntő többsége képképzés alapú (PET-CT, strukturális MRI, funkcionális MRI, stb.) [3]. A PET-CT nagy hátránya, hogy radioaktív izotópok használatával történik. Az MR vizsgálat ezzel szemben ionizáló sugárzás nélkül végrehajtható és relatíve könnyen elérhető. Az MR vizsgálatoknál a felvételek kiértékelése történhet kvalitatív, vagy kvantitatív módon. A kvantitatív kiértékelés alkalmával kerülhet sor többek között a szürkeállomány vastagságnak vagy volumenének meghatározására. A nyugalmi állapotú funkcionális MR vizsgálatokkal pedig feltérképezhetőek a funkcionális agyi kapcsolatok [4], melyek hiánya, megerősödése vagy gyengülése jelezheti MCI, vagy AK jelenlétét. A kiértékelésre jelenleg számos megoldás áll rendelkezésünkre, viszont a strukturális MR felvételek elemzése csak UNIX rendszereken történik, továbbá alap szintű programozói tudásra is szükség van. Tovább nehezíti az elemzést, hogy nincs egységes konszenzus az előfeldolgozási lépésekkel kapcsolatban (pl. szükséges-e minden elemzés előtt mozgáskorrekció stb.). Az előfeldolgozási lépések eltérő alkalmazása pedig különböző eredményt produkálhat [5] [6].

Az fMRI kutatásokban köszönhetően az egyre nagyobb és jobb minőségű adathalmaznak egyre gyakrabban használnak gépi tanuláson alapuló algoritmusokat. A leggyakoribb ezen belül a független komponens analízis, amikor az fMRI jelet komponensekre bontják, hogy a zaj természetű jeleket eliminálják. [7] [8]. Manapság pedig egyre többször használnak betegségek osztályozására felügyelt tanuló algoritmusokat [8] az AK, MCI és az egészséges alanyok elkülönítésére pl. *multi kernel learning* [9] vagy *deep learning* alapú *sparse autoencoder* hálózat [10] is.

Utóbbi csoportba tartoznak az ún. feature selection algoritmusok is, amelyekkel a redundáns vagy irreleváns változókat vagyunk képesek eltávolítani az információvesztés minimalizálásával [8]. Ezek közé tartozik többek között a *Random Forest* végrehajtott *feature selection* [11], *Relieff* filter algoritmus [12], permutáció alapú szűrések [11] vagy ezek kombinációja.

Az eddigi szakiroladmi adatok és a kapott, machine learning alapú elemzések döntő többsége az *Alzheimer's Disease Neuroimaging Initiative* (ADNI) [13] adatbázison alapszik.

A diplomamunka fő célja egy multicentrumos adatbázis létrehozása, amely alapja gépi tanuló algoritmusoknak. További cél egy olyan keretrendszer létrehozása Python nyelv segítségével, amely egyrészt lehetővé teszi az elemzéseket Python-ból, másrészt a kapott eredményeket automatikusan egy előre megadott adatbázisba emeli, harmadrészt olyan, leegyszerűsített függvények létrehozása a cél, amellyel programozói tudás nélkül is elvégezhetőek mind a strukturális, mind a funkcionális MRI felvételek kiértékelései. A keret-

rendszert egy GUI-hoz rendelve pedig leegyszerűsíti annak használatát, illetve a navigáció és a felhasználhatóság is könnyebbé válik.



## 2 Szakirodalmi áttekintés

### 2.1. Az Alzheimer-kór

Az AK egy progresszív neurológiai kórkép, mely az agy atrophijához (zsugorodásához) és idegsejtek elhalásához vezet. Az időskori demencia egyik leggyakoribb oka. Koraibb stádiumban a nehéz emlékezés figyelhető meg, és ahogy a betegség előrehalad ezek a tünetek egyre súlyosabbak lesznek. Először akár maga az elszenvető személy is tudatában lehet annak, hogy csökken az emlékező képessége, de gyakran csak családtagjai, barátai veszik észre, az állapotromlást. A memóriavesztésen kívül a gondolkodás is hanyatlani kezd, romlik a megítélés és a döntéshozó képesség is. A tervezés és az ismert feladatok végrehajtása is nehézséget okozhat és akár az egyén személyisége is megváltozhat. Depresszió, apátia alakulhat ki. A betegség okai nem teljesen ismertek, de általánosságban elmondható, hogy az agyban lévő fehérjék kóros anyagcseréje megzavarja a neuronok szabályos működését, ami toxikus események bekövetkezéséhez vezet. A neuronok sérülnek, a köztük lévő kapcsolatok megszűnnek, majd elhalnak. Rizikó faktort jelenthet rá a 65 év feletti életkor, családban előforduló korábbi AK, Down szindróma megléte és az MCI [1] [14] [15].

### 2.2. Enyhe kognitív zavar

Az MCI a demencia egy korai stádiuma, ahol az érintett egyén kognitív képességei (vizuális/ térbeli vagy beszédbeli észlelés) romlani kezdenek. Az MCI során fellépő kognitív változások már elég súlyosak ahhoz, hogy az elszenvetője vagy az ő közvetlen környezete észlelje a változást, de még ahhoz nem, hogy az egyént gátolják a mindennapi tevékenységekben. A betegség pontos okai nem ismeretek, azonban azért van nagy jelentősége mert AK kialakulásához vezethet, viszont ebben az állapotban a betegség prognózisa még lelassítható, ezzel nagyban javítva az egyén életminőségét. A betegséget két fő csoportba sorolhatjuk az alapján, hogy a páciens, milyen problémákkal küzd:

- *amnesztikus MCI*: MCI, mely elsősorban a memóriát érinti. A beteg elkezdhet fontos információkat elfelejteni.
- *Nonamnesztikus MCI*: MCI, mely elsősorban a gondolkodási képességeket érinti. Ilyen tünetek pl. a megfelelő döntések hozatalának csökkenése vagy, ha egy komplex feladat lépéseit nem megfelelően ítéli meg.

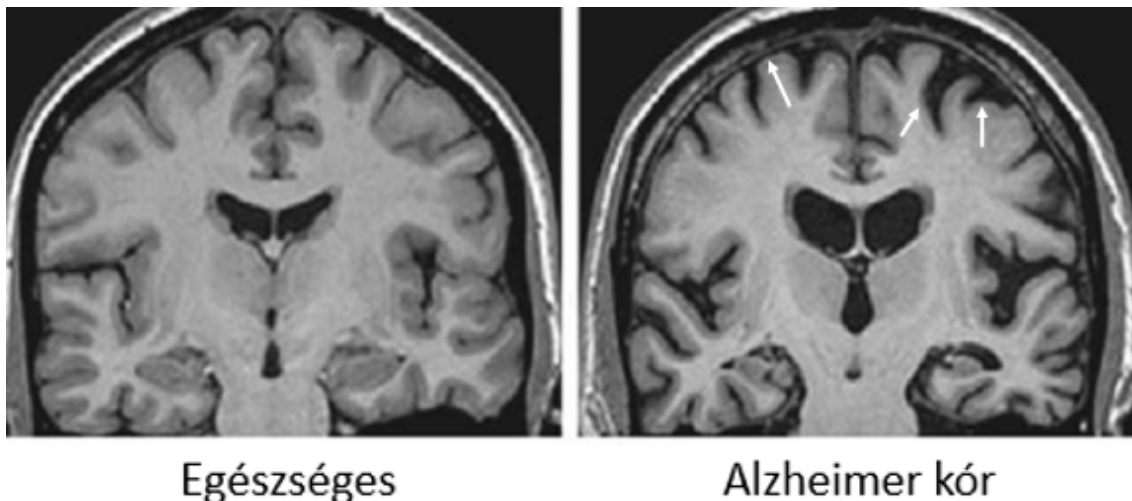
A diagnózis felállítása meglehetősen nehéz feladat hiszen nem pontosan ismertek a betegség okai. Utalhatnak rá az agyban lévő morfológiai változások, kóros liquortenyésztés, ahol az amyloid beta fehérjét vizsgálják vagy csökkent glükóz felvétel az agyban, melyet PET szkennel tudunk mérni. Mikor ezek valamelyike merül fel akkor szinte biztosak lehetünk abban, hogy a betegség AK felé progrediál. A biztos diagnózishoz fontos a megfelelő biomarkerek identifikációja annak érdekében, hogy elkülöníthessük azokat a pácienseket, akik AK miatt szenvednek MCI-ban [2] [16] [17].

A jelenlegi diagnosztizálás a következő főbb elemeken alapszik [2]:

- Kórtörténet, mely során a jelenlegi tüneteket korábbi betegségeket és a familiaritást vizsgálják.
- Neuropszichológiai teszt mely során a kognitív képességeket mérik fel.
- Labor vizsgálatok pl. Liquor tenyésztés.
- Képalkotó vizsgálatok (PET, MRI).

### 2.2.1. Képalkotó vizsgálatok (MRI)

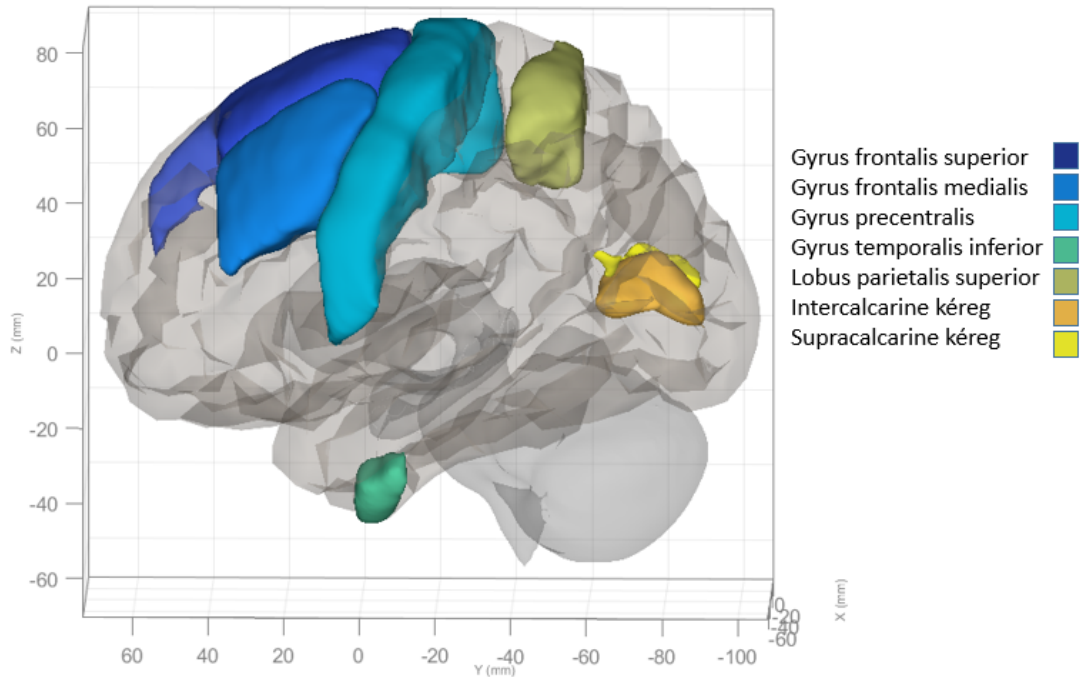
Az elmúlt években az MRI nagy fejlődésen ment keresztül és mára már a diagnosztika egy alappillérenek számít. Kiváló térbeli felbontása mellett jó kontrasztot biztosít a különböző szövetek között. Ehhez a szervezetünkben nagy mennyiségben megtalálható hidrogén atommagot használja fel, a melyek kis mágnesként működnek. Alkalmazva a nukleáris mágneses magrezonancia (NMR) elvét a hidrogén atommagokat úgy tudjuk manipulálni, hogy jelet generáljanak, mellyel képet tudunk alkotni. Képalkotó vizsgálatok közül az MRI és a PET az, amikkel jól diagnosztizálható mind az AK, mind az MCI. Az MRI vizsgálatok esetében agyi atrophíát, tehát az agy zsugorodását és az agyi régiók térfogatának csökkenését vizsgáljuk. Habár az atrophia CT segítségével is megítélhető az MRI sokkal érzékenyebb ezekre a változásokra és jobban kiszűrhetjük vele az egyéb okból bekövetkező atrophíát (pl. infarktus okozta Demencia) [18]. Az atrophia előrehaladott AK esetekben jól látható az MRI felvételeken (2.1 ábra) azonban ez nem mindig ilyen egyértelmű. Korábbi stádiumokban pl. MCI-ban még nem akkora a térfogat csökkenés mértéke így ilyenkor kvantitatív mérésekre van szükség amik számszerűen adják meg ezen térfogatok méretét.



2.1. ábra. egészséges agy (bal) és AK-ban szenvedő páciens agya (jobb) ahol megfigyelhető a szűkreállomány atrophíája

### 2.2.2. Visuo-Spatialis rendszer

A visuo-spatialis (VS) rendszer olyan kognitív funkciók rendszere, amely fontos a vizuális és strukturális információk felismerésében, interpretációjában és elemzésében. A szakirodalmakban egyre több bizonyíték van [19] [20] [21] rá, hogy a VS rendszer hanyatlása jelen van az MCI betegségnél így a diplomamunka során ezt a rendszert külön vizsgáltam az rs-fMRI elemzések esetében. A VS rendszerben lévő régiók a 2.2 ábrán láthatóak



2.2. ábra. Visuo-spatialis rendszerben lévő régiók

### 2.3. funkcionális MRI és elemzés

A 90-es évektől lehetőség nyílt arra, hogy nem csak a strukturális de funkcionális felvételeket is készítsünk az agyról. Ez az ún. BOLD (Blood Oxygen Level Dependent; Vér Oxigén Szint Független Jel) jel miatt lehetséges. Az agyban lévő neuronok aktiválásához energiára van szükség, amit a véráram biztosít. Ilyenkor ezekre a területekre oxigén dús vér érkezik, melynek más mágneses tulajdonságokkal rendelkezik, mint az oxigénben szegény/széndioxidban dús vérnek. Az oxigénben szegény vér paramágneses tulajdonságú, torzítja a mágneses teret, így jelcsökkenést okoz. Ezzel szemben az oxigénben dús vér diamágneses, így mikor oxigéndús vér érkezik a neuronokhoz, jelemelkedés lép fel és ezt hívjuk BOLD jelnek. A BOLD jel egyik felhasználási módja, ha valamilyen feladatot végeztetünk, a pácienssel pl. ütögesse össze az ujjait – ekkor a mozgató neuronokhoz oxigén dús vér fog áramlani így feltérképezhető a mozgató kéreg. Viszont nem csupán cselekvés során lép fel neurális aktivitás, hanem nyugalmi állapotban is. Ezt nevezzük resting state fMRI-nek (rs-fMRI) [4] [22] [23]. Ennek alapja az alacsony frekvenciás BOLD jelek fluktuációja. Amikor két agyi terület BOLD jelének változása korrelál egymással, akkor a két régió funkcionális összeköttetésben van egymással. Ennek megállapításához számos előfeldolgozási lépést kell eszközölnünk, amelyek célja a BOLD jel „tisztítása”. Ezekre azért van szükség mert a felvételek nagyon érzékenyek a zajra pl. a páciens mozgására, fiziológias mozgásokra vagy magából az MR-ből fakadó műtermékekre (pl. inhomogén mágneses tér torzítása) [24]. Másrészt mivel több száz felvétellel dolgozunk szükség van statisztikai módszerekre, hogy a megfelelő információt nyerjük ki ezekből. Az elemzések megvalósítására számos szoftver áll rendelkezésünkre és mindnek megvan az előnye és hátránya is. A diplomamunkám során a MATLAB alapú CONN toolboxot [25] használtam.

Az rs-fMRI felvételek elemzése során a főbb lépések:

- *Előfeldolgozás (Preprocessing)*: Egy több lépésből álló folyamat melynek fő feladata a műtermékek eltávolítása és a felvételek előkészítése további analízisre.

- *Zajmentesítés (Denoising)*: Az előfeldolgozási lépések után még mindig marad zaj a felvételekben, melyek a mozgásból, fiziológiás artefaktumokból erednek. Ezek szűrése nélkül fals eredményeket kaphatunk a későbbi lépések során, így fontos ezek eltávolítása.
- *First-level analyses*: Ebben az lépésben az egyénben lévő funkcionális konnektivitást vizsgáljuk különböző módszerekkel. Ez a lépés ekvivalens az alanszintű vagy másnéven *single subject* elemzéssel
- *Second-level analyses*: Ebben a lépésben a vizsgált csoport egészét vizsgáljuk és nem csak magát az egyént

### 2.3.1. Előfeldolgozás (Preprocessing)

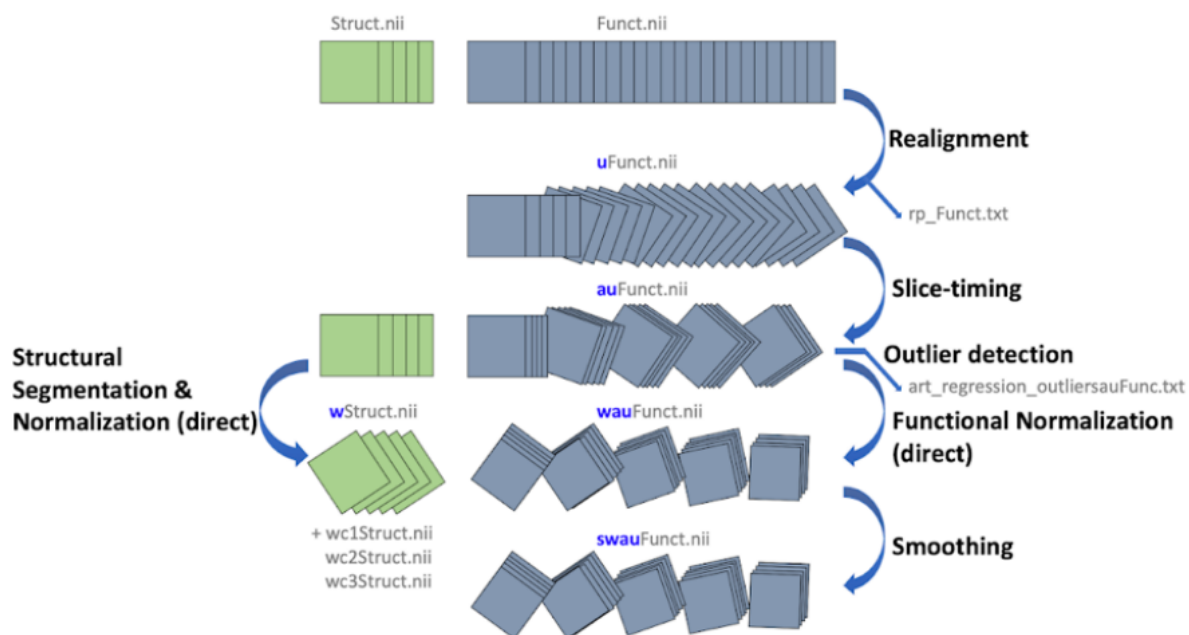
A szkennerből származó mérések nagy valószínűséggel, műtermékekkel tarkítottak lesznek, melyek a páciens mozgásából, fiziológiás mozgásokból és magából a szkennerből származhatnak. Az előfeldolgozás során alkalmazott főbb lépések a következők [24] [26]:

- *Realignment*: Az alany mozgásának kiküszöbölésére szolgáló lépés.
- *Susceptibility Distortion Correction*: Az MR-ben lévő inhomogén mágneses tér torzításának javítására szolgáló lépés
- *Slice Timing Correction*: Az agyról készült szeletek az MR-ben szekvenciálisan, nem egy időben készülnek így ez a lépés ezt kívánja kiküszöbölni.
- *Co-registration, Normalization*: Az anatómiai és funkcionális felvételek regisztrációját a co-registratio alatt végezzük el. Amennyiben csoportos elemzésre van szükség, elkerülhetetlen a normalizáció, melynek során az alanyok strukturális- és funkcionális felvételeit egy koordináta rendszerbe transzformáljuk. A csoportos elemzéshez egy ún. MNI atlaszt alkalmazunk, ami egy 241 normál agy átlagából készült atlasz.
- *Outlier detection*: A BOLD jelben és mozgásban kiugró értékek megtalálása.
- *Smoothing*: A *smoothing* (simítás) során egy kernelt (ált. Gauss kernelt) alkalmazunk a felvételeken, ami elsimítja a képet és eltávolítja a nagy frekvenciás információkat.

A CONN Toolboxban elérhető default preprocessing pipeline a 2.3 ábrán látható

### 2.3.2. Zajmentesítés (Denoising)

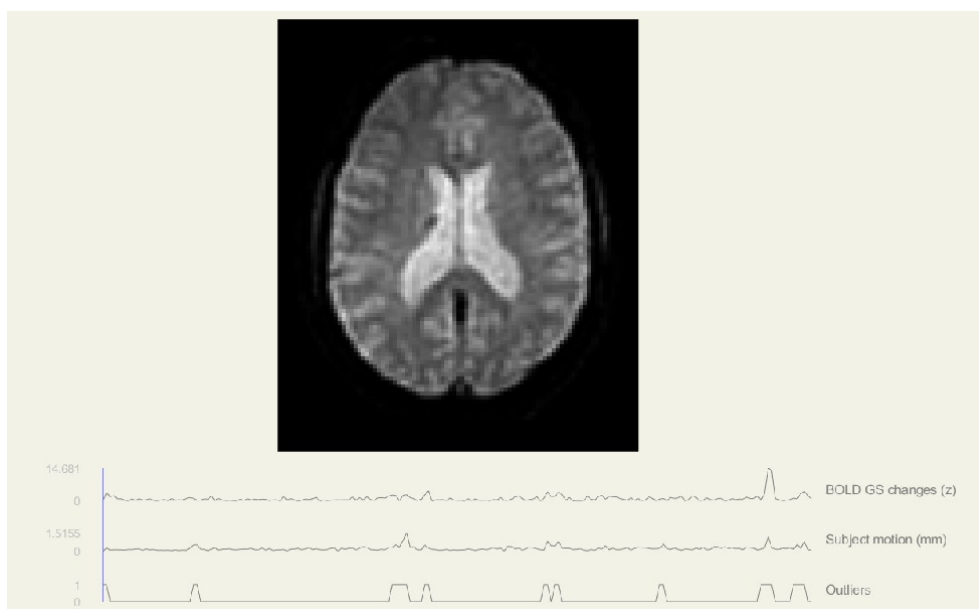
Mivel az előfeldolgozási lépések után is marad artefaktum a felvételekben (pl. fiziológiás mozgásból) így a zajmentesítés lépése is fontossá válik az elemzések során, hogy ne kapjunk fals eredményeket. Ez két fő lépésből áll: A potenciális *confounderek* (jelet befolyásoló zajforrás) lineáris regressziója és időbeli *band pass filter* alkalmazása [27] [28].



2.3. ábra. CONN Toolbox default preprocessing pipeline [27]

### 2.3.3. Minőségellenőrzés

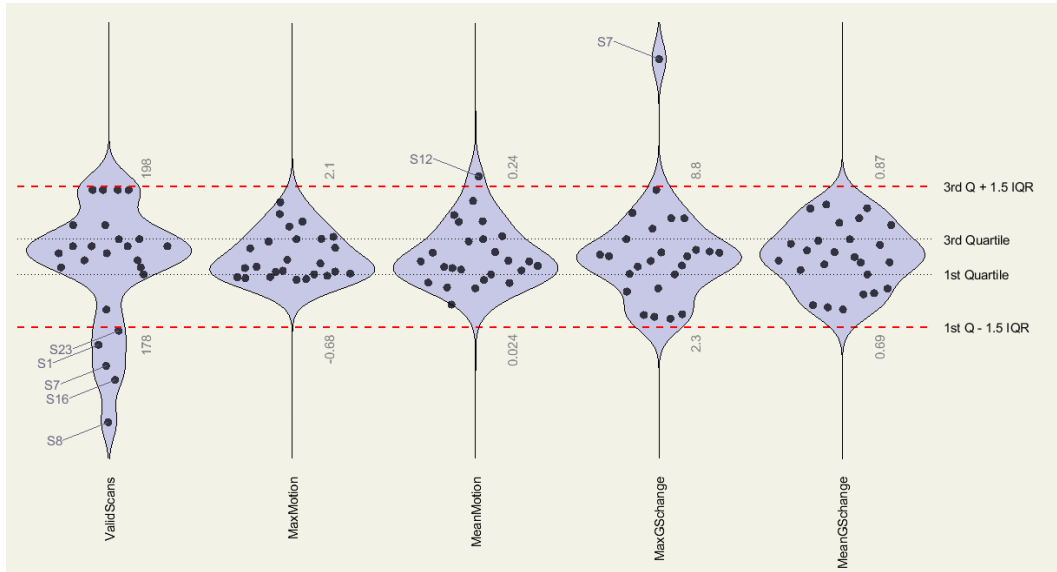
A CONN toolboxon belül rendelkezésre állnak olyan függvények, amelyekkel kvantitatívan meghatározható a felvételek minősége, pl. milyen irányú mozgásokat végzett a beteg, van-e kiugró érték a BOLD jelben, ami műtermékre utalhat stb. Ezt szemlélteti a 2.4 ábra.



2.4. ábra. Becsült BOLD jel változás (*felső*), páciens mozgása (*középső*), ezekből eredő kiugró értékek (*alsó*)

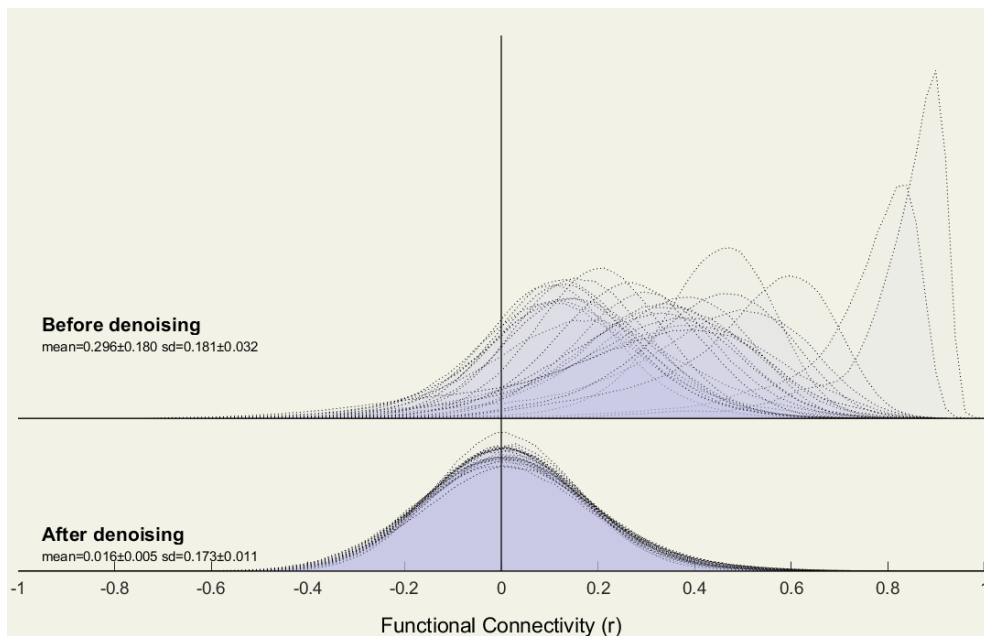
Egy másik ehhez hasonló ábrát mutat a 2.5 ábra, melyen boxplothoz hasonlóan vizsgálhatjuk meg a kiugró értékeket az átlagos és maximális mozgásról, BOLD jel változásról

és a megfelelő szkennek számáról. A kvantitatív minőségellenőrzés után eldönthetjük, hogy egy-egy alany elemzését kizárjuk-e, vagy sem.



2.5. ábra. kiugró értékek vizsgálata boxploton

Végül pedig a Zajmentesítést is ellenőrizhetjük. Ennek ellenőrzése végett a funkcionális konnektivitás eloszlását vizsgáljuk random voxelek közt az agyban 2.6. Zajmentesítés előtt (Before denoising) megállapítható, hogy a funkcionális konnektivitás értékek a pozitív korreláció felé húzódnak. Zajmentesítés után (After denoising) viszont egy normál eloszlást kapunk [27].



2.6. ábra. Funkcionális konnektivitás értékek eloszlása (random voxelek között), zajmentesítés előtt (felső), és után (alsó)

### 2.3.4. *First Level Analysis*

A First level analyses különböző agyi területek között méri az időbeli korrelációt BOLD jel fluktuációban. Ennek megállapítására számos eljárás létezik, melyek közül az egyik legelterjedtebb az ún. ROI-to-ROI konnektivitás [27]. A módszer előre meghatározott Region of Interest-ek (ROI-k) között vizsgálja a konnektivitást. A ROI-k egy alapértelmezett atlaszból származnak, amiben a Harvard-Oxford [29] és az AAL [30] atlaszokból származó anatómiai területek szerepelnek. A diplomamunka során három hálózati metrikákat választottam ki, amiket a gépi tanulás alapú klasszifikációra használtam fel. A három mérték az *Intrinsic Connectivity (IC)*, *Local Correlation (LCOR)* és *Fractional Amplitude of Low-Frequency Fluctuations (fALLF)*.

#### *Intrinsic Connectivity (IC)*

Az IC megmutatja, hogy egy voxel konnektivitása, milyen erős az összes többi voxelhez képest. (egy voxel hány másik voxelrel van összeköttetésben egy bizonyos küszöb érték felett) [31]:

$$IC(x) = \sqrt{\int_{y \in M} r^2(x, y) dy} \quad (2.1)$$

Ahol  $r$  a voxelek közötti korrelációs térkép,  $M$  az előre meghatározott maszk, és  $IC$  az *Intrinsic Connectivity* térkép [31]

#### *Local Correlation (LCOR)*

Ez a mérőszám megmutatja minden egyes voxel lokális koherenciáját. Tehát egy voxel konnektivitását a szomszédos területeken lévő voxelekével ahol a szomszédság mértékét egy Gauss súlyfüggvény adja majd meg [31]:

$$LCOR(x) = \frac{\int w(x - y)r(x, y)dy}{\int w(x - y)dy} \quad (2.2)$$
$$w(z) = e^{-\frac{|z|^2}{2\sigma^2}}$$

Ahol  $r$  a voxelek közötti korrelációs térkép,  $w$  az izotróp Gauss súlyfüggvény, melynek a mérete  $\sigma$  és ez határozza meg a szomszédság méretét.  $LCOR$  pedig a *Local Correlation* térkép [31].

#### *Fractional Amplitude of Low-Frequency Fluctuations (fALLF)*

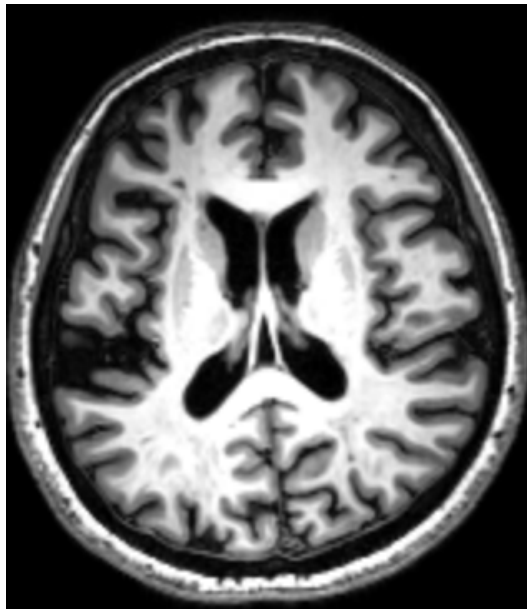
A fALLF térkép egy relatív mértéket ad a BOLD jel erejére. Egy arányszám az általunk vizsgált sávszűrő sávban (0.01-0.1 Hz) és az egész frekvenciaspektrumban vizsgált BOLD jel között [32]:

$$f_{ALLF}(x) = \sqrt{\frac{\sum_t (h(t) * S(x, t))^2}{\sum_t S(x, t)^2}} \quad (2.3)$$

Ahol  $S$  az eredeti BOLD idősor sávszűrés előtt,  $h$  a sávszűrő, és  $f_{ALLF}$  a *fractional Amplitude of Low-Frequency Fluctuations* térkép [32].

## 2.4. Strukturális MRI és elemzés

A strukturális MR felvételek segítségével megítélhető az agy anatómiája és pathológiája. AK esetében jól megítélhető az atrophia az MCI esetében pedig egyszerűen alkalmazhatóak rajta a kvantitatív mérések [33] [34]. A 2.7 ábrán egy konvencionális T1 súlyozott felvétel látható a koponyáról.



2.7. ábra. T1 súlyozott axiális irányú MR felvétel a koponyáról

Ezen felvételek elemzésére az egyik leggyakrabban használt szoftver, melyet a diplomamunkám során is használtam a Freesurfer.

### 2.4.1. Freesurfer

A program az agyat fehér-, és szürkeállományra, illetve a lágyagyhártyát alkotó felületre szegmentálja majd egy atlasz segítségével felosztja. Ezeken több mérést is el tudunk végezni pl.: adott terület vastagságának megállapítása, felület, vagy akár görbület mérése. A szoftver egyik nagy hátránya, hogy csak UNIX rendszereken érhető el és a működtetését is terminálon keresztül parancssorokkal tudjuk irányítani tehát alapszintű programozói tudás is szükséges hozzá. Az egyik leggyakrabban használt parancssor a *recon-all*, mely minden lépést elvégez, felhasználói beavatkozás nélkül.

## 2.5. Gépi tanulás

A gépi tanulás egyre gyakrabban jelenik meg a szakirodalomban annak érdekében, hogy betegséget klasszifikáljanak vagy biomarkereket mutassanak ki különböző feature selection algoritmusokkal [35] [36] [37]. A diplomamunka egyik célkitűzése a megfelelő biomarker



identifikáció gépi tanulás alapú módszerek segítségével. Ehhez elsődlegesen a fMRI elemzések során készült IC, LCOR és FALLF metrikákat használtam fel, illetve ezen térképek átlagolt értékeit a már említett 132 ROI-ra. A bemenő változók dimenzióinak csökkentése érdekében egy Sequential Floating Forward Selection (SFFS) nevű algoritmust használtam fel, ami meglévő 132 ROI-ból egy kisebb halmazt választ ki. Ezzel a módszerrel kiszűrhetőek azok a változók melyek irrelevánsak vagy redundánsak. Az algoritmust két gépi tanulás alapú modellel futtattam, illetve validáltam az eredményeket, melyek közül az egyik a *Random Forest* (RF), a másik pedig a *Support Vector Machine* (SVM).

### 2.5.1. feature selection

Célja csökkenteni a bemenő változók számát, ezzel csökkentve a modell számításigényességét és javítva a teljesítményét. A feature selection algoritmusokat két fő csoportba sorolhatjuk: Filter és Wrapper módszerek. A Filter módszerek gyorsabbak és még a tanulás előtt az előfeldolgozás során választják ki a megfelelő változókat. A Wrapper módszerek a modell tanítását a bemenő változók egy részével hajtják végre iteratív módon. Ebben az esetben az algoritmus mindig azokat a változókat választja ki, amikkel a modell a legjobb eredményt éri el. A *Filter* módszereknél, gyakori a túltanulás [38] és nem mindig nyújtanak megfelelő pontosságot a wrapper módszerekhez képest [39].

Tehát a feature selection módszerek célja, hogy egy  $D$  változó halmazból kiválasszon egy  $Y$  halmazt úgy, hogy  $D > Y$  annak érdekében, hogy csökkentsük a bemenő változók dimenzióinak számát és robusztusabb modellt építsünk, aminek jó általánosító képessége van.

$$D = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_D \end{bmatrix} \qquad Y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_y \end{bmatrix} \qquad (2.4)$$

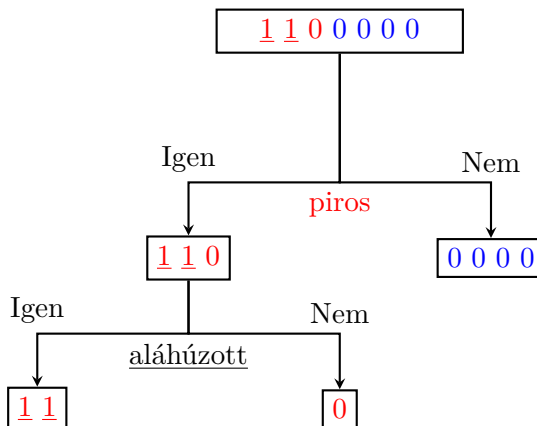
Az általam alkalmazott SFFS algoritmus egy Wrapper módszer, ami a Sequential Forward Selection (SFS) egy tovább gondolt változata. Az SFS algoritmus során a bemenő változóink vektora  $Y$  üres. Először ehhez hozzáadunk egy változót a  $D$  halmazból és az algoritmus végig iterál az összes változón ( $x_1, x_2 \dots x_D$ ). Amelyik változóval a modell a legjobb eredményt éri el azt megtartja. Ezután a modell a választott változó mellé helyez egy újabb változót a  $D$  halmazból ismét végig iterál ezen, majd a legjobb modellt elérő változóval tovább megy. Ez az iteráció addig halad amíg el nem éri az algoritmus az előre megadott változó számot, amit a használat előtt előre definiálnunk kell [40] [41].

## 2.5.2. Random Forest

A Random Forest algoritmus egy robusztus klasszifikációra is használható módszer. Az algoritmus a döntési fákon alapszik, mégpedig oly módon, hogy több döntési fát is épít, innen ered a neve is.

### Döntési fa

A Random Forest megértéséhez szükség van a döntési fák bemutatására is, ami egy olyan modell, amely a bemenő vektoroknál elágazásokat csinál, így próbál következtetni a megfelelő címkére. A 2.8 ábra példáján egyszerű megérteni a működését.

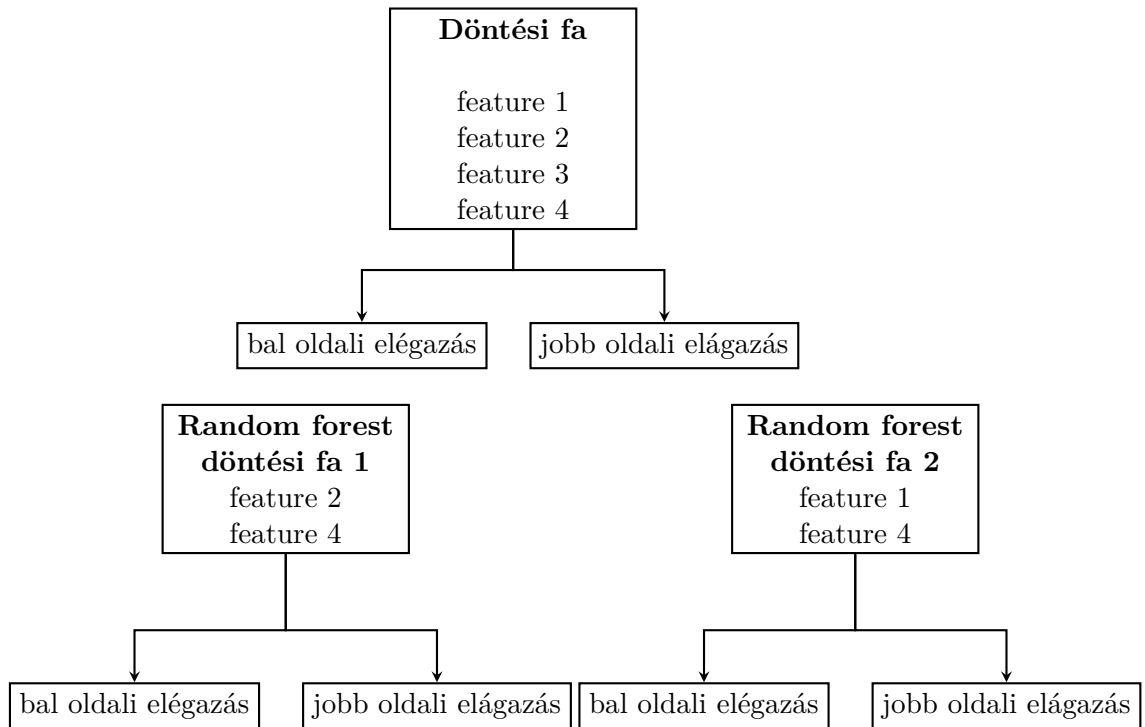


2.8. ábra. Döntési fa ábrája

Ebben az esetben címkék a számok lesznek. Azt akarjuk meghatározni, hogy az adott szám 1 vagy 0 a döntési fa pedig ezen a számok tulajdonságai alapján (szín, aláhúzotttság) hoz létre elágazásokat és hoz döntést. A döntési fa egy jó módszer a klasszifikációra vagy akár regresszióra is használhatjuk. Nagy hátránya azonban a túltanulás. Túltanulásról beszélünk, mikor a modell nem tud jól általánosítani és csak azokra az adatokra ad jó eredményt, amelyeket már látott. Erre próbál megoldást nyújtani a Random Forest azzal, hogy nem csak egy, hanem egyszerre több Döntési fát is használ, így egy robusztusabb döntést hozva, ezzel jobb klasszifikációt elérve.

## Random Forest

A Random Forest-et több modell (több döntési fa) alkotja. Ezeket az eljárásokat Ensemble-nek nevezzük. A modellben minden fának megvan a saját kimenete a végső kimenetet pedig az együttes döntésekből jön létre. A fák között nincs korreláció így egymástól független kimenetet fognak adni. Ezt kétféle módszerrel érjük el: az egyik a bagging (vagy bootstrap aggregáció). Ennek lényege, hogy a fák más-más adatsorokon dolgoznak. A másik módszer pedig az, hogy a bemenő változókat is véletlenszerűen választják ki (2.9 ábra).

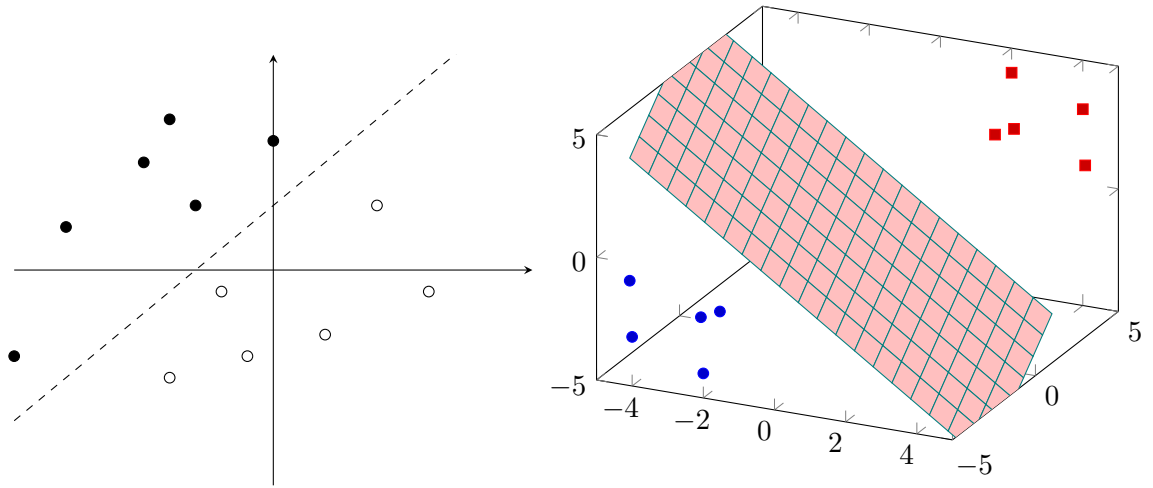


2.9. ábra. Random forest és döntési fa működési elve

Ezzel a módszerrel elérhetjük, hogy a döntési fák más-más adaton és más-más változókon tanuljanak és döntsenek. Ahogy az ábrán is látható nem ugyan azon a változó halmazon dolgoznak így a döntési fák egymástól függetlenek lesznek.

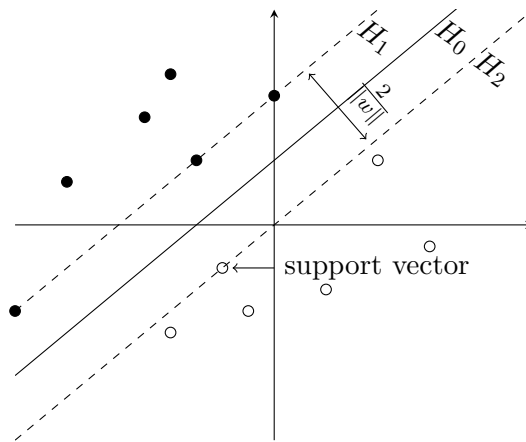
### 2.5.3. Support Vector Machine (SVM)

Az SVM egy széles körben alkalmazott algoritmus, melyet regresszióra és klasszifikációra egyaránt használhatunk. Az algoritmus lényege, hogy egy olyan hipersíkot találjon egy  $N$ -dimenziós térben, amely elkülönítve klasszifikálja az adatpontokat. Ez a hipersík  $R^2$ -ben egy egyenes,  $R^3$ -ban pedig egy sík. Ez látható a 2.10 ábrán.



2.10. ábra. hipersík  $R^2$ -ben (bal) és  $R^3$ -ban (jobb)

Ahhoz, hogy robusztusan el lehessen különíteni az adatpontokat ún. *support vector*-okat alkalmazunk, melyek a hipersíkhöz közeli adatpontok és ezek befolyásolják, annak helyzetét és orientációját. A pontokat elválasztó sík mellett 2 további hipersíkot is elkülönítünk. A kettő közötti teret pedig *margin*-nak nevezzük. A cél, hogy az említett *margin*t maximalizáljuk [42].



2.11. ábra. Az SVM működésének ábrája  $R^2$ -ben

Ha két csoportunk van melyet klasszifikálni szeretnénk akkor ezeknek két fajta címkéjük lesz:  $y_i = +1$  vagy  $y_i = -1$ . A support vector-oknak a lehető legközelebb kell esniük a hipersíkokhoz ezért a két hipersíkot a következőképp lehet felírni:

$$\begin{aligned} w^T * x + b &\geq 1, \text{ mikor } y_i = +1 \\ w^T * x + b &\leq -1, \text{ mikor } y_i = -1 \end{aligned} \quad (2.5)$$

Ahol  $w$  egy súlyvektor, ami a sík orientációját adja,  $x$  a bemenő változó vektora és  $b$  a bias ami a sík eltolását adja.

$H_1$  és  $H_2$  pedig a síkok amik egyenlete:

$$\begin{aligned} H_1 : w^T * x + b &= 1 \\ H_2 : w^T * x + b &= -1, \end{aligned} \quad (2.6)$$

A support vectorok a sík egyenletére eső vektorok és ezek helyzete határozza meg a síkok egyenletében lévő  $w$  súlyokat és a  $+b$  bias-t is. Ebből következik, hogy a két felületet elválasztó sík a  $H_1$  és  $H_2$  mediánja lesz  $H_0$  aminek az egyenlete:

$$w^T * x + b = 0 \quad (2.7)$$

Az algoritmus célja, hogy a  $H_1$  és  $H_2$  távolságát maximalizáljuk. A távolság felírható a következő egyenlettel:

$$2 * \frac{|w * x + b|}{\|w\|} = \frac{2}{\|w\|} \quad (2.8)$$

Annak érdekében, hogy maximalizáljuk a két sík közötti távolságot  $\|w\|$ -t kell minimalizálnunk úgy, hogy  $H_1$  és  $H_2$  közé nem esik pont tehát:

$$y_i(x_i * w) \geq 1 \begin{cases} w^T * x + b \geq 1 \\ w^T * x + b \leq -1 \end{cases} \quad (2.9)$$

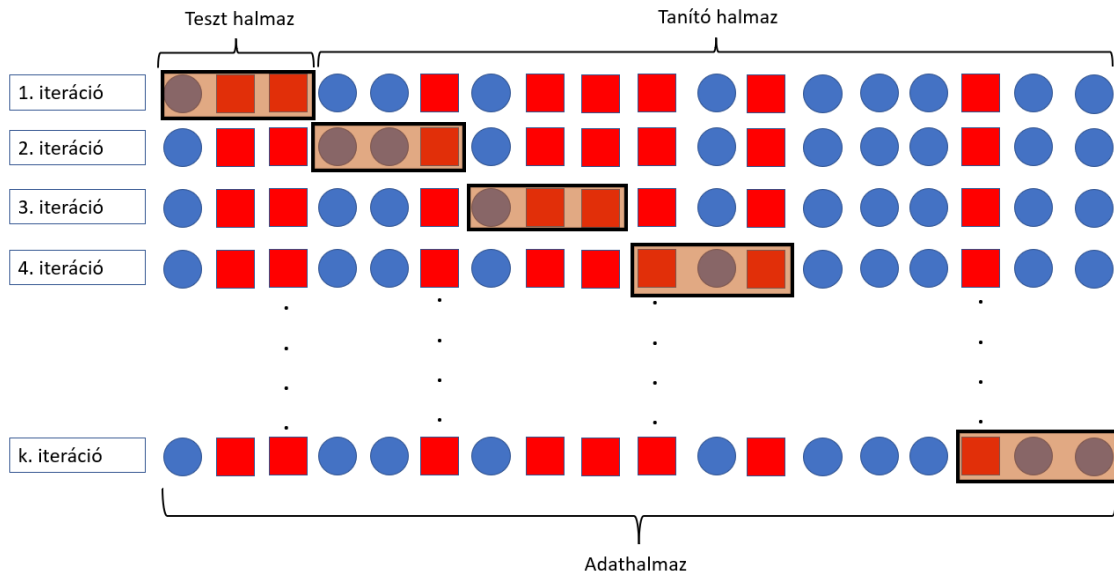
Ez az SVM legegyszerűbb esete, amikor a pontok egymástól lineárisan elválaszthatók. A legtöbb probléma azonban nem ilyen egyszerű, ilyenkor megengedhetünk egy bizonyos szintű hibát is, ahol a pontok a két síkon belülre is eshetnek. Egyéb kernel függvényeket is alkalmazhatunk, hogy a teret szeparábilis problémává transzformáljuk. Illetve az elválasztásra is több függvény is használható az egyenes egyenletén kívül [42].

### 2.5.4. Kereszt Validáció (KV)

Mikor a feature selection algoritmusokkal megkapjuk a kiválasztott változókat fontos ezek validálása. Erre egy kiváló módszer a kereszt validáció (KV). Minden gépi tanulás alapú modellnél azt szeretnénk elérni, hogy jó általánosító képességű legyen tehát ne csak azokon az adatokon teljesítsen jól, amiken tanult, hanem az új, még nem látott adatokon is jó eredményeket érjen el. Ezért is fontos a modell validációja. Ahhoz, hogy meghatározzuk a modell teljesítményét olyan adatokon kell tesztelnünk, amik nem szerepeltek a tanító halmazban, ebből látható lesz, hogy a modell túl-, vagy alul tanul-e, vagy jól általánosít. Mikor sok adat áll rendelkezésünkre akkor elég egy tanító és egy teszt adathalmazra szétszedni a meglévő adatinkat. Ilyenkor akár egy nagyobb részt is lehetünk a tanító adatok közül a teszt adathalmazba. Azonban, amikor kevés adatunk van, akkor minden további adatponttal javítható a modell pontossága ezért nem előnyös nagyobb részhalmazokra szétszedni az adathalmazt. Ebben az esetben lehet számunkra hasznos a kereszt validáció, mivel ezzel minden adatpont szerepelni fog a tesztelő és tanító adathalmazban is [43]. A KV az alábbi lépéseket követi:

1. Az adathalmazunkat szétválasztjuk több részhalmazra, amelynek a számát mi határozzuk meg (ez adathalmaztól függ, de a leggyakoribban az 5 vagy nagyobb adathalmaznál 10).
2. Tanítjuk a modellt úgy, hogy a választott részhalmaz egy része a teszt adathalmaz lesz, a többi pedig tanító.
3. Megismételjük a tanítást úgy, hogy a következő részhalmaz adja a teszt adathalmazt a többi pedig a tanító halmazt.
4. Addig iterálunk amíg végig nem mentünk a teljes adathalmazon

Ennek menete a 2.12 ábrán látható:



2.12. ábra. Kereszt validáció

### 2.5.5. Hiperparaméter optimalizáció

A gépi tanulás alapú modelleknek két paraméterük van

- *Hiperparaméter*: Olyan paraméterek, melyeket a felhasználó állít be a tanítás előtt
- *Modell paraméter*: Olyan paraméterek, melyeket a modell megtanul (SVM esetében pl. az egyenes orientációja)

A Hiperparaméterek azt határozzák meg, hogyan épül fel maga a modell. Ennek a beállítása egy optimalizációs feladat. Több hiperparaméter is van és a mi célunk az, hogy megtaláljuk azok kombinációit, amivel a legjobb eredményt érjük el. Erre több lehetőségünk is van. Az egyik legegyszerűbb mikor ezeknek a beállítását kézzel végezzük el algoritmus használata nélkül. Azonban ez a módszer nem mindig elegendő. Gyakran a hiperparaméter tér, amiben keressük a megfelelő megoldást túl nagy ahhoz, hogy kézzel kipróbáljuk az összes lehetőséget. Ilyenkor használhatunk algoritmusokat, amelyek automatikusan egy előre megadott hiperparaméter térből válogatják ki azokat a paramétereket, amikkel a legjobban teljesít majd a modell. Az egyik ilyen algoritmus a *Random search* ami véletlenszerűen válogatja ki a változókat egy előre megadott térből. Az algoritmus gyors viszont hátránya, hogy mivel random keres lesznek olyan paraméterek, melyek lehet, hogy jobb eredményt adtak volna viszont nem kerültek tesztelésre. Erre megoldás lehet a *Grid search* nevű algoritmus, ami az összes megadott paramétert kipróbálja. Ez akkor lehet nagyon hasznos, ha nem akarunk túl bonyolult modellt építeni a túltanulás elkerülése végett [44] [45], de a hiperparaméter tér még mindig túl nagy egy kézzel való beállításhoz.

## 3 Anyag és módszer

### 3.1. Elemzések automatizációja

A diplomamunka során a MATLAB szoftvert és Python programozási nyelvet használtam, hogy létrehozzak egy olyan keretrendszert, amelyből elérhetőek mind a funkcionális mind a strukturális MR felvételek elemzései, ezen elemzések automatikus futtatása, beteg adatainak automatikus felvétele, illetve az elemzési lépések minőség ellenőrzése is.

A keretrendszert egy GUI-ba ágyaztam, melyhez a Tkinter Python könyvtárat [46] használtam fel.

Annak érdekében, hogy elérhetőek legyenek az rs-fMRI felvételek elemzései a *MATLAB Engine API for Python* nevű alkalmazásprogramozási felületet [47] használtam, aminek segítségével elérhetővé váltak a CONN Toolbox függvényei Pythonból.

A strukturális MRI felvételek elemzéséhez pedig a *Windows Subsystem for Linux* (WSL) alrendszer [48] alkalmaztam, ahova telepítettem a Freesurfert majd *bash script*-ekkel elértem azt. A *subprocess* Python könyvtár [49] pedig lehetővé teszi bash scriptek futtatását így a GUI-ból használható lesz a Freesurfer program is.

Az általam készített programmal a kutatás során kapott MR vizsgálatokat elemeztem annak érdekében, hogy majd fel tudjam használni ezen adatokat a gépi tanuláshoz.

### 3.2. Alanyok

A diplomamunka során 32 MCI-ban szenvedő és 46 egészséges kontroll alanyt (összesen 78 páciens) elemeztem. Az adatokat két fő kutatóközpont biztosította: az Országos Mentális, Ideggyógyászati és Idegsebészeti Intézet (OMII), és a Semmelweis Egyetem Neurológiai Klinikája.

### 3.3. MR vizsgálat

Minden páciens átesett egy koponya MRI felvételen, amelyben mind a strukturális mind a funkcionális felvételeket tartalmazta. A két különböző kutatóközpontban két különböző szkenneren készültek a felvételek. A felvételek az Országos Mentális, Ideggyógyászati és Idegsebészeti Intézetben egy Siemens Magnetom Verio 3T-s készüléken készültek a felvételek egy standard 12 csatornás fejtekerccsel. A protokollban szerepelt egy T1 súlyozott 3D MPRAGE (*magnetization prepared rapid gradient echo*) szekvencia (TR = 2300 ms; TE = 3.4 ms; TI = 100 ms; flip angle: 12°; voxel méret: 1.0 x 1.0 x 1.0 mm) a strukturális felvételek miatt. Az rs-fMRI felvételekhez pedig egy EPI alapú MRI szekvenciát került alkalmazásra (TR = 2,000 ms; TE = 30 ms; flip angle = 79°; voxel méret = 3 x 3 x 3 mm). A pácienseket megkértük, hogy csukott szemmel feküdjenek nyugodtan az fMRI felvételek során.

A Semmelweis Egyetem Neurológiai Klinikáján pedig egy Philips Ingenia 3T-s MRI berendezésen készültek a felvételek egy 8 csatornás SENSE fejtekerccsel. A strukturális MRI felvételek egy T1 súlyozott 3D Turbo Field Echo (TFE) szekvenciával készültek (TR = 9,7 ms; TE = 4.6 ms; flip angle: 8°; voxel méret: 1.0 x 1.0 x 1.0 mm). Az rs-fMRI felvételek pedig egy T2\* súlyozott EPI szekvenciával (TR = 2,0 ms; TE = 30 ms; flip



angle: 70°; voxel méret: 3 x 3 x 4 mm). A pácinseket megkértük, hogy a képernyőn lévő keresztire fókuszáljanak az fMRI felvételek során.

## 3.4. Felvételek elemzése

### 3.4.1. Strukturális elemzés

A T1 súlyozott 3D-s anatómiai felvételeket a Freesurfer 6.0 szoftverrel elemeztem [50]. A programon belül a *recon all* parancsot alkalmaztam minden páciensen. Az elemzés kimeneteként pedig az adott régiók átlagos vastagságát és térfogatát ezek szórását kapjuk meg. Ezek az információk 2 fontosabb fájlba kerülnek, amik az *lh.aparc.stats*, illetve az *rh.aparc.stats*. A két fájl két agyi félteke kortikális parcelláció eredményeit tartalmazza. Ezekből szakirodalmak szerint [51] [52] kiválogattam két olyan régiót amik a legjobban reprezentatívak a betegségcsoportra. A Parahippocampust és a Entorhinális kérget, illetve ezek vastagságát és térfogatát vizsgáltam. A két csoport közötti különbség megítélésére két mintás T próbát alkalmaztam amit a Pythononban elérhető *Scipy* [53] nevű könyvtáron keresztül implementáltam.

### 3.4.2. rs-fMRI felvételek elemzése

Az rs-fMRI felvételek elemzéséhez a CONN MATLAB toolboxot használtam ami egy nyílt forráskódú MATLAB és SPM [54] alapú szoftver (az SPM-ben lévő fMRI elemzésre alkalmas függvények implementációjával). A szoftver számos kiértékelési módszert nyújt nyugalmi állapotú fMRI vizsgálatok elemzéséhez [25]. Az elemzéshez a *standard preprocessing pipeline*-t alkalmaztam; *functional realignment and unwrap*, *slice-timing correction* (*interleaved* a Siemens készülék esetében és *ascending* a Philips szkennernél), *outlier identification*, *direct segmentation and normalization* és *smoothing* 8mm Gauss kernellel. Az előfeldolgozás után minőségellenőrzést hajtottam végre. A zajmentesítés lépésben egy 0,008-0,09 Hz közötti sávszűrőt alkalmaztam illetve lineáris regressziót a potenciális confounderek (fehérállományból, Liquirból származó jelek) szűrésére. Végezetül pedig a már említett három voxel szintű metrikát futattam le: *Intrinsic Connectivity*, *Local Correlation* és *fractional Amplitude of Low Functional Fluctuations*.

### 3.5. Gépi tanulás

A gépi tanulás alapú algoritmusokat az rs-fMRI adatokon hajtottam végre. A ROI-k a CONN Toolboxban elérhető atlaszon lettek definiálva. *Sequential Floating Forward Selection* nevű algoritmust alkalmaztam, annak érdekében, hogy megtaláljam 132 ROI közül azokat, amelyek a legjobban leírják a különbséget az MCI és a kontroll alanyok között. Az SFFS algoritmus implementálására az *mlxtend* Python könyvtárat [40] használtam fel.

A hiperparaméter optimalizációra a *scikit-learn* Python könyvtárból [55] elérhető *Grid Search* algoritmust alkalmaztam.

Az eredmények értékelésére kereszt validációt hajtottam végre. A KV-t a Random Forest és Support Vector Machine algoritmusokkal végeztem el. A modellek a *scikit-learn* Python könyvtárból [56] [57] kerültek implementálásra. A tanítás és tesztelés végén egy átlag értéket számoltam a KV-ben mért pontosságból. Azt, hogy a modell mennyire volt pontos az átlagos pontossági értékekkel konfúziós mátrixsal (KM) határoztam meg (3.1 ábra).

		<u>Predikált adat</u>	
		Kontroll	MCI
<u>Valós adat</u>	kontroll	TN	FP
	MCI	FN	TP

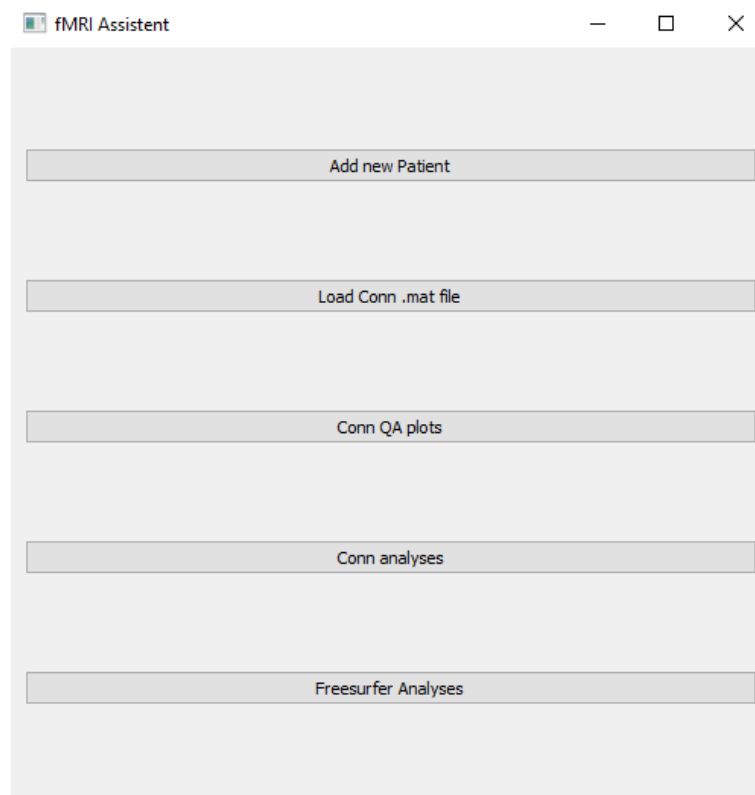
**3.1. táblázat.** konfúziós mátrix, TN: True negative, FP: False Positive, FN: False Negative, TP: True Positive

A táblázatból pedig a pontosság kiszámolása:  $Pontosság = \frac{TP+TN}{TP+TN+FP+FN}$

# 4 Eredmények

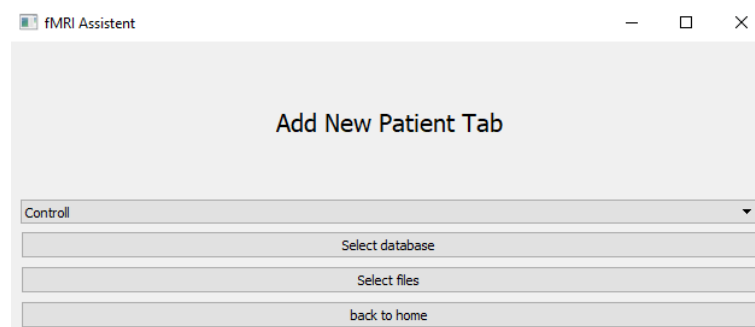
## 4.1. Elemzések automatizálása

A félév során a már előző félévben elkészített Python keretrendszert bővítettem további funkciókkal, illetve egy grafikus felülettel annak érdekében, hogy minél felhasználóbarátabb legyen, illetve, hogy javítsam az elemzések automatizációját. A felhasználói felület az 4.1 ábrán látható.



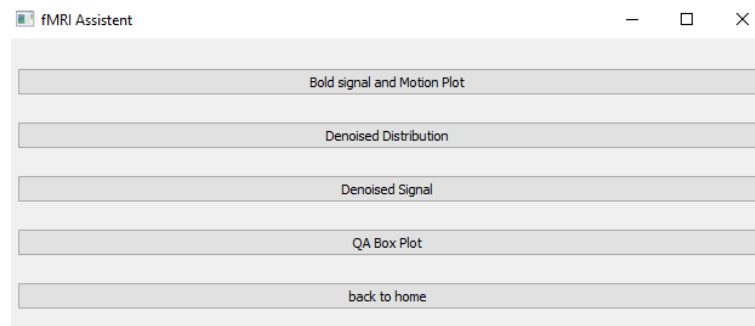
4.1. ábra. felhasználói felület

Az első gombra „Add new patient” kattintva a 4.2 látható felületre kerülünk.



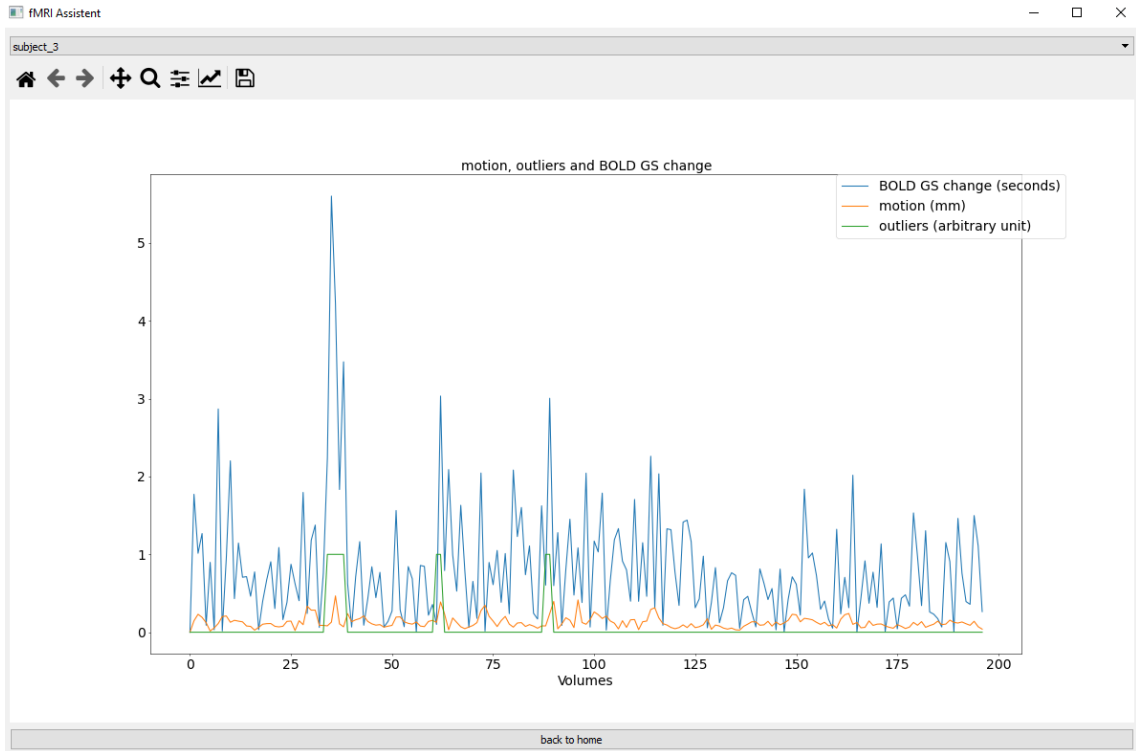
4.2. ábra. Add new patient gomb felülete

A felületen kiválaszthatjuk, hogy a páciens az MCI vagy a kontroll csoportba tartozik-e. Ennek megfelelően 1, ha MCI és 0 számot rendel az adott pácienshez, ha kontroll. A „Select database” nevű gombbal a kiválasztott adatbázisba helyezi majd a fájlokat, melyeket a „select files” nevű gomb segítségével tehetünk meg. Az 4.1 látható második „Load .mat file” gombbal egy funkcionális felvételek elemzése után létrejött .mat kiterjesztésű fájlt olvashatunk be. A CONN toolbox miután végzett az elemzéssel egy ún. „MATLAB struct”-ot hoz létre. Ebben a fájlban tárolódnak az elemzés információ, mint pl. a készült fájlok elérési helyei és a megadott paraméterek. Ennek segítségével a program ebből a .mat kiterjesztésű fájlból kiolvassa a megfelelő információkat és ennek segítségével ún. „Quality Assurance plot”-okat tudunk vele készíteni mellyel az fMRI során szükséges előfeldolgozási lépések minőségét tudjuk ellenőrizni.



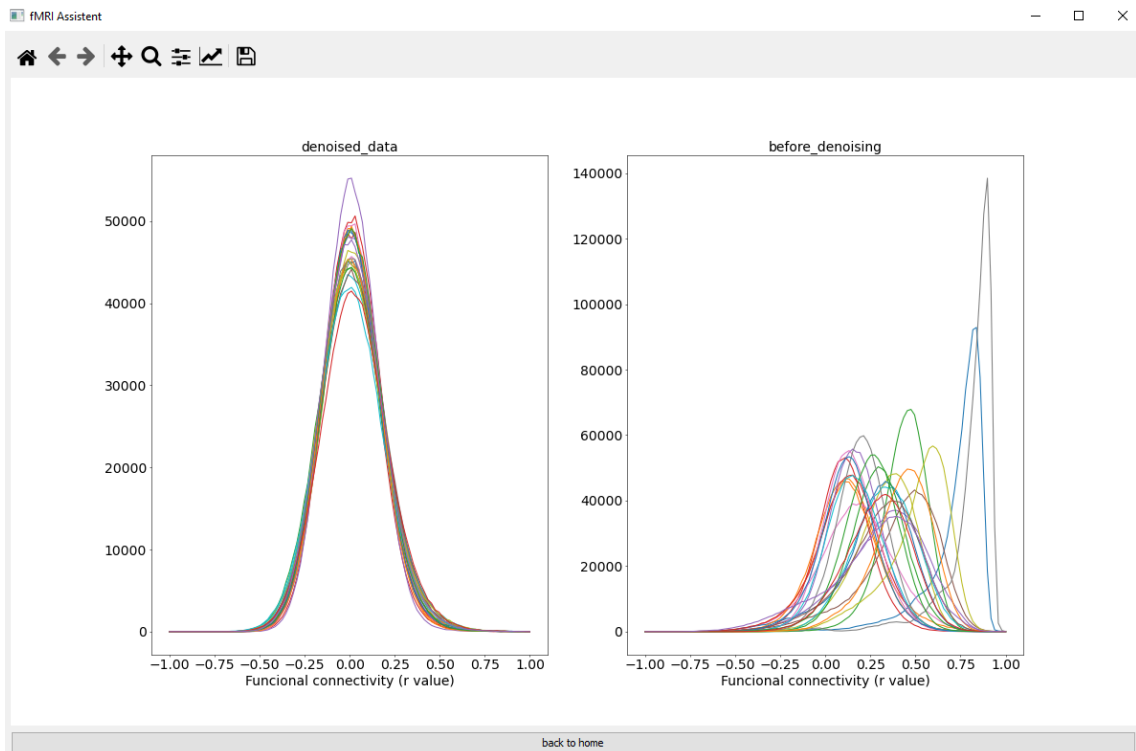
4.3. ábra. QA plot lehetőségek

A felületen az első menüpont a „*Bold signal and Motion Plot*” ábrája látható a 4.4 ábrán. Az ábrán a felső görbe megmutatja az átlagos BOLD idősor változását, az alatta lévő középső görbén a páciens elmozdulását láthatjuk. Ebből a kettőből pedig láthatóak a kiugró értékek, amelyek az alsó görbén figyelhetőek meg. Ezeket a görbéket, ahogy az ábrán is látható minden alanyra meg tudjuk jeleníteni.



4.4. ábra. BOLD jel és mozgás ábrája

A második lehetőség a „Denoised Distribution”. Ebben az esetben a zajmentesítés előtti (jobb ábra) és utáni (bal ábra) BOLD jel eloszlását figyelhetjük meg random voxelek között (4.5 ábra).



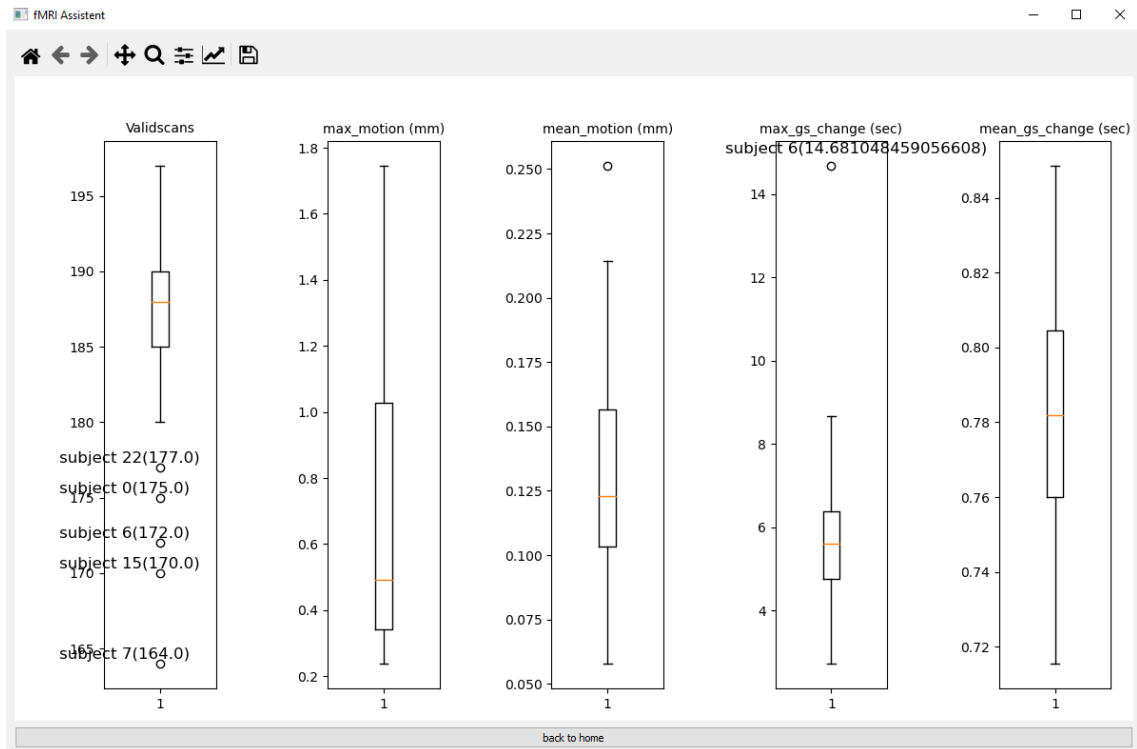
4.5. ábra. Zajmentesítés utáni (jobb) és előtti (bal) korrelációk eloszlása

A harmadik lehetőség a „Denosed signal”. Ebben a menüpontban két régió zajmentesített BOLD jelét hasonlíthatjuk össze Pearson korrelációval. A programban megvizsgálhatjuk a két idősort, illetve azt is meg tudjuk nézni, hogy mennyi a Pearson korreláció  $R$  illetve  $p$  értéke. A menüpont felső részében tudunk kiválasztani két anatómiai területet (ROI-t), majd megvizsgálni ezek kapcsolatát.



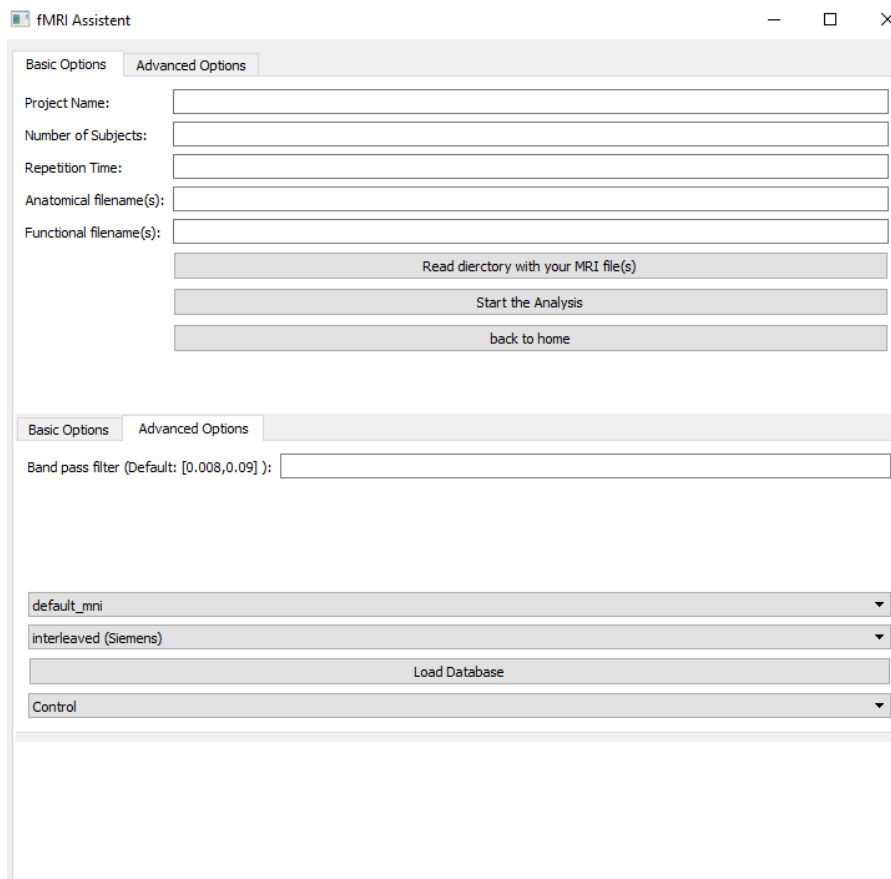
**4.6. ábra.** két ROI közötti korreláció és idősor vizsgálat

Végül pedig az utolsó menüpont a „QA box plot” opció. Ebben az esetben az adott vizsgálatban lévő kiugró értékeket vizsgálhatjuk meg mind a BOLD jel változásában illetve mozgásban és az elkészült felvételekben is. Ezt boxplot formájában jeleníthetjük meg. A program a kiugró értékek mellé kiírja az adott értéket, illetve, hogy ezt melyik alanynál tapasztalta. Ezt a 4.7 ábra mutatja.



**4.7. ábra.** boxplot az adott vizsgálatról (érvényes szkennelések, maximum és átlagos mozgás, BOLD jel változás). A kiugró értékek mellett a program jelzi, hogy melyik páciensnél mérte ezt az értéket illetve a mért értéket is kiírja.

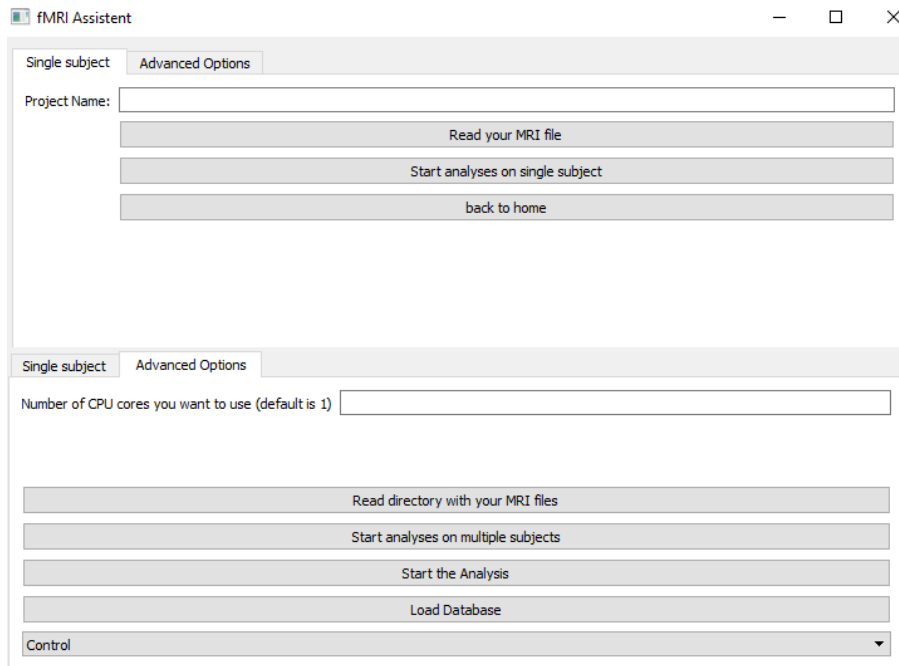
A fent említett funkciók mellett az általam készített programból lehetőség van mind a strukturális mind a funkcionális felvételek elemzésére. Az első ábrán is látható „Conn analysis” gomb segítségével a 4.8 felületére kerülünk, ahonnan az fMRI vizsgálatok elemezhetőek.



4.8. ábra. CONN elemzés felülete

Ezen keresztül tudjuk elindítani a funkcionális felvételek elemzését néhány paraméter megadásával. Az első bemenet melyet a program kér a mappa neve melyben majd a különböző fájlokat fogja helyezni. Ezt követi az alanyszám és a repetíciós idő (az fMRI felvételek időbeli felbontása). A következő két paraméter pedig az anatómiai és a funkcionális fájlok neve lesz. A “Load folder that your files are in” gombbal egy Windows fiedialog felületen keresztül választhatjuk ki azt a mappát melyben az anatómiai és funkcionális felvételek találhatóak. A program automatikusan, olyan nevű fájlokat fog keresni a mappán belül melyet az előbbiekben megadtunk. További elemzési opciók beállítására is van lehetőség az *advanced options*, menüpont alatt mint például a sávszűrő, az előfeldolgozási lépések, illetve a MR szeletek sorrendje. Ezekhez a program egy alapértelmezett paramétert is rendel melyek egy általános elemzéshez teljesen megfelelnek, így, ha a felhasználó nem választ ki semmit, akkor ezeket fogja használni. A menüponton belül lehetőség van kiválasztani, hogy az alany, melyik csoportba tartozik (MCI, Kontroll). Ha ezt kiválasztottuk a elemzés után az adatokat automatikusan felveszi a megadott táblázatba amit a *Load Database* gomb segítségével adhatunk meg. A GUI-n belüli utolsó menüpontban az anatómiai felvételek elemzése válik elérhetővé a programon keresztül.





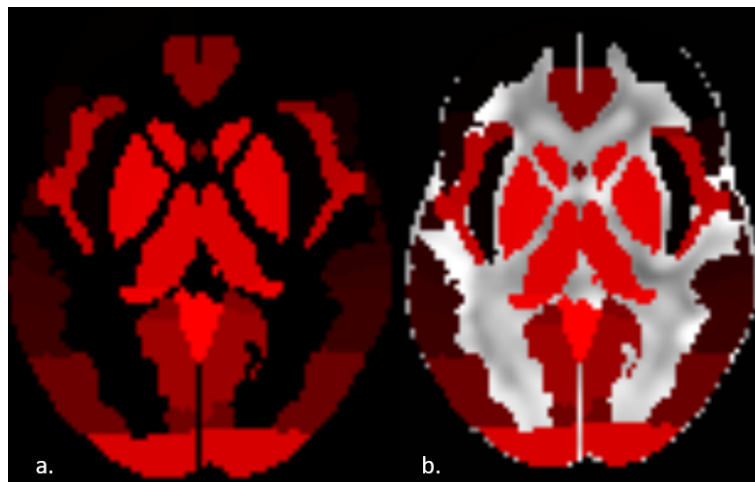
4.9. ábra. Strukturális felvételek elemzésére szolgáló ablak

Ahogy azt a korábbi fejezetben említettem a Freesurfer nevű szoftver csak UNIX alapú operációs rendszerekre érhető el Windowsra nem. Azonban Windowson belül egy elérhető és jól működő lehetőség Linux rendszerek futtatására a Windows Subsystem for Linux (WSL). WSL-en belül a Freesurfer nevű program is telepíthető Windows operációs rendszereken is. A program melyet készítettem ezt használja ki. Ha telepítve van a Freesurfer a WSL-en keresztül akkor egyéb dolgot nem kell tennünk, mint megadni a programnak a kért paramétereket. Ilyen a mappa neve melyben a fájlokat fogja helyezni és beolvasni az anatómiai fájlt melyen az elemzést el kívánjuk végezni. Ha ezután elindítjuk az elemzést, ez automatikusan lefut. További lehetőség az fMRI elemzésekhez hasonlóan, ha kiválasztjuk az adatbázist, illetve azt, hogy az adott páciens mely csoportba tartozik. Ekkor a program a készített fájlok információit automatikusan hozzárendeli ehhez az adatbázishoz. További lehetőség, hogy az alanyokat párhuzamosan futtassuk több CPU szálon is. Mivel egy alany elemzési ideje több óra (7-24 óra) így ez jelentősen meggyorsíthatja az elemzéseket. Ilyenkor nem egy fájlt, hanem egy egész mappát kell adnunk a programnak innen automatikusan kiolvassa a szükséges fájlokat majd lefuttatja az elemzést.

## 4.2. Adattisztítás

Mivel a Freesurfer különböző területek térfogatát és felületét adja meg kimeneteként, így ebben az esetben nem kellett ezeken az adatokon tisztítást végezni már készen álltak az elemzés következő lépésére. Azonban a CONN Toolbox kimenő adatainál szükség volt erre a lépésre is.

Mind három CONN által számolt metrika voxel szintű információt tartalmaz, amiből összesen  $96 \times 96 \times 96$  van. Emiatt annak érdekében, hogy az agyi területekről kapjunk információt ezeknek megfelelően átlagoltam a voxelértékeket mind három mérésnél. Először készítettem egy maszkot, melynek felbontása megegyezik a készült felvételek felbontásával. Ebben a maszkban minden terület intenzitás értéke egy számmal van jelölve 1-132-ig. Így a 132 ROI jól elkülönül egymástól. A 4.10 ábra mutatja a ROI-kat, illetve, ezeket rávetítve egy IC map-ra.



4.10. ábra. a. Maszk b. Maszk az IC map-ra vetítve

Az átlagoláshoz pedig írtam egy MATLAB kódot, amely megnézi a ROI-k koordinátáit és ezeken a helyeken átlagolja az fMRI felvételek voxeleiben lévő értékeket majd ezeket egy MATLAB cell tömbbe helyezi.

MATLAB kódrészlet mely elvégzi az átlagolást az IC map esetében:

```
%Mappa melyben a fájlok találhatóak
locationpre="conn_arithm_Graph"
%páciensek száma a for loophoz
nsubs=78
%maszk beolvasása
Mask=niftiread('mask.nii');
%atlasz beolvasása
ROIs=table2array(readtable("atlas.txt"));
Intrinsic_connect={}
for i=1:nsubs
    Intrinsic_connect{i,3}=table2array(readtable("atlas.txt"));
    if i<10
        location=locationpre+"/results/firstlevel/ICC_01/BETA_Subject00"+i+"_Condition001_Measure001_Component001.nii";
        Intrinsic_connect{i,1}="sub_0"+i;
        Intrinsic_connect{i,2}=niftiread(location);
    else
        location=locationpre+"/results/firstlevel/ICC_01/BETA_Subject0"+i+"_Condition001_Measure001_Component001.nii";
        Intrinsic_connect{i,1}="sub_"+i;
        Intrinsic_connect{i,2}=niftiread(location);
    end
    for k=1:max(Mask, [], 'all')
        Intrinsic_connect{i,3}{k,3}=find(Mask==k);
        Intrinsic_connect{i,3}{k,4}=mean(Intrinsic_connect{i,2}(Intrinsic_connect{i,3}{k,3}));
    end
end
```

A CONN toolboxban található egy .txt fájl, amiben a ROI-k nevei találhatóak. Ezek sorrendje megegyezik az intenzitás értékek sorrendjével. Ezt felhasználva egyszerűen behelyezhetők az átlagértékek egy *MATLAB cell* tömbbe ahonnan exportálhatóak .csv fájlba. Ez a 4.11 ábrán látható.

	1	2	3	4
1	'FP r'	'Frontal Pole Right)	8081x1 double	0.5747
2	'FP l'	'Frontal Pole Left)	6926x1 double	0.6528
3	'IC r'	'Insular Cortex Right)	1344x1 double	0.1158
4	'IC l'	'Insular Cortex Left)	1334x1 double	0.0358
5	'SFG r'	'Superior Frontal Gyrus Right)	2680x1 double	0.6328
6	'SFG l'	'Superior Frontal Gyrus Left)	2831x1 double	0.8094
7	'MidFG r'	'Middle Frontal Gyrus Right)	2740x1 double	0.5766
8	'MidFG l'	'Middle Frontal Gyrus Left)	2927x1 double	0.5885
9	'IFG tri r'	'Inferior Frontal Gyrus, pars triangularis Right)	555x1 double	0.6575
10	'IFG tri l'	'Inferior Frontal Gyrus, pars triangularis Left)	650x1 double	1.0523
11	'IFG oper r'	'Inferior Frontal Gyrus, pars opercularis Right)	686x1 double	0.6509
12	'IFG oper l'	'Inferior Frontal Gyrus, pars opercularis Left)	766x1 double	0.8534
13	'PreCG r'	'Precentral Gyrus Right)	4275x1 double	0.2931
14	'PreCG l'	'Precentral Gyrus Left)	4359x1 double	0.0175

4.11. ábra. MATLAB cell melyben a ROI-k rövidítése (1. oszlop), ROI-k nevei (2. oszlop), ROI-k koordinátái (3. oszlop) és ROI-k átlagértékei (4. oszlop) találhatóak

## 4.3. Gépi tanulás

### 4.3.1. rs-fMRI

A RF és SVM modellek közül a *Grid Search* algoritmus a következő két modellt találta a legjobbnak.

```
#A legjobb random forest modell
Random_forest=RandomForestClassifier(bootstrap=False, max_depth=7, min_samples_leaf=2,
random_state=42)

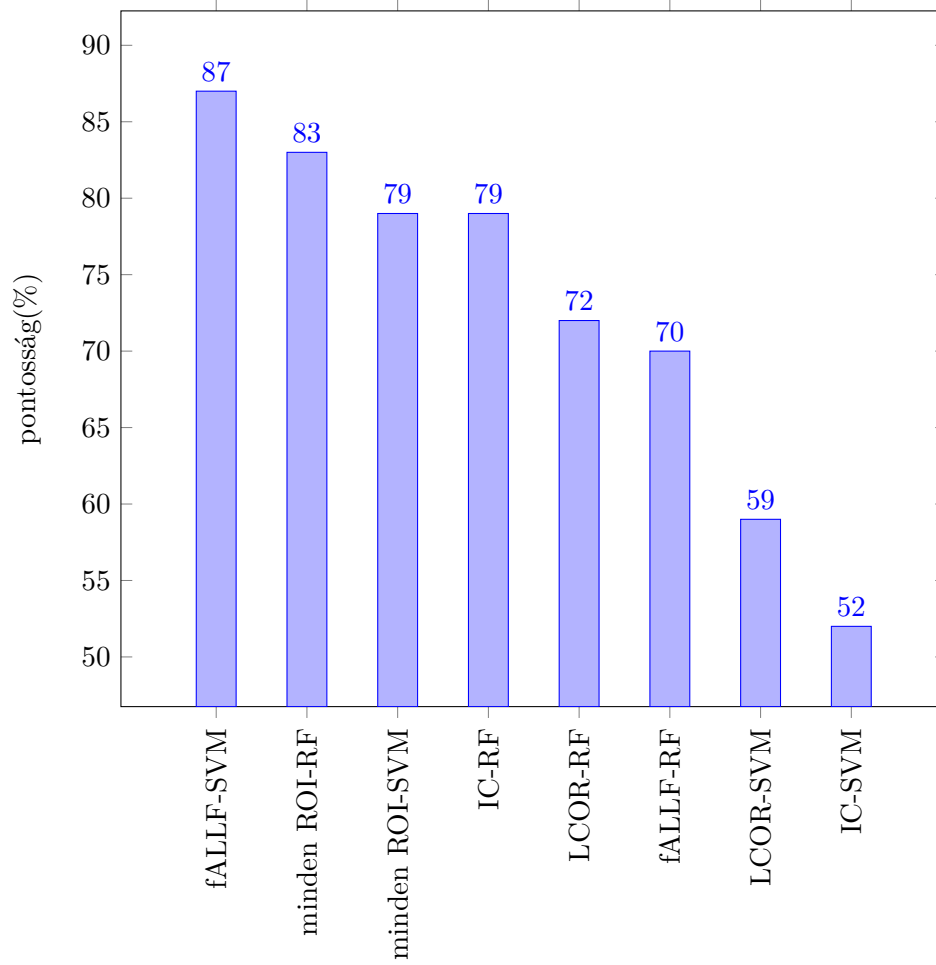
#A legjobb SVM modell
SMV= svm.SVC(C=39, break_ties=True, class_weight=None, degree=1, gamma="auto", kernel="rbf",
shrinking=True, probability=True,random_state=42)
```

A legnagyobb pontosságot (87%) az SVM algoritmussal értem el az fALLF metrikával. A második legjobb eredmény pedig a RF algoritmussal sikerült elérni, ahol 83%-os volt a pontosság, abban az esetben mikor az összes metrikát vizsgáltam egyben. Habár a fALLF érte el a legjobb eredményeket 87%-kal az IC esetében is sikerült 79%-ot elérni a RF algoritmus segítségével. A kapott eredményeket a 4.1 táblázat tartalmazza. Itt megfigyelhető a modellek kereszt validációval mért átlagos pontossága, illetve az ezekből számolt konfúziós mátrix is. A 4.12 ábrán pedig láthatóak a pontosságok oszlopdiagram formájában.

	RandomForest		pontosság(%)	SVM		pontosság (%)
IC	38	8	79	46	0	52
	8	24		30	2	
LCOR	43	12	72	46	0	59
	12	20		32	0	
fALLF	34	12	70	45	1	87
	14	18		9	23	
Egyben a 3 metrika	42	4	83	43	3	79
	9	23		13	19	

**4.1. táblázat.** Gépi tanulás algoritmusok eredményei és ebből számolt konfúziós mátrix

A 4.12 ábrán pedig láthatóak a pontosságok oszlopdiagram formájában.



**4.12. ábra.** oszlopdiagram a modellek pontosságáról

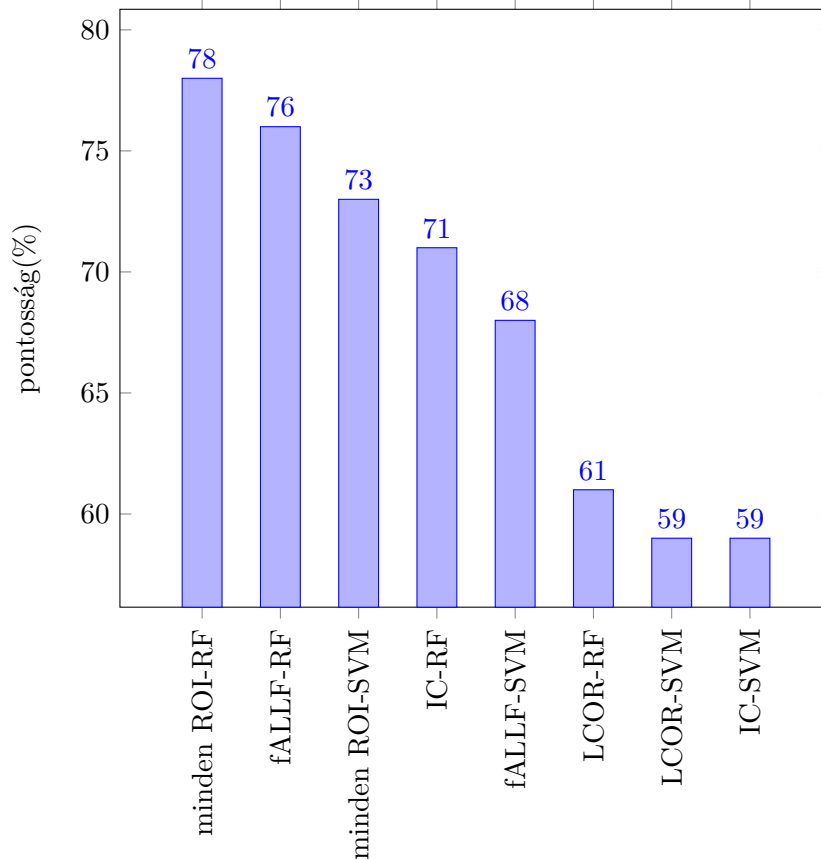
## Visuo-Spatialis rendszer

Mivel az AK gyakran érinti a visuo-spatialis rendszert [58] így azokat a ROI-kat, amik ezt a rendszert írják le külön vizsgáltam. Ebben az esetben a legjobb eredményt akkor sikerült elérnem mikor a 3 metrikát kombináltam. A RF érte el a legjobb eredményt 78%-os pontossággal. Az eredmények a 4.2 táblázatban láthatóak.

	RandomForest		pontosság(%)	SVM		pontosság (%)
IC	34	12	71	46	0	58
	10	22		32	0	
LCOR	34	12	61	46	0	59
	18	14		32	0	
fALLF	39	7	76	38	8	68
	11	21		19	13	
Egyben a 3 metrika	38	8	78	41	5	73
	13	19		16	16	

**4.2. táblázat.** Gépi tanulás algoritmusok eredményei és ebből számolt konfúziós mátrix (Visuo-Spatialis rendszer)

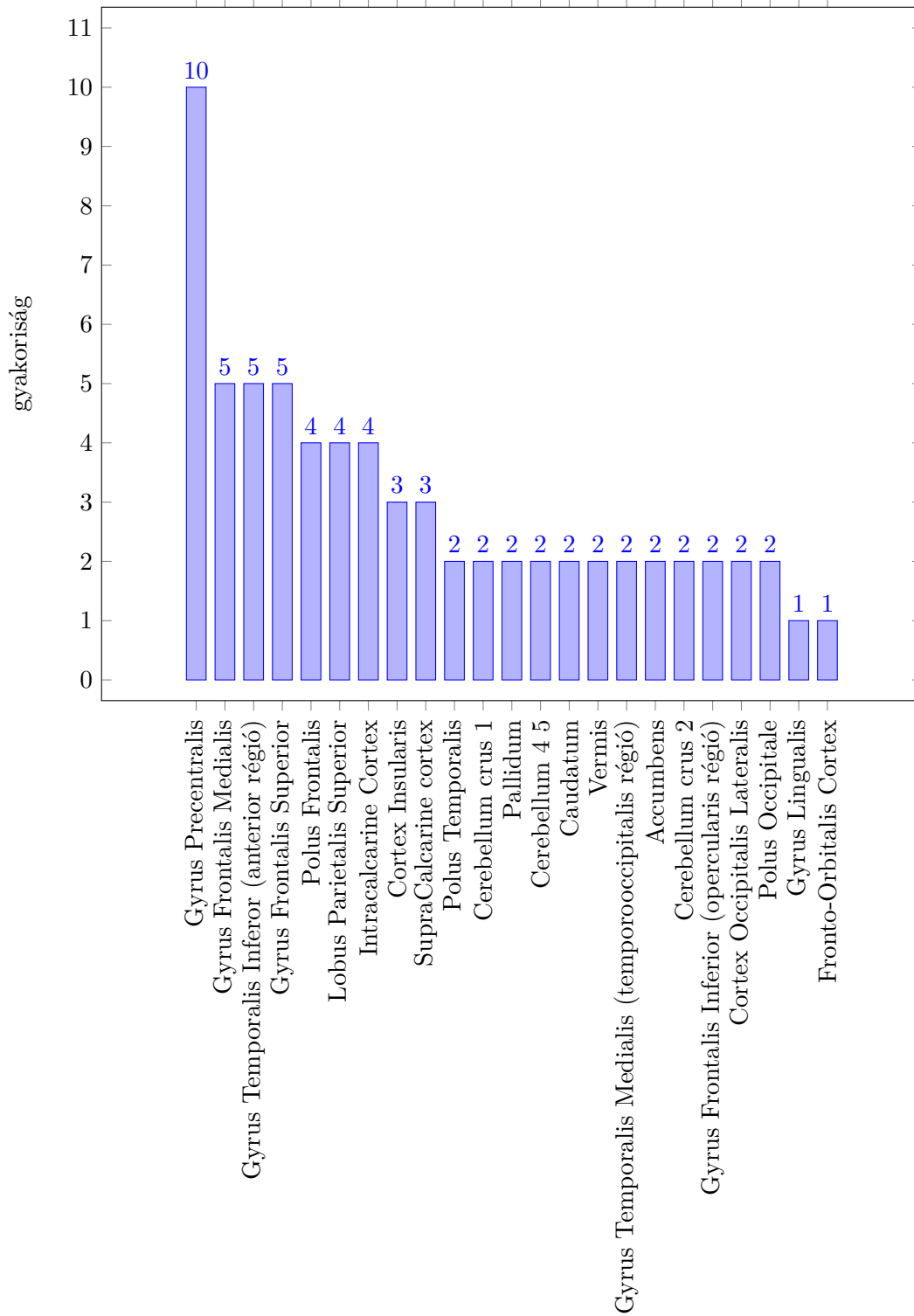
A 4.13 ábrán pedig láthatóak a pontosságok oszlopdiagram formájában.



**4.13. ábra.** oszlopdiagram a modellek pontosságáról (Visuo-Spatialis rendszer)

#### 4.4. Feature selection

A *feature selection* algoritmusok több változót is kiválasztottak, melyek gyakoriságának eloszlásáról látható egy ábra a 4.14 ábrán.



4.14. ábra. oszlopdiaagram az SFFS algoritmus által kiválasztott változókról

A legjobb 3 modell kiválasztott változói pedig a 4.3 táblázatban láthatóak mind a visuo-spatialis és összesített ROI-k esetében.

	<b>Visuo-Spatialis rendszer</b>	
<b>Egyben a 3 metrika (RF)</b>	<b>fALLF (RF)</b>	<b>Egyben a 3 metrika (SVM)</b>
Gyrus Temporalis Medialis (anterior régió)	Gyrus Frontalis Medialis	Gyrus Temporalis Medialis (anterior régió)
Gyrus Precentralis	Gyrus Precentralis	Gyrus Precentralis
Gyrus Frontalis Medialis	Intracalcarine Cortex	Gyrus Frontalis Medialis
Supracalcarine Cortex		Gyrus Frontalis Superior
	<b>Minden ROI</b>	
<b>fALLF (SVM)</b>	<b>minden ROI (RF)</b>	<b>minden ROI (SVM)</b>
Gyrus Temporalis Medialis (temporooccipitális régió)	Gyrus Parahippocampalis	Gyrus Temporalis Medialis (temporooccipitális régió)
Gyrus Temporalis Inferior (temporooccipitális régió)	Accumbens	Gyrus Frontalis Inferior (triangularis és opercularis régió)
Gyrus Precentralis	Supracalcarine Cortex	Gyrus Frontalis Superior
Gyrus Frontalis Superior	Fronto-Orbitális Cortex	Polus Frontalis
Polus Frontalis		Gyrus Precentralis
Caudatum		Gyrus Lingualis
Accumbens		Cortex Occipitalis Lateralis
Polus Occipitalis		
Preconeus		
Cerebellum Crus1		
Cerebellum 4 5		

4.3. táblázat. A legjobb 3 modell által kiválasztott ROI-k



#### 4.4.1. Struktúrális MRI

A két kiválasztott ROI (Parahippocampus, Entorhinális kéreg) átlagos paramétere a 4.15 ábrán láthatóak.



4.15. ábra. oszlopdiagram a régiók átlagos térfogata (felső) és vastagsága (alsó) között (Kontroll vs MCI)

Az ábrákon jól látható, hogy a két csoport Entorhinális kéreg vastagsága és térfogata között nagy különbség van. A két mintás T próba kimutatta, hogy szignifikáns különbség van a két terület között. Szignifikánsan nagyobb az MCI-ban szenvedő páciensek Entorhinális kéreg vastagsága mind két oldalon ( $p < 0.001$ ), illetve a kontroll csoport Entorhinális kéreg térfogata volt szignifikánsan nagyobb mind két oldalon ( $p < 0.001$ ).

## 5 Következtetés

A diplomamunka célja, hogy létrehozzak egy olyan Python felületet, ami leegyszerűsíti és automatizálja az fMRI és strukturális MRI vizsgálatok elemzését, illetve ezek eredményeinek táblázatba felvételét. Továbbá feladataim közé tartozott egy adatbázis létrehozása MCI-ban szenvedő betegekkel és nemben, valamint korban megegyező kontroll alanyokkal. A végső cél pedig az adatbázisban lévő alanyok elemzett felvételeinek a feldolgozása és gépi tanulás alapú klasszifikációja és a betegséget leíró biomarkerek meghatározása.

### Automatizáció

A Python keretrendszert egy GUI-hoz rendeltem. Benne MATLAB-on keresztül elérhetőek a CONN Toolbox függvényei ezzel pedig végrehajtható az fMRI felvételek elemzése. WSL-en keresztül pedig a Freesurfer szoftvert tudjuk használni a strukturális MRI adatok elemzéséhez. Az analízis után az adatok automatikusan bekerülnek egy előre megadott táblázatba így ezek automatikus felvétele is lehetséges. Továbbá az előfeldolgozási-, és a zajmentesítés lépésének a minőség ellenőrzését is végre tudjuk hajtani a GUI-n keresztül. Mind két elemzéshez csak minimális adatokat kell megadnunk amik után az elemzés elindul és automatikusan végigmegy minden szükséges lépésen.

A GUI segítségével az MRI adatok elemzése és az ebből született adatok felvétele nagyban leegyszerűsödik. A program lehetővé teszi egyszerre több alany elemzését is, mind a strukturális, mind a funkcionális felvételek elemzése esetén, amely tovább könnyíti a felhasználó munkáját. A minőségellenőrző ábrákkal pedig az előfeldolgozás és zajmentesítés minőségét is meghatározhatjuk. Ha már vannak meglévő adataink akkor ezeket is könnyen hozzáadhatjuk az adatbázishoz egy külön menüben. A program használata Windows operációs rendszeren is elérhető és nem szükséges hozzá programozói tudás sem.

### rs-fMRI analízis eredményei

A biomarker identifikáció érdekében egy ún. *feature selection* algoritmust használtam, amely képes kiszűrni az irreleváns vagy redundás változókat ezzel egy pontosabb és jobb általánosító képességű modellt eredményezve. Ezt két modellel végeztem el: Random Forest és Support Vector Machine.

A klasszifikáció kritikus kérdése a megfelelő *feature selection* algoritmus kiválasztása. Az összes változó használatával nem kapunk megfelelő eredményeket és a változók összes kombinációinak kipróbálása nem lehetséges mert meglehetősen számításigényes folyamat. Ennek érdekében dolgozták ki a szekvenciálisan kereső technikákat. Ezek mindig a legjobb változót választják ki majd ehhez adnak hozzá folyamatosan egy előre meghatározott számig. Az általam választott SFFS algoritmus pedig nem csak előre halad hanem folyamatosan néz visszafelé így egy robusztusabb részhalmazt biztosítva.

A diplomamunka során sikerült egy 87%-os eredményt elérni a *fALLF* metrikával mikor mind a 132 ROI-t vizsgáltam. A Visuo-Spatiális rendszernél pedig 78%-os eredményt mikor kombináltam a *fALLF*, *ICC* és *LCOR* metrikákat és 76%-os pontosságot a *fALLF* metrikával.

A Visuo-Spatialis rendszernél a kiválasztott ROI-k a következők voltak:

- Gyrus Temporalis Medialis (anterior régió)
- Gyrus Precentralis
- Gyrus Frontalis Medialis
- Supracalcarine Cortex
- Intracalcarine Cortex

Mikor pedig minden ROI-t kombináltam akkor a továbbiak:

- Gyrus Temporalis Medialis (temporooccipitális régió)
- Gyrus Frontalis Inferior (triangularis és opercularis régió)
- Gyrus Frontalis Superior
- Polus Frontalis
- Gyrus Precentralis
- Gyrus Lingualis
- Cortex Occipitalis Lateralis

Az algoritmus szerint ezek azok a ROI-k, amik a legjobban különbséget tesznek az MCI és a kontroll csoport között.

A Gyrus Temporalis medialis mind 3 esetben jelen van mint kiválasztott ROI. Korábbi kutatások arra utalnak, hogy a régióknak több kognitív funkcióban is szerepe van pl. beszéd feldolgozás, szemantikus emlékezés, (anterior és mediális régió) illetve a vizuális felfogás (temporooccipitális régió) [59] [60] [61]. Az említett funkciók hanyatlása pedig jellemzőek az AD betegségekre is.

A Gyrus Frontalis szintén szerepel mind 3 esetben a kiválasztott ROI-k között. A Gyrus Frontalis Superior főleg az öntudatosságban van szerepe koordinációban az érzékelő rendszerrel [62]. A Gyrus Frontalis Medialis trianguláris és opercularis régió a Broca területnek felel meg aminek főleg a nyelvi produkcióban és megértésben játszik szerepet [63]. A Gyrus Frontalis Medialis pedig a műveltségben és a számolásban (számolások megértésében) van szerepe [64]. Az AK a kognitív képességek romlásával járó betegség így ezek a területek is leírhatóak lehetnek a betegségekre.

A másik anatómiai területet, amit mind három esetben kiválasztásra került a Gyrus Precentralis. A terület a primer motoros cortex tehát a mozgásért felelős [65]. A betegség ezeket a területeket is érintheti.

A Visuo-Spatialis rendszerben az Intracalcarine és Supracalcarine kéreg szerepel ami abban az esetben nem mikor mind a 132 ROI szerepel. A Calcarine kéreg a vizuális rendszer részét képezi. Szerepe van az orientáció és térbeli frekvenciák meghatározásában [66] [67]. Ezen funkciók elvesztése és romlása szintén részét képezi a betegségnek.

A Polus Frontalis az egyik leggyakrabban kiválasztott ROI ami nem szerepel a Visuo-Spatialis rendszerben. Fontos szerepet játszik az emberi viselkedésben és a kognitív képességekben (pl. cselekvés megválasztás) [68]. Ezen képességek szintén romlanak az AK során.

A Gyrus Lingualis ROI-t az algoritmus mindössze ebben az egy esetben választotta ki. A gyrus fő feladata a vizuális információ feldolgozásában és főleg a szavak betűk feldolgozásában [69]. Ez szintén lehet a betegség romlása miatt és voltak már korábbi kutatások is amik kimutatták a csökkent funkcionális konnektivitást AK-ban szenvedő páciensek esetében [70].

A Cortex Occipitalis Lateralis ROI-t szintén csak ebben az egy esetben választotta ki az algoritmus. A kéreg fő feladata a tárgy és arc felismerésben van [71]. Az emlékezés csökkenésével az arcok és tárgyak felismerése is nehezebbé válik a betegség során. Koárbbi kutatások pedig mutattak ki csökkent térfogatot a cortexben [72].

Mind a Visuo-Spatiólis rendszernél mind pedig abban az esetben mikor az összes ROI-t vizsgáltam akkor értem el a legjobb eredményeket mikor a három metrikát egyben vizsgáltam illetve fALLF metrikával. Ez arra utal, hogy a fALLF a két csoport között informatívabb lehet és pontosabb klasszifikációt ad azonban a másik két metrikában is vannak olyan információk amik javíthatnak a pontosságon. A fALLF azért adhat pontosabb eredményt, mivel ez lényegében a BOLD jel erősségét adja vissza az IC és az LCOR-al szemben amik inkább a BOLD konnektivitás dinamikáját (szomszédos és az összes voxelrel) [35].

A diplomamunka egyik limitációja a kevés alanzszám (78) volt, ami egy alapvető probléma az legtöbb fMRI kutatás során, hiszen magas a vizsgálatok költsége és az alanyok bevonása [73]. A gépi tanulás alapú algoritmusok jobban teljesítenek nagyobb adathalmazon, mert minél több alany áll rendelkezésre annál jobban megtalálja a két csoport közötti különbséget és annál jobb lesz az általánosító képességük. Egy másik limitáció a csoport heterogenitása. Az MCI és az AK sem egyformán érintik az egyének agyát és más-más területek lehetnek defektívek.

A diplomamunka során meghatároztam, hogy az rs-fMRI által nyújtott változók SFFS algoritmus segítségével megbízhatóan jelzik-e, hogy az adott alany szenved-e MCI megbetegedésben. Ehhez RF és SVM modelleket alkalmaztam és KV-val vizsgáltam meg a modellek pontosságát. A kevés alanzszám és a csoport heterogenitása ellenére sikerült magas pontosságot elérni a választott metrikákkal. A változók, amiket az SFFS algoritmus választott pedig azokat a régiókat érintik amik főleg a kognitív képességekért felelősek, így ezek potenciális biomarkerek lehetnek, amiket nagyobb alanzszámmal tovább lehet vizsgálni, hogy a diplomamunka során született eredményeket validálják. Továbbá az általam választott módszert egyéb betegségek illetve csoportok között is lehet alkalmazni.

## Strukturális MRI analízis eredményei

A biomarker identifikációt nem csak az rs-fMRI eredményeken vizsgáltam meg, hanem a strukturális MRI adatokon is. Ebben az esetben szakirodalmak alapján válaszottam ki két régiót amik térfogatának csökkenése leírhatja betegséget. Ez a két régió a:

- Parahippocampus és a
- Entorhinális kéreg

Mind két agyi terület az emlékezésért ezek kódolásáért és dekódolásáért felelősek [74] [75] és ezért van fontos szerepük az AD és MCI során. A két régió átlagos vastagságát és térfogatát néztem meg a két csoport esetében és két mintás T próbát alkalmaztam annak érdekében, hogy megtudjam van-e szignifikáns különbség a két csoport között. Ebből kiderült, hogy az Entorhinális kéreg szignifikánsabb kisebb az MCI-ban szenvedő betegek esetében. Ez szintén potenciális biomarker lehet későbbi vizsgálatok és kutatások során, hiszen már ebben a korai stádiumban is szignifikáns eltérés van a két csoport között.

# Irodalomjegyzék

- [1] Rudy J Castellani, Raj K Rolston, and Mark A Smith. Alzheimer disease. *Disease-a-month: DM*, 56(9):484, 2010.
- [2] Mild cognitive impairment - Symptoms and causes, . URL <https://www.mayoclinic.org/diseases-conditions/mild-cognitive-impairment/symptoms-causes/syc-20354578>.
- [3] Structural Magnetic Resonance Imaging - an overview | ScienceDirect Topics, . URL <https://www.sciencedirect.com/topics/medicine-and-dentistry/structural-magnetic-resonance-imaging>.
- [4] H. Lv, Z. Wang, E. Tong, L.M. Williams, G. Zaharchuk, M. Zeineh, A.N. Goldstein-Piekarski, T.M. Ball, C. Liao, and M. Wintermark. Resting-State Functional MRI: Everything That Nonexperts Have Always Wanted to Know. *AJNR: American Journal of Neuroradiology*, 39(8):1390–1399, August 2018. ISSN 0195-6108. DOI: 10.3174/ajnr.A5527. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6051935/>.
- [5] Nathassia K. Aurich, José O. Alves Filho, Ana M. Marques da Silva, and Alexandre R. Franco. Evaluating the reliability of different preprocessing steps to estimate graph theoretical measures in resting state fMRI data. *Frontiers in Neuroscience*, 9, 2015. ISSN 1662-453X. URL <https://www.frontiersin.org/articles/10.3389/fnins.2015.00048>.
- [6] William R. Shirer, Heidi Jiang, Collin M. Price, Bernard Ng, and Michael D. Greicius. Optimization of rs-fMRI Pre-processing for Enhanced Signal-Noise Separation, Test-Retest Reliability, and Group Discrimination. *NeuroImage*, 117:67–79, August 2015. ISSN 1053-8119. DOI: 10.1016/j.neuroimage.2015.05.015. URL <https://www.sciencedirect.com/science/article/pii/S105381191500395X>.
- [7] Martin J McKeown, Lars Kai Hansen, and Terrence J Sejnowski. Independent component analysis of functional MRI: what is signal and what is noise? *Current opinion in neurobiology*, 13(5):620–629, October 2003. ISSN 0959-4388. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2925426/>.
- [8] Meenakshi Khosla, Keith Jamison, Gia H Ngo, Amy Kuceyeski, and Mert R Sabuncu. Machine learning in resting-state fmri analysis. *Magnetic resonance imaging*, 64:101–121, 2019.
- [9] Saruar Alam, Goo-Rak Kwon, and Alzheimer’s Disease Neuroimaging Initiative. Alzheimer disease classification using kpca, lda, and multi-kernel learning svm. *International Journal of Imaging Systems and Technology*, 27(2):133–143, 2017.
- [10] Siqi Liu, Sidong Liu, Weidong Cai, Sonia Pujol, Ron Kikinis, and Dagan Feng. Early diagnosis of alzheimer’s disease with deep learning. In *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, pages 1015–1018. IEEE, 2014.
- [11] Jaime Gómez-Ramírez, Marina Ávila-Villanueva, and Miguel Ángel Fernández-Blázquez. Selecting the most important self-assessed features for predicting conversion

- to mild cognitive impairment with random forest and permutation-based methods. *Scientific Reports*, 10(1):1–15, 2020.
- [12] Gloria Castellazzi, Maria Giovanna Cuzzoni, Matteo Cotta Ramusino, Daniele Martinelli, Federica Denaro, Antonio Ricciardi, Paolo Vitali, Nicoletta Anzalone, Sara Bernini, Fulvia Palesi, et al. A machine learning approach for the differential diagnosis of alzheimer and vascular dementia fed by mri selected features. *Frontiers in neuroinformatics*, 14:25, 2020.
- [13] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer’s disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.
- [14] Jeffrey L Cummings and Greg Cole. Alzheimer disease. *Jama*, 287(18):2335–2338, 2002.
- [15] Marco Pagani, Flavio Nobili, Silvia Morbelli, Dario Arnaldi, Alessandro Giuliani, Johanna Öberg, Nicola Girtler, Andrea Brugnolo, Agnese Picco, Matteo Bauckneht, Roberta Piva, Andrea Chincarini, Gianmario Sambuceti, Cathrine Jonsson, and Fabrizio De Carli. Early identification of MCI converting to AD: a FDG PET study. *European Journal of Nuclear Medicine and Molecular Imaging*, 44(12):2042–2052, November 2017. ISSN 1619-7089. DOI: 10.1007/s00259-017-3761-x.
- [16] Serge Gauthier, Barry Reisberg, Michael Zaudig, Ronald C Petersen, Karen Ritchie, Karl Broich, Sylvie Belleville, Henry Brodaty, David Bennett, Howard Chertkow, et al. Mild cognitive impairment. *The lancet*, 367(9518):1262–1270, 2006.
- [17] Ronald C Petersen. Mild cognitive impairment. *CONTINUUM: Lifelong Learning in Neurology*, 22(2 Dementia):404, 2016.
- [18] Frank Gaillard. Alzheimer disease | Radiology Reference Article | Radiopaedia.org. URL <https://radiopaedia.org/articles/alzheimer-disease-1>.
- [19] András Horváth, Dalida Berente, Janos Zsuffa, Tom Werber, Máté Kiss, Anita Drotos, Anita Kamondi, and Gabor Csukly. Alteration of visuo-spatial system as an early marker of cognitive decline: a double-center neuroimaging study. *Frontiers in Aging Neuroscience*, page 503, 2022.
- [20] Clifford R Jack Jr, Val J Lowe, Stephen D Weigand, Heather J Wiste, Matthew L Senjem, David S Knopman, Maria M Shiung, Jeffrey L Gunter, Bradley F Boeve, Bradley J Kemp, et al. Serial pib and mri in normal, mild cognitive impairment and alzheimer’s disease: implications for sequence of pathological events in alzheimer’s disease. *Brain*, 132(5):1355–1365, 2009.
- [21] Puneet Talwar, Suman Kushwaha, Monali Chaturvedi, and Vidur Mahajan. Systematic review of different neuroimaging correlates in mild cognitive impairment and alzheimer’s disease. *Clinical Neuroradiology*, 31(4):953–967, 2021.
- [22] What Is fMRI? - Center for Functional MRI - UC San Diego, . URL <http://fmri.ucsd.edu/Research/whatisfmri.html>.
- [23] Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

- [24] Russell A Poldrack, Jeanette A Mumford, and Thomas E Nichols. *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.
- [25] CONN toolbox, . URL <https://web.conn-toolbox.org/home>.
- [26] Stephen C Strother. Evaluating fmri preprocessing pipelines. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):27–41, 2006.
- [27] Alfonso Nieto-Castanon. *Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN*. Hilbert Press, 2020.
- [28] César Caballero-Gaudes and Richard C Reynolds. Methods for cleaning the bold fmri signal. *Neuroimage*, 154:128–149, 2017.
- [29] Harvard-Oxford cortical and subcortical structural atlases, . URL <https://neurovault.org/collections/262/>.
- [30] AAL / AAL2 / AAL3 – Neurofunctional Imaging Group (GIN-IMN), . URL <https://www.gin.cnrs.fr/en/tools/aal/>.
- [31] CONN toolbox - Networks (voxel-level), . URL <https://web.conn-toolbox.org/fmri-methods/connectivity-measures/networks-voxel-level>.
- [32] CONN toolbox - Other, . URL <https://web.conn-toolbox.org/fmri-methods/connectivity-measures/other>.
- [33] Giovanni B Frisoni, Nick C Fox, Clifford R Jack, Philip Scheltens, and Paul M Thompson. The clinical use of structural mri in alzheimer disease. *Nature Reviews Neurology*, 6(2):67–77, 2010.
- [34] Seyed Hani Hojjati, Ata Ebrahimzadeh, Ali Khazaei, Abbas Babajani-Feremi, Alzheimer’s Disease Neuroimaging Initiative, et al. Predicting conversion from mci to ad by integrating rs-fmri and structural mri. *Computers in biology and medicine*, 102:30–39, 2018.
- [35] Arkan Al-Zubaidi, Alfred Mertins, Marcus Heldmann, Kamila Jauch-Chara, and Thomas F. Münte. Machine learning based classification of resting-state fmri features exemplified by metabolic state (hunger/satiety). *Frontiers in Human Neuroscience*, 13, 2019. ISSN 1662-5161. DOI: 10.3389/fnhum.2019.00164. URL <https://www.frontiersin.org/article/10.3389/fnhum.2019.00164>.
- [36] Gloria Castellazzi, Maria Giovanna Cuzzoni, Matteo Cotta Ramusino, Daniele Martinelli, Federica Denaro, Antonio Ricciardi, Paolo Vitali, Nicoletta Anzalone, Sara Bernini, Fulvia Palesi, Elena Sinforiani, Alfredo Costa, Giuseppe Micieli, Egidio D’Angelo, Giovanni Magenes, and Claudia A. M. Gandini Wheeler-Kingshott. A machine learning approach for the differential diagnosis of alzheimer and vascular dementia fed by mri selected features. *Frontiers in Neuroinformatics*, 14, 2020. ISSN 1662-5196. DOI: 10.3389/fninf.2020.00025. URL <https://www.frontiersin.org/article/10.3389/fninf.2020.00025>.
- [37] Xiaoke Hao, Yongjin Bao, Yingchun Guo, Ming Yu, Daoqiang Zhang, Shannon L. Risacher, Andrew J. Saykin, Xiaohui Yao, and Li Shen. Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer’s disease. *Medical Image Analysis*, 60:101625, 2020. ISSN 1361-8415. DOI: <https://doi.org/10.1016/j.media.2019.101625>. URL <https://www.sciencedirect.com/science/article/pii/S1361841519301616>.



- [38] Lauren Burrell, Otis Smart, George K Georgoulas, Eric D Marsh, and George J Vachtsevanos. Evaluation of feature selection techniques for analysis of functional mri and eeg. In *DMIN*, pages 256–262, 2007.
- [39] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.
- [40] SequentialFeatureSelector: The popular forward and backward feature selection approaches incl. floating variants - mlxtend, . URL [http://rasbt.github.io/mlxtend/user\\_guide/feature\\_selection/SequentialFeatureSelector/](http://rasbt.github.io/mlxtend/user_guide/feature_selection/SequentialFeatureSelector/).
- [41] Petr Somol, Jana Novovicova, and Pavel Pudil. Efficient Feature Subset Selection and Subset Size Optimization. In *Pattern Recognition Recent Advances*, volume 56. February 2010. ISBN 978-953-7619-90-9. DOI: 10.5772/9356. Journal Abbreviation: Pattern Recognition Recent Advances.
- [42] MLMath.io. Math behind Support Vector Machine(SVM), February 2019. URL <https://ankitnitjsr13.medium.com/math-behind-support-vector-machine-svm-5e7376d0ee4d>.
- [43] Sanjay.M. Why and how to Cross Validate a Model?, August 2020. URL <https://towardsdatascience.com/why-and-how-to-cross-validate-a-model-d6424b45261f>.
- [44] Dealing with very small datasets, . URL <https://kaggle.com/code/rafjaa/dealing-with-very-small-datasets>.
- [45] 7 Effective Ways to Deal With a Small Dataset | HackerNoon, . URL <https://hackernoon.com/7-effective-ways-to-deal-with-a-small-dataset-2gyl407s>.
- [46] Fredrik Lundh. An introduction to tkinter. URL: [www.pythonware.com/library/tkinter/introduction/index.htm](http://www.pythonware.com/library/tkinter/introduction/index.htm), 1999.
- [47] Call MATLAB from Python - MATLAB & Simulink - MathWorks Nordic, . URL <https://se.mathworks.com/help/matlab/matlab-engine-for-python.html>.
- [48] craigloewen msft. Install WSL. URL <https://docs.microsoft.com/en-us/windows/wsl/install>.
- [49] subprocess — Subprocess management — Python 3.10.6 documentation, . URL <https://docs.python.org/3/library/subprocess.html>.
- [50] FreeSurferWiki - Free Surfer Wiki, . URL <https://surfer.nmr.mgh.harvard.edu/fswiki>.
- [51] Hiroshi Matsuda. The role of neuroimaging in mild cognitive impairment. *Neuropathology*, 27(6):570–577, 2007.
- [52] Anqi Qiu, Christine Fennema-Notestine, Anders M Dale, Michael I Miller, Alzheimer’s Disease Neuroimaging Initiative, et al. Regional shape abnormalities in mild cognitive impairment and alzheimer’s disease. *Neuroimage*, 45(3):656–661, 2009.
- [53] Statistical functions (scipy.stats) — SciPy v1.8.1 Manual, . URL <https://docs.scipy.org/doc/scipy/reference/stats.html>.
- [54] SPM - Documentation, . URL <https://www.fil.ion.ucl.ac.uk/spm/doc/>.

- [55] 3.2. Tuning the hyper-parameters of an estimator, . URL [https://scikit-learn/stable/modules/grid\\_search.html](https://scikit-learn/stable/modules/grid_search.html).
- [56] sklearn.ensemble.RandomForestClassifier, . URL <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>.
- [57] sklearn.svm.SVC, . URL <https://scikit-learn/stable/modules/generated/sklearn.svm.SVC.html>.
- [58] Ina Iachini, Alessandro Iavarone, Vincenzo Paolo Senese, Francesco Ruotolo, and Gennaro Ruggiero. Visuospatial memory in healthy elderly, AD and MCI: a review. *Current Aging Science*, 2(1):43–59, March 2009. ISSN 1874-6128. DOI: 10.2174/1874609810902010043.
- [59] Linda L Chao, James V Haxby, and Alex Martin. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience*, 2(10):913–919, 1999.
- [60] Roberto Cabeza and Lars Nyberg. Imaging cognition ii: An empirical review of 275 pet and fmri studies. *Journal of cognitive neuroscience*, 12(1):1–47, 2000.
- [61] Daniel Tranel, Hanna Damasio, and Antonio R Damasio. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, 35(10):1319–1327, 1997.
- [62] Ilan I Goldberg, Michal Harel, and Rafael Malach. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron*, 50(2):329–339, 2006.
- [63] David Caplan. Why is broca’s area involved in syntax? *Cortex*, 42(4):469–471, 2006.
- [64] Maki S Koyama, David O’Connor, Zarrar Shehzad, and Michael P Milham. Differential contributions of the middle frontal gyrus functional connectivity to literacy and numeracy. *Scientific reports*, 7(1):1–13, 2017.
- [65] TA Yousry, UD Schmid, H Alkadhi, D Schmidt, Aurelia Peraud, Andreas Buettner, and P Winkler. Localization of the motor hand area to a knob on the precentral gyrus. a new landmark. *Brain: a journal of neurology*, 120(1):141–157, 1997.
- [66] Isabelle Klein, Anne-Lise Paradis, Jean-Baptiste Poline, Stephen M Kosslyn, and Denis Le Bihan. Transient activity in the human calcarine cortex during visual-mental imagery: an event-related fmri study. *Journal of cognitive neuroscience*, 12(Supplement 2):15–23, 2000.
- [67] Clive R Rosenthal, Samantha K Andrews, Chrystalina A Antoniadis, Christopher Kennard, and David Soto. Learning and recognition of a non-conscious sequence of events in human primary visual cortex. *Current Biology*, 26(6):834–841, 2016.
- [68] Sebastian Bludau, Simon B Eickhoff, Hartmut Mohlberg, Svenja Caspers, Angela R Laird, Peter T Fox, Axel Schleicher, Karl Zilles, and Katrin Amunts. Cytoarchitecture, probability maps and functions of the human frontal pole. *Neuroimage*, 93:260–275, 2014.
- [69] Andrea Mechelli, Glyn W Humphreys, Kate Mayall, Andrew Olson, and Cathy J Price. Differential effects of word length and visual contrast in the fusiform and lingual gyri during. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1455):1909–1913, 2000.

- [70] Xiaozheng Liu, Wei Chen, Hongtao Hou, Xingli Chen, Jiangtao Zhang, Jian Liu, Zhongwei Guo, and Guanghui Bai. Decreased functional connectivity between the dorsal anterior cingulate cortex and lingual gyrus in alzheimer’s disease patients with depression. *Behavioural brain research*, 326:132–138, 2017.
- [71] Kalanit Grill-Spector, Zoe Kourtzi, and Nancy Kanwisher. The lateral occipital complex and its role in object recognition. *Vision research*, 41(10-11):1409–1422, 2001.
- [72] CR McDonald, LK McEvoy, L Gharapetian, C Fennema-Notestine, DJ Hagler, D Holland, A Koyama, JB Brewer, AM Dale, et al. Regional rates of neocortical atrophy from normal aging to early alzheimer disease. *Neurology*, 73(6):457–465, 2009.
- [73] Russell A Poldrack, Chris I Baker, Joke Durnez, Krzysztof J Gorgolewski, Paul M Matthews, Marcus R Munafò, Thomas E Nichols, Jean-Baptiste Poline, Edward Vul, and Tal Yarkoni. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature reviews neuroscience*, 18(2):115–126, 2017.
- [74] Gary W Van Hoesen, Bradley T Hyman, and Antonio R Damasio. Entorhinal cortex pathology in alzheimer’s disease. *Hippocampus*, 1(1):1–8, 1991.
- [75] Gary W van Hoesen, Jean C Augustinack, Jason Dierking, Sarah J Redman, and Ramasamy Thangavel. The parahippocampal gyrus in alzheimer’s disease: clinical and preclinical neuroanatomical correlates. *Annals of the New York Academy of Sciences*, 911(1):254–274, 2000.