

TDK-dolgozat

Németh Ágnes Sára

**Szegedi Tudományegyetem
Természettudományi és Informatikai Kar
Informatikai Intézet
Műszaki Informatika Tanszék**

**Intra-breath oszcillometriás mérések feldolgozása és gépi
tanulósos osztályozása**

Készítette:

Németh Ágnes Sára
Mérnökinformatikus MSc
szakos hallgató
II. évfolyam

Témavezetők:

Dr. Kalmár György
adjunktus

Makan Gergely
tanársegéd

Szeged
2024

Tartalomjegyzék

Tartalomjegyzék.....	3
1. BEVEZETÉS	4
2. IRODALMI ÁTTEKINTÉS	6
2.1. Intra-breath oszcillometria.....	6
2.2. Kilégzési áramláskorlátozás	7
2.3. Gépi tanuló módszerek az oszcillometria területén.....	9
3. A MÉRÉSEK FELDOLGOZÁSA.....	10
3.1. ADAT BEOLVASÁS ÉS KALIBRÁCIÓ	12
3.2. Adatfeldolgozás	13
3.3. Impedancia hurkok.....	15
3.4. Kiugró értékek felismerése (outlier detection).....	17
4. OSZTÁLYOZÁS.....	20
4.1. Jellemzők kinyerése	21
4.2. Osztályozó algoritmusok.....	25
5. ALKALMAZÁS.....	29
5.1 ILD tesztanyag.....	29
5.2 OHS beteganyag.....	31
6. ÖSSZEFOGLALÁS.....	35
Irodalomjegyzék.....	37
MELLÉKLETEK	40

1. BEVEZETÉS

A dolgozat célja elemezni a nyugodt légzés közbeni kilégzési áramláskorlátozás (tEFL- tidal expiratory flow limitation) jelenlétét obezitás-hipoventiláció szindrómában (OHS) szenvedő betegek oszcillometriás mérései alapján, jelfeldolgozási és gépi tanulási eszközökkel segített osztályozási módszerekkel. A kilégzési áramláskorlátozás egy markáns légzésmechanikai elváltozás, ahol a belégzés akadálymentes, viszont a kilégzési folyamat a perifériás légutak megváltozott mechanikája miatt rendkívül akadályozott [1]. A légzőizmok kilégzési aktivitása ugyanis nem csak a légzőhólyagocskák kiürülését váltja ki, hanem a kivezető légutak szűkülését is eredményezi, nyugodt légzés esetében is. Ennek a két jelenségnek a dinamikai egyensúlya miatt a kilégzési folyamat elnyújtott és alacsony áramlású lehet. A tEFL egy jellemző fő tünete a krónikus obstruktív tüdőbetegségnek (COPD) [2], melyben a perifériás légutak stabilitását főként kémiai hatások (dohányzás) legyengítik: a tüdőszövet roncsolódása következtében csökken a légutakat kifeszítő elasztikus erő. Az OHS páciensek esetében viszont a nagy testtömeg ($BMI \geq 30 \text{ kg/m}^2$) nyomást fejt ki a mellkasfalra és a hasüreg közvetítésével a rekeszizomra, ezzel csökkenti a tüdő kapacitását és a légutak átmérőjét közelebb viszi az elzáródási küszöbhez, megkönnyítve a tEFL fellépését [3].

A légzőrendszer mechanikájának oszcillometriás vizsgálata a nyugodt légzésre szuperponált külső vizsgálójel alkalmazásával a légzőrendszer bemeneti impedanciájának (komplex Z_{rs}) meghatározását teszi lehetővé [4]. Ez a mérési technika szokásos módon többfrekvenciás vizsgálójelet tartalmaz, melynek sávja a légzési frekvenciánál nagyságrendileg magasabb tartományban (4-40 Hz) helyezkedik el. A mérés különösebb együttműködést nem igényel, emiatt széles kortartományban és számos betegcsoportnál előnyös vizsgálati módszer. A bemeneti impedancia értékét általában annak disszipatív (valós rész vagy rezisztencia, R_{rs}) és az energiatárolást jellemző szöveti elasztikus és inertív (képzetes rész vagy reaktancia, X_{rs}) részével adják meg. Az eredmények szokásos módon 20-30 s hosszúságú mérésekből számolt R_{rs} és X_{rs} átlagértékek [5].

Az intra-breath oszcillometria nem a Z_{rs} átlagos értékeire, hanem annak a légzés során bekövetkező ciklikus változásaira összpontosít [6]. Ennek érdekében és a kellő jel-zaj viszony elérése céljából egyfrekvenciás (felnőtteknél 10 Hz, csecsemőknél 16 Hz) vizsgálójelet alkalmaz, ezzel felfedve az R_{rs} és az X_{rs} változása mögötti mechanikai meghatározókat, mint például az R_{rs} légúti turbulencia miatti áramlásfüggése, valamint a nyomás- és térfogatfüggésre visszavezethető nemlinearitások. A vizsgálójel periódusidejének megfelelő rövid időszaka alatt

kiszámolva a Zrs értékeit az idővariáns és nemlineáris rendszer szakaszonkénti lineáris megközelítését alkalmazzuk.

Az intra-breath technika a Szegedi Tudományegyetemen a Műszaki Informatika Tanszék munkatársai részvételével folyó kutatómunka tárgya és eredménye, e csoporthoz nyílt alkalmam csatlakozni és hallgatói résztvevőként dolgozni az NKFIH-OTKA 128701 sz. pályázatban.

A dolgozatban intra-breath impedancia hurkokkal foglalkozom, ami a szakaszonként számolt Rrs és Xrs ábrázolása a légzési térfogat és a légzési áramlás függvényében. Ez az ábrázolási mód látványosan jeleníti meg a légzés közbeni ciklikus mechanikai folyamatokat és a mérési műtermékeket is (például a hangrés összeszűkülése vagy nyelés), valamint képes megjeleníteni a légzési ciklusok variabilitását és lehetővé teszi a szelektálás után az átlagos hurok definiálását.

A mérési adatok a légutak bemeneténél mért nyomás és áramlás mintavételezett sorozatai, melyek a spontán légzésből származó és az oszcillációs vizsgálójel eredményeképp létrejött komponensek szuperpozíciói. E jelek feldolgozásához először egy beolvasó programot írtam, amely a nyers áramlási és nyomás adatokból, a megfelelő kalibráció után, kiszámolja az impedanciát, majd légzési ciklusokra bontja a mérési adatokat. Ezt követően automatizáltan kiszűrtem a mérési artefaktumokat, és a tisztított adatokból előállítottam egy átlagos hurkot minden mérési szakaszra, amelyre a tEFL-t leíró jellemzőket definiáltam.

A gépi tanuló módszerek segítségével elsősorban a tEFL fellépését detektáltam az OHS beteganyagban a folyamatos pozitív légúti nyomástámogatás (CPAP: continuous positive airways pressure) egyes értékeihez tartozó mérési szakaszok során. A modell tanításához egészséges kontroll és COPD-s személyek intra-breath méréseit használtam fel, mivel a COPD egy fő tünete a kilégzési áramláskorlátozás. Ezeken a jellemzőkön végeztem a tanításokat, ahol a hiperparaméterek hangolását kereszt-validációs módszerekkel végeztem.

A Semmelweis Egyetemen az Otthoni lélegeztetési program keretében végeztek egy vizsgálatot OHS-ben szenvedő betegeken, amelyben minden beteget ülő és fekvő testhelyzetben megmértek különböző CPAP mellett. A pozitív légúti nyomás ugyanis segít nyitva tartani a tüdőt, így csökkentve a testsúly által okozott nyomás hatását. Ezeket a méréseket kézzel kiértékeltek és az eredményeket publikálták [3]. Munkám során az adathalmaz újraértékelést végeztem el az algoritmusommal minden légzéstámogatási szint mellett. A dolgozat elkészítése során a feladatom az volt, hogy átfogóan megismerjem az intra-breath oszcillometriai mérés technikát és készítsek egy programot, amely az adatok előfeldolgozása és vizualizálása után a gépi tanuló módszerek integrált alkalmazásával elvégzi az intra-breath

oszcillometria automatikus kiértékelését, tipizálását változatos pulmonológiai fenotípusok, kohorszvizsgálatok és követéses stúdiumok számára. A módszer használhatóságát demonstrálandó egy modellt kellett betanítanom a tEFL felismerésére, majd tesztelnem a meglévő mérési adathalmazon.

A következő, 2. fejezetben az intra-breath oszcillometriával kapcsolatos irodalmat mutatom be. Itt kitérek a mérés technika elveire, történelmére, majd specifikusan a COPD és tEFL kapcsolatára és a CPAP hatására COPD-s betegeknél, és bemutatom az OHS-es adathalmazon végzett eddigi kutatást. A 3. fejezetben az adatfeldolgozás lépéseit írom le, az adatbeolvasás, kalibráció, impedancia számítás, hurkok értelmezése és artefaktum szűrés kap helyet. A 4. fejezet a gépi tanuló algoritmus tanítását taglalja, itt a tanító adathalmazt, a jellemzők kinyerésének lépéseit és az algoritmusok eredményeinek összehasonlítását írom le. Az 5. fejezetben az algoritmus tesztelését végzem el egy speciális betegcsoport (intersticiális tüdőbetegség: ILD) mérési adatait használva, majd az alkalmazását mutatom be az OHS-es betegek mérésein. A dolgozatot a munkám összefoglalásával zárom.

2. Irodalmi áttekintés

Ebben a fejezetben elsősorban az intra-breath oszcillometriát mutatom be. Először a mérési módszert mutatom be és hasonlítom össze más oszcillometriás technikákkal, majd kitérek a kilégzési áramlás korlátozás mintázataira a COPD-s betegek esetén, illetve a CPAP hatására ebben a betegcsoportban. Végül bemutatom az általam elemzett adathalmazt eredetét és az eddigi eredményeket. Az utolsó alponthoz más oszcillometriás mérések gépi tanulós elemzését ismertetem. Az egyszerűbb algoritmusoktól kiindulva a mély neurális hálózatokig terjedő módszerekkel végzett elemzéseket és azok eredményeit ismertetem.

2.1. Intra-breath oszcillometria

Az oszcillometria egy olyan mérési módszer, amely a légzőrendszer mechanikai tulajdonságait vizsgálja külső vizsgálojel segítségével. A szokásos oszcillometriás mérések során a páciens nyugodt légzésére szuperponálnak kis amplitúdójú, nyomáshullámokat, amelyeket általában egy hangszóró vagy rezgő szitamembrán állít elő. A vizsgálat alatt az áramlási sebességet és a nyomásváltozásokat folyamatosan rögzítik. Az így kapott adatokból számolják ki a légzőrendszer mechanikai tulajdonságait, mint például a légutak és a szövetek rezisztenciáját és reaktanciáját különböző frekvenciákon [4]. Ezek az értékek segítenek a légúti betegségek

diagnosztizálásában, például a gyakori krónikus obstruktív tüdőbetegség (COPD) vagy az asztma esetében [7].

Az első oszcillometriával kapcsolatos publikációt 1956-ban adták ki, amely innovatív megközelítést kínált a légzőrendszer alapvető megértéséhez [8]. A kutatók szinuszos nyomáshullámokat használtak mérőjelként 1 és 35 Hz között, így felfedezték, hogy a légzőrendszer bemeneti impedanciája frekvenciafüggő. Alacsony frekvenciákon a reaktancia a szöveti elasztikus (térfogatfüggő) tulajdonságokat tükrözi, magasabb frekvenciákon viszont a gyorsuláshoz kapcsolódó tehetetlenségi (inertív) reaktancia a meghatározó. DuBois és munkatársai [8] részletesen elemezték az áramlás és a nyomás kapcsolatát (ezt a 3.1 fejezetben fogom kifejteni); párhuzamot vontak mechanikai-akusztikus és az elektronikai analogonok között és ezzel szemléltették a légzőrendszer biomechanikájának modelljét.

A pulmonológiai gyakorlatban használatos oszcillometria az előző bekezdésben ismertetett elvek alapján működik [9]. Általában többfrekvenciás a vizsgálójel, 20-30 s hosszú mérések során egy átlagértéket ad a R_{rs} -re és X_{rs} -re is. Ez az időskála viszont nem teszi lehetővé a légzőrendszer finom dinamikájának vizsgálatát a légzési cikluson belül [4]. Az időbeli felbontásra koncentrálnak érdekében az impedancia értékeit érdemes egy rövid időablakra számolni, például 0,1-0,2 s hosszú csúszóablak alkalmazásával. Ezen felül célszerű egy frekvenciát használni, hogy javítsuk a jel/zaj viszonyt és elkerüljük a mérőfrekvenciák modulációs oldalsávjainak átfedését. Az intra-breath oszcillometria fő célja tehát az egyéni légzésciklusokon belül a Z_{rs} változásának vizsgálata, ezzel feltérképezve a Z_{rs} nemlineáris függését a légzési áramlástól és a térfogattól. Különös fontosságúak a nulla légzési áramláshoz tartozó (belégzéskezdeti és -végi) Z_{rs} értékek, melyeknél minimális a nemlineáris felső légutak befolyása és ahol a tüdő mechanikai állapota így tisztábban látszik. Az első intra-breath mérésekről 1971-ben számoltak be [10], és e módszer az elmúlt évtizedben gyorsan terjed [6].

2.2. Kilégzési áramláskorlátozás

Tidális kilégzési áramlás-korlátozásról (tEFL) akkor beszélünk, amikor nyugodt légzés során megnehezül a kilégzési fázis: az áramlás eléri egy alacsony csúcserőértéket és tovább csökken még a fokozott légzőizom aktivitás ellenére is. Ilyenkor a légszomj lép fel és fokozott légzési munka szükséges. Ez az obstruktív tüdőbetegségek - mint pl. a COPD - egy tipikus tünete. Dellacà és munkatársai [2] a tEFL impedancia mintázatait tanulmányozták, a tüdőn keresztüli (transzpulmonális) nyomás mérésével együtt. A vizsgálatban stabil COPD-s betegek és megfelelő életkorú egészséges személyek vettek részt. A mérés során 5 Hz-es vizsgálójellett alkalmaztak 200 Hz-es mintavételi frekvencia mellett, és minden mintavételi pontban

kiszámolták az impedanciát egy 0,2 s-os csúszóablak alkalmazásával. Négy indikátort használtak az EFL detektálására: átlag kilégzési Xrs, minimum kilégzési Xrs, a kilégzési és belégzési átlag Xrs különbsége, illetve a különbség a belégzés maximális Xrs értéke és a kilégzés minimális Xrs értéke között. Ezeket az értékeket minden légvételre kiszámították és kiértékeltek, majd három csoportba sorolták a pácienseket: áramláslimitált, nem áramláslimitált és bizonytalan. A megfelelő küszöbszintek kiválasztása után a besorolás 100%-os specificitással és érzékenységgel történt. A tEFL élettani alátámasztását is nyújtó tanulmány bebizonyította, hogy ez az egyszerűen használható, érzékeny, specifikus és neminvaszív mérés technika megfelelő a tEFL megbízható kimutatásához.

A tEFL oszcillometriával is kimutatható jelenségét különböző súlyossági szintű COPD akut exacerbációjakor és egészséges kontroll személyekben vizsgálták folyamatos pozitív légúti nyomás mellett [11]. A tanulmány során 8 Hz-es vizsgálójelet és 0, 4, 8, 14, 20 vízcm értékű CPAP légzéstámogatást alkalmaztak. Kimutatták, hogy egészséges személyek esetén XeE és XeI értéke közel azonos, viszont COPD-ben XeE jóval kisebb, mint az XeI. Emellett a belégzési szakaszban az Xrs növekedése egy nagyfokú inhomogenitás enyhülését jelzi a COPD-s betegek esetén. A CPAP növelésének hatására a tEFL mintázatok szintről-szintre gyengültek, betegenként eltérő mértékben, mivel a folyamatos pozitív nyomás a hörgők kilégzési elzáródását mérsékli. Ebben tanulmányban jelenítik meg elsőként az impedancia hurkokat a légzéstámogatás növekvő szintjein - így szemmel követhető a változás, amely során az Rrs kevésbé változik, Xrs pedig megmutatja, hogyan csökken a tEFL és a COPD-sekre jellemző inhomogenitás a légzéstámogatás növelésével. A tEFL mintázata ugyanakkor elfedhet más fontos Zrs mintázatokat, ami esetleg egyéb tünetekért felelős.

Az előzőhöz hasonló tanulmányt végeztek obezitás-hipoventiláció szindrómában (OHS) érintetteknél [3]. A tEFL náluk is jelentős tünet, de az ok különbözik. A túlzott mellkasfali és hasi zsírtömeg összenyomja a tüdőt, korlátozva annak teljes kitágulási képességét (tüdőszűkület). Emellett a felső légutak problémái, például a beszűkült légutak tovább akadályozhatják a légáramlást. A tanulmány során 35 beteget mértek meg ülő és fekvő testhelyzetben 0, 5, 10, 15, 20 vízcm légzéstámogatás mellett. A vizsgálat végén megismételték a támogatás nélküli mérést, hogy elemezni tudják a CPAP hatásának esetleges maradandóságát. A mérés során 10 Hz-es vizsgálójelet alkalmaztak, 256 Hz-es mintavételi frekvencia mellett. A CPAP terápia hasonló hatást ért el, mint a COPD-s betegeknél, szintről szintre gyengítette a tEFL mintázatot. Korábbi vizsgálatok eredményeivel összhangban a fekvő testhelyzet Zrs emelkedésével járt; megfigyelték ugyanakkor, hogy e betegeknél a tEFL nagyobb mértékben lép fel a fekvő testhelyzetben. A méréseket kézzel értékelték ki és elemezték. Az eredmények

alapján finomítható az otthoni (főleg éjszakai) légzéstartogatás optimális CPAP szintjének beállítása.

2.3. Gépi tanuló módszerek az oszcillometria területén

Az irodalmi áttekintés zárásaként olyan tanulmányokat mutatok be, amelyek célja és felhasznált eszközei hasonlítanak az én céljaimhoz, csak intra-breath helyett más oszcillometriás adatokon végezték az elemzéseket. Intra-breath mérésekkel ilyen publikációk még nem születtek. Egy áttekintő közlemény összefoglalja az eddigi eredményeket ezen a területen, és kifejti a COPD korai észlelésének kihívásait és fontosságát, hiszen az első tünetektől átlagosan 3,6 év telik el a diagnózisig, ezzel elveszítve a korai beavatkozás esélyét [12].

Amaral és munkatársai gépi tanuló módszerek segítségével detektálták a COPD-megnyilvánulásait hagyományos oszcillometriás adatokban [13]. A kísérlet során 150 mérést elemeztek, amelyek 25 COPD-s és 25 egészséges személytől származtak. A mérések során olyan vizsgálójelet alkalmaztak, ami 4 Hz és 32 Hz között a 2 Hz minden egész számú többszörösén tartalmaz szinuszos komponenst, ezt a jelet 1024 Hz-el mintavételezték, majd 16 másodperces ablakokra Fourier transzformációt alkalmaztak, és minden frekvenciára kaptak egy átlag impedanciaértéket. Öt féle jellemzőt használnak, ezek közül egyik sem egyezik azokkal, amiket én használtam. Ezek a jellemzők a rezisztencia extrapolált értéke 0 Hz-re, a mért frekvenciákon a rezisztenciák és a reaktanciák átlaga, illetve a rezisztencia görbe meredeksége, rezonancia frekvencia, és effektív compliance. Ezekből a jellemzőkből a lineáris korrelációs együtthatók alapján választották ki leginformatívabbakat, itt először megkeresték azt a jellemzőt, amely legjobban korrelál a kimenettel, majd eltávolították azokat a jellemzőket, amelyek jobban korrelálnak a legjobb jellemzővel, mint a kimenettel, mivel ezek nem rendelkeznek releváns plusz információ tartalommal. Végül öt féle osztályozó teljesítményét hasonlították össze: lineáris normál Bayes osztályozó (LBNC – linear Bayes Normal Classifier), K legközelebbi szomszéd (KNN – K nearest neighbor), döntési fa, mesterséges neuronháló és support vector machines. A kísérlet során a legjobb eredményt a KNN érte el, 87%-os érzékenységgel és 97% specificitással.

Srivastava és munkatársai légzési hangok alapján különböztették meg a COPD-s betegeket és az egészséges személyeket [14]. Ehhez 126 alany címkézett tüdő hangfelvételeit használták, amelyeket adatnövelési (*data augmentation*) technikákkal bővítettek ki. Az osztályozáshoz Konvolúciós Neurális Hálózatokat (CNN-eket) alkalmaztak. A jellemzők között szerepelnek a hangfeldolgozásban elterjedt Mel-Frequency Cepstral Coefficients (MFCC), a Mel-spektrogram és a Chroma (Constant-Q). Az osztályozásnál 93%-os

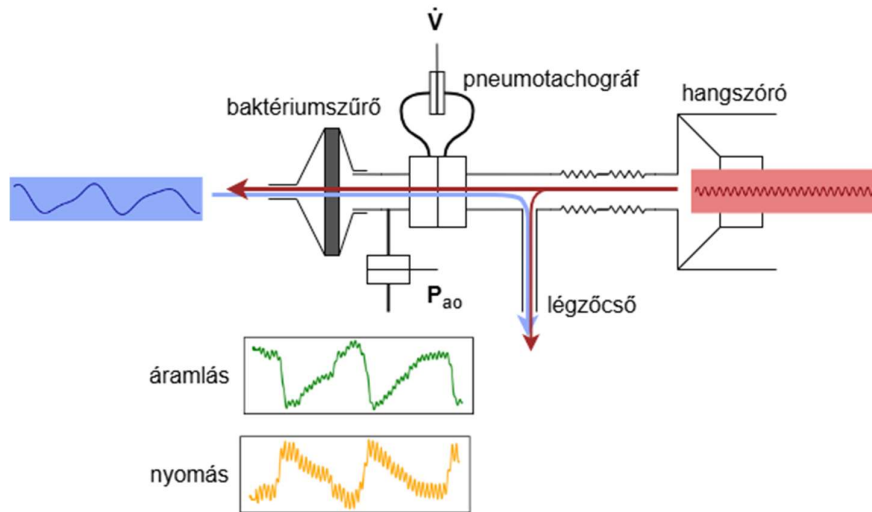
pontosságot értek el, emellett a COPD súlyosságát is meghatározták enyhe, mérsékelt és súlyos kategóriákba sorolva. A tanulmány hangsúlyozza, hogy a mélytanulás ígéretes eszköz lehet a COPD légzési hang alapú diagnosztikájában, amely nem invazív és hatékony módszert kínál az orvosi ellátás támogatására és a szakemberhiány kezelésére.

Összességében az egészséges és COPD-s betegek megkülönböztetése gépi tanulási algoritmusokkal számos megközelítésben megvalósult már. Az oszcillometriás mérések és a gépi tanulási módszerek alkalmazása széles körű, azonban ezek is általában kis méretű adathalmazokra támaszkodnak. Mindezek ellenére intra-breath oszcillometriával eddig még nem végeztek hasonló jellegű vizsgálatot. Hasonló hiányosság figyelhető meg az OHS-ben szenvedők légzőrendszeri impedancia elemzése terén is, amelyet eddig kizárólag Baglyas és munkatársai tanulmányozták.

3. A mérések feldolgozása

Több oszcillometriás mérőeszköz is elérhető a kereskedelemben, azonban lényegében megegyezik a működési alapelvük, amit a 3.1-es ábrán mutatok be. Felnőttek mérése során a páciens általában ülő testhelyzetben van: csipesszel zárja az orrpályát, a baktériumszűrő szájrészére zárja az ajkakat és az arcát megtámasztja.

Az eszköz hangszóró segítségével állítja elő az oszcillációs mérőjelet, amely egyfrekvenciás is lehet és így alkalmas intra-breath mérésekre is. Ez felnőttek esetében célszerűen 10 Hz frekvenciájú; az amplitúdó a normál légzéséhez képest közel ötször kisebb. A mérőfejben összegződik a mérőjel és a páciens légzéséből származó spontán áramlásjel is. Ezen a helyen mérjük a légúti bemeneti nyomást (Pao – airway opening pressure), amely magába foglalja a baktériumszűrőn eső nyomást is, hiszen a vizsgálójel ezen keresztül jut a légzőrendszer bemenetére. Az áramlás ($\dot{V}$) méréséhez egy pneumotachográfot használnak, amely egy mérőellenállás (általában műanyag vagy fém szita) két oldalán lévő nyomáskülönbséget érzékeli. A szenzorjeleket szinkronban mintavételezik, az általam vizsgált mérések esetében 256 Hz-el.



3.1. ábra: Mérési elrendezés szemléltetése: A kék jel a páciens nyugalmi légzésének áramlás jele, a piros a hangszóró által generált 10 Hz-es szinusz jel, a szenzorok ezen jelek szuperpozícióját méri (zöld és narancsszín).

A légzőrendszer, mint mért rendszer nemlineáris és idővariáns viselkedésű, de rövid időszakokra feltételezhető a linearitás. Így a viselkedése leírható differenciálegyenletekkel és érvényes rá a szuperpozíció elve. Ez előnyös, mivel így minden áramlás komponens külön kezelhető más bemeneti jelektől függetlenül, így a vizsgálójel és a rendszer vizsgálójelre adott válasza spontán légzés mellett is elkülöníthető. Feltételezzük, hogy a rendszer időinvariáns, tehát a rendszer struktúrája és a paraméterei nem változnak az időben. Ez a valóságban nem igaz a légzőrendszerre, mivel annak a mechanikai paraméterei ciklikusan változnak a légzési áramlás és térfogat függvényében, azonban elfogadható az állandósult állapot (azaz a szakaszonkénti linearitás) feltételezése egy kontrollált körülmények között történő *rövid* idejű mérésnek.

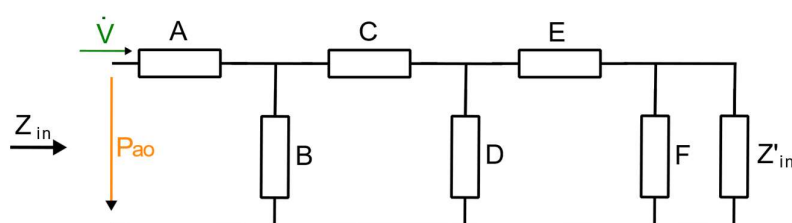
A légzésmechanika területén a mechanikai impedancia $[Z(f)]$ egy frekvenciatérbeli kapcsolat az áramlás $[\dot{V}(f)]$ és a nyomás $[P(f)]$ között (ezek az elektromos áram és feszültség analogonjai). A kapcsolatot a 3.1. egyenlet írja le, ezen értékek frekvenciatérben értelmezettek és meghatározzák az amplitúdót és az eltolást (fázist) is, vagy más alakban a valós és képzetes részt.

$$Z(f) = \frac{P(f)}{\dot{V}(f)}$$

3.1. egyenlet: A légzőrendszeri mechanikai impedancia számításának képlete.

3.1. Adat beolvasás és kalibráció

A $P(t)$ és $\dot{V}(t)$ jelek mintavételezett adatait a mérőeszközök szöveges (*txt*) vagy *inpx* kiterjesztésű fájlokban exportálják; az első egy általános szövegformátum, amely a nyers adatokat tartalmazza a hozzájuk tartozó időbélyeggel. A második egy bináris fájl formátum, amelyben először egy fejléc található a mérőeszközhöz és a mérési alkalomra jellemző adatokkal, majd maga az adat. A beolvasott szenzoradat viszont önmagában nem elegendő. A mérés közben a P_{ao} értékébe belemért, a baktériumszűrőn eső nyomás és az oszcillációs áramlásnak a csőszakaszok gázkompresszibilitásból adódó csillapítása az alább részletezett kalibrációs lépéssel korrigálható.



3.2. ábra: Kalibrációs modell.

A 3.2. ábrán mutatom be a kalibráció számításánál használt modellt. Itt a Z'_{in} a tényleges légzőrendszeri bemeneti impedancia, viszont az általunk mért Z_{in} tartalmazza a képen látható impedanciahálót. A létrahálózat részei:

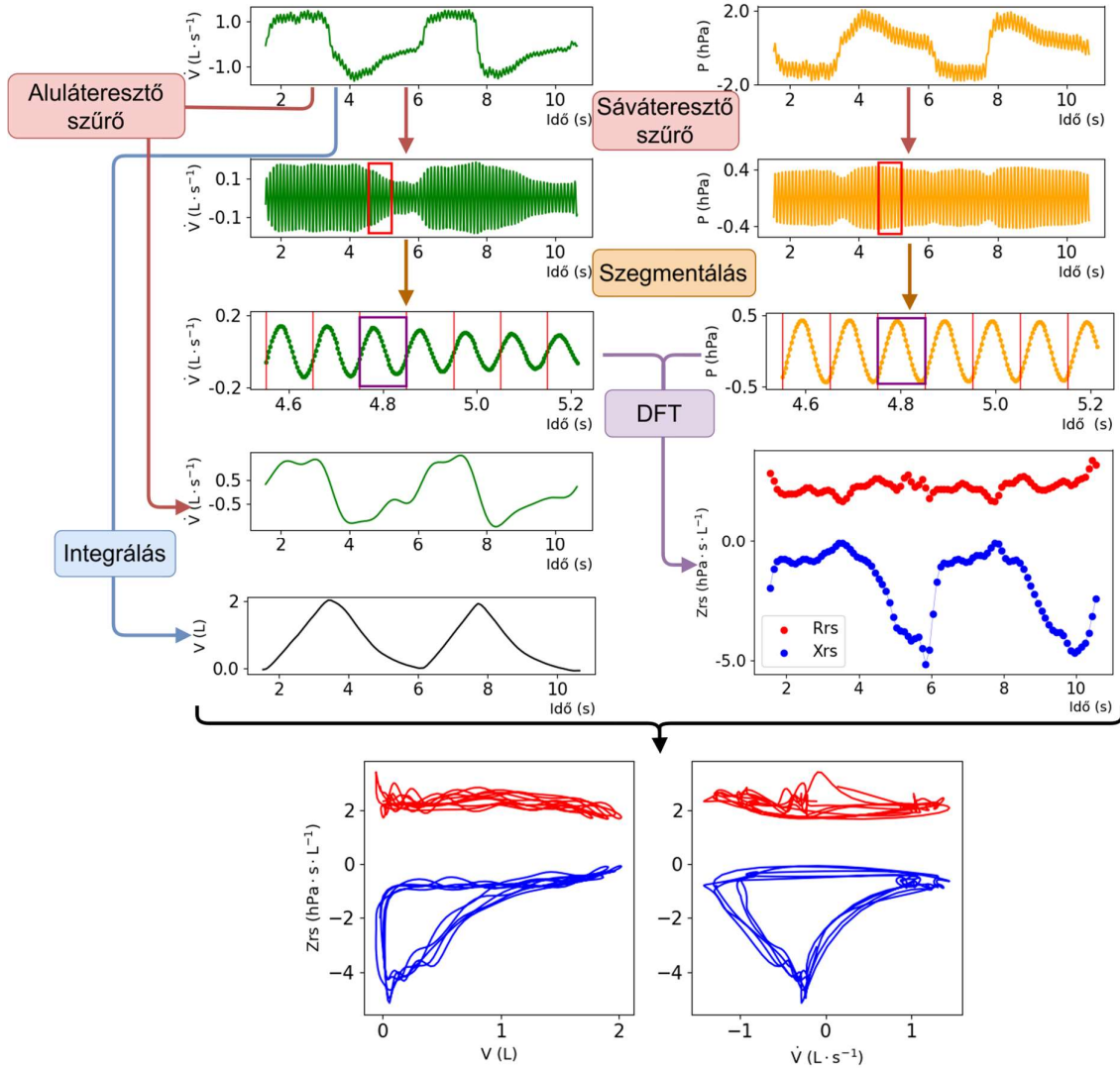
- **A:** Az áramlás- és nyomásmérési ponttól a baktériumszűrőig terjedő cső impedanciája (rezisztív valós rész és inertív képzetes rész)
- **B:** Az áramlás- és nyomásmérési ponttól a baktériumszűrőig a gázterek kompresszibilitása, lényegében egy kapacitív impedancia
- **C:** a szűrő rezisztív komponense
- **D:** a filterréteg és a légúti bemeneti pont között a csőszakasz koncentrált gázkompresszibilitása
- **E:** a filterréteg után, és a légúti bemeneti pont között a csőszakasz impedanciája
- **F:** a maszk impedanciája (amennyiben arcmaszkot használtak)

Ezeknek az előtét impedanciáknak a befolyása a terhelő impedancia nagyságának növekedésével egyre jelentősebb. Főként különböző mérési elrendezésekkel készült adatok összehasonlításánál fontos a korrekció elvégzése az eszközkomponensek leválasztására.

Az *inpx* fájlok integráltan tartalmazzák a kalibrációs adatokat, de a szövegfájlok esetén is a programom ad lehetőséget külső .cal típusú fájlok beolvasására, amelyeknek tartalmazzák a mérési frekvenciákhoz tartozó paramétereket.

3.2. Adatfeldolgozás

A beolvasott $P(t)$ és $\dot{V}(t)$ adatokból több lépésen keresztül állítom elő az impedanciát. Ezt a folyamatot mutatja be a 3.3 ábra, amelynek a fő pontjait részletezem.



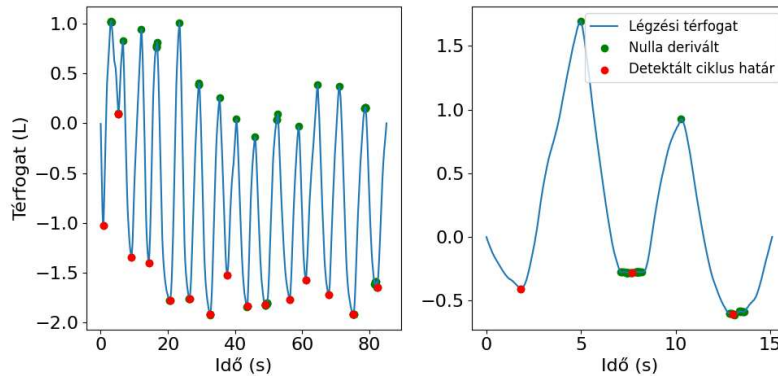
3.3 ábra: Az adatfeldolgozás lépései: a mért nyomás és áramlás adatokból az impedancia hurkok előállítása.

Az intra-breath technika lényege, hogy az impedancia változását folyamatosan kövessük, ellenben a hagyományos oszcillometriával (a teljes mért jel spektrális analízise). A mérések során használt 10 Hz-es mérőjel minimális időbeni felbontása 0,1 s, ami azt jelenti, hogy egy megfelelő csúszóablakkal, 0,1 s hosszú időszeltekre lehetséges az impedancia értékének meghatározása. Esetünkben a 256 Hz-es mintavételi frekvencia kedvezőtlen, ugyanis a 100 ms hosszú intervallumok nem egész számú mintából álló ablakokat eredményeznének. Emiatt a jeleket a feldolgozás kezdetén újramintavételezem 1 kHz-re. Így már a jelből a mérés során

használt 10 Hz-es impedancia komponens pontosan kinyerhető egy Fourier transzformáció segítségével. A short-time Fourier transzformáció (STFFT) segítségével tehát a teljes mért jelekből egy időfüggő impedanciagörbét állítok elő a 3.1. egyenlet alapján. Érdemes még megemlíteni, hogy számunkra csak a 10 Hz-es komponens és modulációs oldalsávjai a fontosak, minden más frekvenciától megszabadulhatunk. Ezért az FFT számolás megkezdése előtt a jelet egy sáváteresztő digitális szűrővel (8-12 Hz) megsűrösöm, mivel szélesebb frekvencia tartomány esetén a légzési frekvencia felharmonikusai, valamint egyéb biológiai (pl.: keringési eredetű) jelek elfedik a keresett dinamikát.

Ezután előállítom a légzési térfogat értékét, amit bár a rendszerek nem mérnek, az áramlási jel numerikus integráljaként elő lehet állítani. A térfogat előállításához először ki kell vonni az áramlás értékekből az átlagukat, mivel a mérések az esetek többségében tartalmaznak egy trendet, ami az idő elteltével eltolja a térfogatértékeket (integrális hiba). A térfogatot persze csak becsülni tudjuk, mert mint minden esetben, itt is ismerni kellene a kezdőfeltételt, amelytől indulva az integrálást megkezdjük. Ezt a kezdőfeltételt úgy tudom meghatározni, hogy a mérés elemzését egy belégzés kezdetekor indítom és egy kilégzés végén hagyom abba. Ebben az esetben csak teljes légvételeket vizsgálok, és feltételezem, hogy minden mérés a kiindulási térfogathoz visszatérve ér véget. A légzési ciklusok határait ezen fűrészelre hasonlító térfogat grafikonok minimum pontjai határozzák meg, ahogy a 3.4. ábrán is látható.

E határpontok robosztus meghatározása nem triviális, mivel a jelen lévő alaptrend nagysága meghaladhatja az átlag légvételek nagyságát; esetleg nem várt viselkedések (például nyelés) új csúcsokat tud létrehozni a jelben, illetve a konkavítás gyakori változása megnehezíti a derivált vizsgálatát. Az ilyen jelenségek által okozott, alégvétel határt tévesen kijelölő minimum pontokat több iteráción keresztül távolítom el. Figyelembe veszem az egymásutáni minimumpontok értékét a rákövetkező maximum ponttal, hogy csak megfelelő távolságú pontok maradjanak meg. Vizsgálom a minimum pontok között eltelt időt, amit egy átlag légzési periódusidőhöz hasonlítok. Miután megtaláltam az első légvétel kezdetét és az utolsó kilégzés végét, ezen pontok alapján kivágom a jelből a releváns szakaszt és újra kiszámolom a térfogat értékeket.

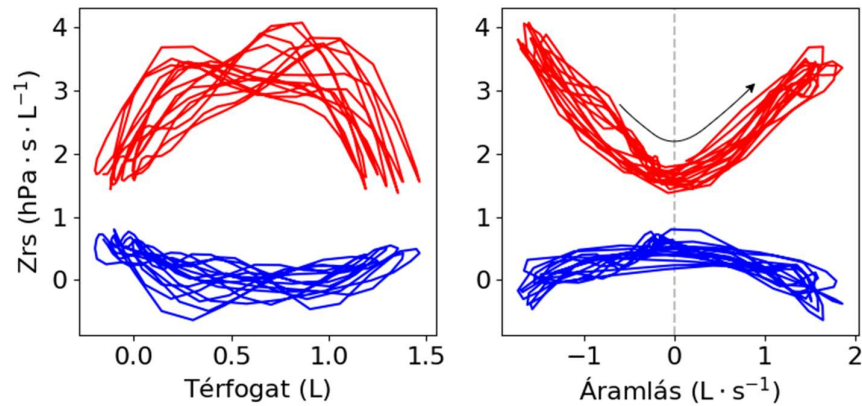


3.4. ábra: A légzési térfogat grafikonja. Zölddel a derivált előjelváltozásából előállított szélsőértékek, pirossal a szűrés után kiválasztott légzési ciklus határt definiáló minimum pontok láthatóak.

Utolsó lépésként az új adatsoron meg kell keresni a légzési ciklusok határait. Ezt az áramlás adatok alapján csinálom. Minden nullátmenet az áramlásban meghatároz egy belégzés vagy kilégzés végét. A mérési artefaktumokból adódóan szintén lehetnek hamis nullátmenetek, amelyek nem határoznak meg egy teljes légvételt. Ezeknek a szűrését hasonlóan végzem, mint a térfogatjel esetén. A nulla áramlású pontokban az impedancia értékek nagy jelentőséggel rendelkeznek, fontos orvosi mutatók meghatározására használatosak, ezért a minél pontosabb érték meghatározásához lineáris interpolációt alkalmazok minden nullátmenet két pontjára. Többek között így megkapom a belégzés végi rezisztenciát (ReI), és reaktanciát (XeI), a kilégzés végi rezisztenciát (ReE) és reaktanciát (XeE) valamint a megfelelő tidális különbségeket: $\Delta R = ReE - ReI$ és $\Delta X = XeE - XeI$.

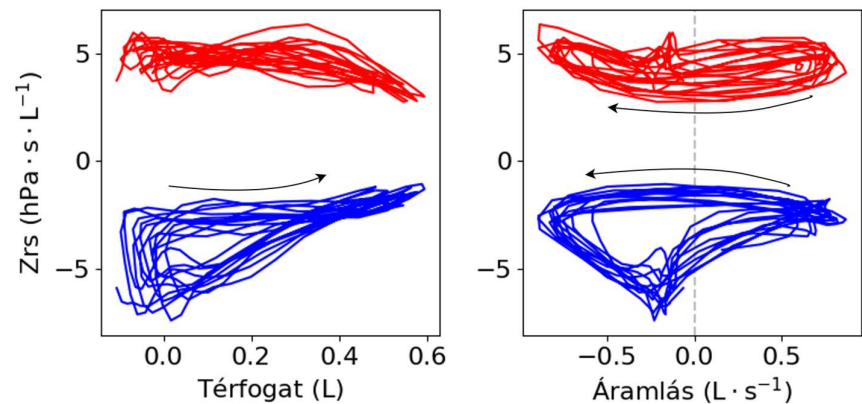
3.3. Impedancia hurkok

Az adatbeolvasásból és -feldolgozásból megkaptam a 10 Hz-es impedancia idősorait, ezt elemezni önmagában nehéz, ezért elterjedt az intra-breath mérések esetén az impedancia hurkok használata. Amennyiben adott időpontban az impedanciaértéket a légzési térfogat vagy áramlás függvényében ábrázoljuk, akkor észrevehető az egy légzési ciklus szakaszainak ismétlődése. Szemléletesen látszanak a nulla áramláshoz tartozó pontok, követhetők a belégzési és kilégzési útvonalak és megítélhető az egyes ciklusok változékonyságának mértéke is. Ez az ábrázolásmód előnyös, mivel betekintést nyújt a légzőrendszer dinamikus tulajdonságaiba, beleértve az áramlás- és a térfogatfüggésből adódó nemlinearitásokat is.



3.5. ábra: Egészséges alany impedancia hurkai (kék: Xrs és piros: Rrs).

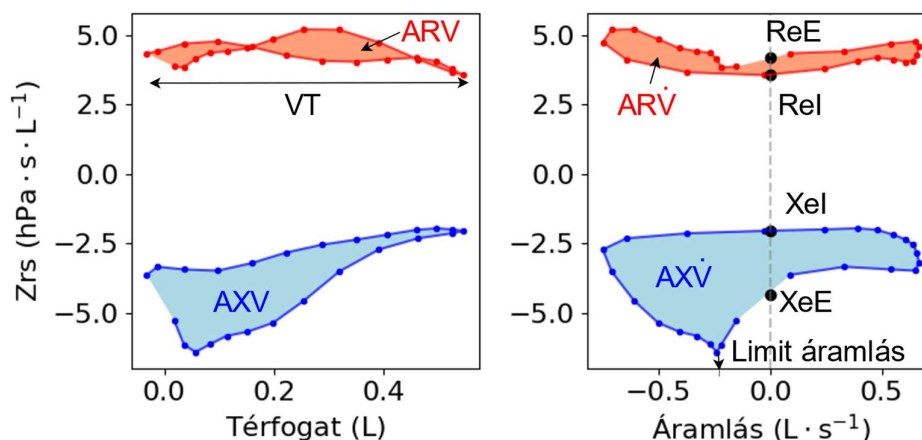
A 3.5. ábrán egy egészséges személy mérési adataiból készített hurkok láthatóak, irányítottsággal együtt. Megfigyelhető, hogy az Rrs a belégzési és főleg a kilégzési áramlás növekedésével megnő, ami a felső légutak nemlinearitására utal. Az Xrs tükrözi ezt a viselkedést. A 3.6. ábra ezzel szemben egy COPD-ben szenvedő páciens mérési adatait illusztrálja, bemutatva a tEFL jellegzetes, elsősorban az Xrs-ben megnyilvánuló mintázatát.



3.6. ábra: COPD-ben szenvedő alany impedancia hurkai (kék: Xrs és piros: Rrs).

Az áramláskorlátozással rendelkező betegek esetén a belégzési és kilégzési csúcsáramlás között az Xrs nem sokat változik. A kilégzési csúcsáram elérése után azonban Xrs eltér a korábbi pályától és elkezd drámaian csökkenni. Egy alacsony áramlásértéknél (áramláskorlát) stagnálva lassan eléri a minimumát, egy elhúzódó kilégzésvégi fázis részeként. Eközben az Rrs értékeinél nem figyelhető meg jelentős változás. A kilégzés végén elkezdődik a belégzési aktivitás, ennek következtében az áramláskorlátozás feloldódik (miközben a nettóáramlás még negatív értékű), ezután az Xrs visszatér normális tartományba.

A Zrs hurkok használata nem csak a folyamatok szemléltetésére hasznos, sok diagnosztikailag hasznos mérőszámot is vizualizál. Ezekből mutatok példákat a 3.7. ábrán. Itt a 3.2. alfejezetben említett interpolált ReI, XeI, ReE és XeE mellett ábrázolom a légzési térfogatot (VT), a hurkok területét, amely jelen esetben egy előjeles szám, mivel függ a hurok irányítottságától (az óramutató járásával ellentétes bejárási irány negatív területet eredményez). A tEFL-lel rendelkező alanyok esetén lehet definiálni a limit áramlást, amely az Xrs minimum helye az áramlás tengelyen, ez általában egy -0.2 L/s körüli érték. Későbbiekben ezek az értékek a gépi tanuló modell tanításánál jellemzőként is szerepelnek.



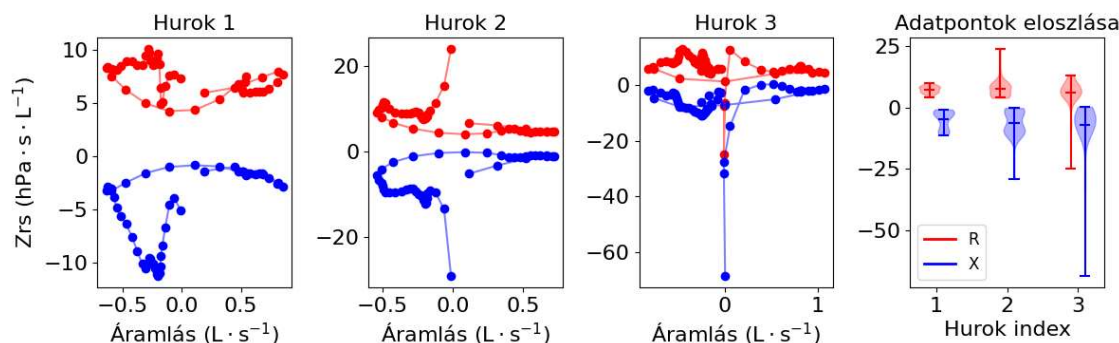
3.7. ábra: Hurokleírók: VT: légzési térfogat, ReE: kilégzés végi rezisztencia, ReI: belégzés végi rezisztencia, XeE: kilégzés végi reaktancia, XeI: belégzés végi reaktancia, ARV, AXV, AR $\dot{V}$, AX $\dot{V}$: a hurkokhoz tartozó előjeles terület.

3.4. Kiugró értékek felismerése (*outlier detection*)

A mérések során a nyugodt, egyenletes légzés fenntartása nem mindig sikerül. Vannak esetek, amikor a páciens sűrűbb és kisebb légvételeket vesz, máskor lassabbakat és mélyebbeket. Ilyen esetekben a kétféle légzési mód során más hurokmintázat léphet fel, mivel egy kisebb légvétel nem váltja ki a térfogatnövekedéssel járó nemlinearitásokat. A légzési adatokban olykor szabálytalanságok is megmutatkozhatnak, mint például a nyelés, csuklás, hangképzés vagy a levegő szivárgása. Mivel gyakran egy mérés tartozik egy pácienshez, így ezen jelenségeket, amelyek akár a légzésfunkciós mutatókat is befolyásolhatják, fel kell ismerni és eltávolítani. Ebben a részben a hurkok tisztítását mutatom be, azaz hogyan lehet közülük kizárni azokat, amelyek artefaktumokat tartalmaznak. Így egy homogén, megbízható részhalmazt hozok létre az elemzéshez. Ez segíteni fog a megfelelő, reprezentatív hurok létrehozásában, amelyek önmagukban képesek a páciens légzőrendszeri dinamikáját megjeleníteni. A hibás adatok válogatása sokban függ a rendelkezésre álló hurkok mennyiségétől, általános szabályként

megszabtam, hogy maximum a teljes darabszám egyharmada szűrhető ki és minimum három darab huroknak maradnia kell a szűrt adatban. A szakirodalomban ilyen jellegű adat tisztításra nem találtam példát.

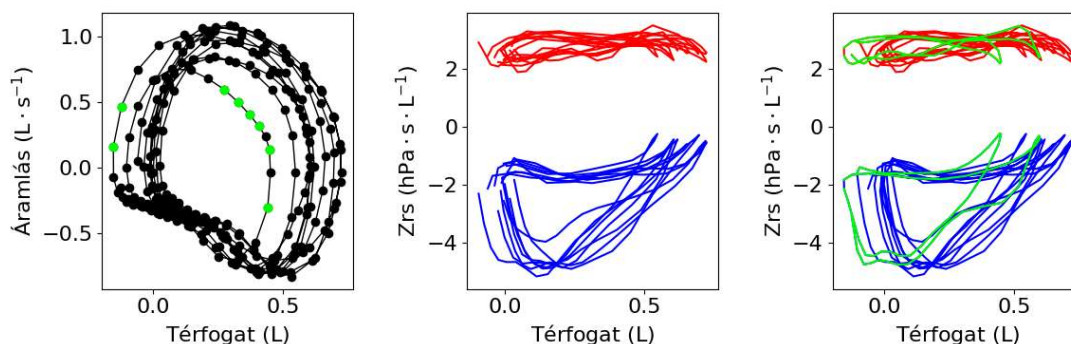
A mérési adatokban előforduló hibákat három csoportba soroltam be. Az első csoportba azok a mérések tartoznak, amelyekben egy-egy hibás pont vagy csúcs jelenik meg. Ez azt jelenti, hogy az ismétlődő alapmintázat ellenére található egy kis pontcsoportból álló csúcs, amely akár nagyságrendileg is eltér a mérés többi adatától. Ebben az esetben az egyéni légzési ciklusok Rrs és Xrs adatainak a szórását összehasonlítom az egyéni szórások átlagával. A kritikus eltérés arányát az egészséges alanyok mérési adatai alapján állítottam be, mivel náluk a legkisebb az Rrs és Xrs csúcstól-csúcsig változása, így arányaiban náluk fordul elő a legnagyobb eltérés a szórásban helyes mérések esetén. A 3.8. ábra bemutat egy példát hibás csúcs detektálására a szórás alapján. A mérés során mindössze három hurok keletkezett, amelyekben viszont sejteni lehet, hogy kettő hibás. Az y tengely értékeire figyelve észrevehető a szórásokban is a nagyságrendi különbség, illetve a hurok ábrákon megfigyelhető, hogy a letörések hirtelen keletkeznek, kevés pontból állnak.



3.8. ábra: Kiugró értékkel rendelkező mérés impedancia hurkai és a hurkok pontjainak eloszlása.

A második csoportba a légvételek nagyfokú változásai tartoznak, ahol szükséges a nagyon kicsi és túl nagy légvételek kiszűrése. Ehhez a V- $\dot{V}$ diagramot elemeztem, ami csak az áramlás és a térfogat összefüggését és tartományát mutatja. Itt az átlagtól eltérő pályák azonosítása a célom, ehhez a DBSAN klaszterező algoritmust használtam [15], mivel ez alkalmas amorf klaszterek létrehozására és definiál kívülálló pontokat is. A módszer lényege, hogy a klasztereket a sűrű pontok köré hozza létre, ezek olyan pontok, amelyek körül epsilon távolságban van minimum N darab pont. A klaszterhez ezután hozzáadja a sűrű pont szomszédait, majd ezen pontok szomszédait is, így növelve a klaszter méretét. A kimaradt pontok tehát távol esnek az átlag pályától, így még a pálya közepén átszelő szakaszok

detektálása is lehetséges. Az algoritmust az összes hurok pontjára lefuttattam, majd minden hurokhoz beazonosítom hány pontja lett kívülálló, ez alapján döntök. Két paraméter beállítása volt fontos, az epsilon környezet, amely a terjedés mértékét határozza meg, illetve a kívülálló pontok mennyisége egy hurokban, ami a hurkot átlagtól eltérőnek azonosítja. Az epsilon beállítása a mögöttes jelentése alapján történt. Mivel az áramlás és a térfogat is hasonló értéktartományban mozog, így a 0,1 epsilon jelentése 0,1 literrel nagyobb térfogat elérése vagy 0,1 L/s-nál nagyobb áramlás elérése. Ezen megfontolás alapján választottam ki a 0,1-et, amely a térfogat hurkok esetén már jelentős pályaelterést okoz. Ezt demonstrálja a 3.9. ábra. Ahol az első panelen található az áramlás – térfogat diagram pontjai. Zölddel vannak jelölve a DBSCAN által kiválasztott kívülálló pontok. Az utolsó panel az összes mérést mutatja, zölddel jelzi a kívülállónak nyilvánított pontokat. A középső panel a szűrt kimenetet ábrázolja, érdemes megfigyelni, hogy az eltérés a térfogat tengely mentén volt és így a térfogat hurkokban szabad szemmel is láthatóak ezen abnormális pályák. A jól megválasztott epsilon után a pontok számára vonatkozó határérték megválasztása egyszerűen következett, mivel már kettő pont esetén is nem véletlen találta meg az algoritmus az adott pályát.



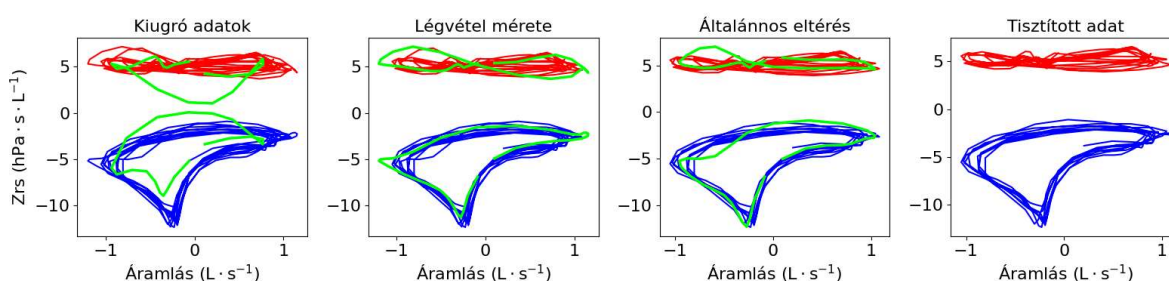
3.9. ábra: Hurok szelekció a légvétel mérete alapján: Az első panel a légzés közbeni áramlás változást ábrázolja térfogat függvényében, ezen zölddel vannak kijelölve a DBSCAN által kívülállónak ítélt pontok. A második panel a már szűrt impedancia hurkokat jeleníti meg szokásos módon (kék: Xrs és piros: Rrs). A harmadik panelen a detektált kívülálló hurkok zöld színűek.

A harmadik módszeremben általánosan kerestem az átlagos viselkedéstől eltérő hurkokat, melyek az eddigi kategóriákba nem illenek bele, mégis más pályát írnak le, mint a többi hurok. Itt az előző esethez hasonló megközelítést alkalmaztam, azonban itt a *LocalOutlierFactor* eljárást alkalmaztam, amely az adatpontokhoz egy lokális sűrűséget vizsgálja [15]. A kívülálló pontokat most az R-V, X-V, R- $\dot{V}$ és X- $\dot{V}$ hurkokban kerestem. Miután mindegyikben azonosítottam a hibás adatpontokat, és minden hurokhoz hozzárendeltem a hibás pontok számát, utána páronként összevetettem az R-V és R- $\dot{V}$, illetve

az X-V és X- $\dot{V}$ hurkok esetén talált pontok számát, olyan hurkokat keresve, amelyek mindkét típusú impedancia hurokban más pályát járnak be, mint a többség.

Zavaros légzés esetén jelentősen megnövekedhet a hurkok változatossága, ami problémát jelent, ha csak egy felvétel készült. Jobb esetben ezt a mérést végző operátor észleli, és megismétli a mérést, de ez mindig történik meg. Ilyen körülmények között a hiba szűrés a hurkok többségében jelezhet, ezért úgy döntöttem, hogy a rendelkezésre álló hurkok közül maximálisan csak a felük választható ki. Amennyiben a talált kívülálló száma meghaladja ezt a küszöböt, fontossági sorrendet alakítok ki, ami alapján elvégezhetem a szűrést. Ezt a fontossági sorrendet az Rrs és Xrs adatsorok korrelációs összehasonlításával állítom fel, így a sokaságtól legjobban eltérő hurkokat szűröm ki.

A 3.10. ábra bemutat egy mérést, amely mindhárom típusú hibára mutat példát. A szűrés első lépésében az Rrs eloszlása alapján azonosít egy eltérő szórású hurkot, a második lépésben olyan hurkot talál, amely nagyobb maximális áramlás értékeket tartalmaz a többinél, ezzel a nagyobb légvételre utalva. A harmadik lépésben pedig olyan hurkot detektál, ami egy szakaszon szemmel láthatóan eltér a többitől. Mivel tíz hurokból három volt rossz, így ezek a kívülálló hurkok törlésre kerülnek.



3.10. ábra: A kívülállók szűrési folyamatának szakaszai: 1. hibás pontokat tartalmazó ciklusok azonosítása, 2. a légvétel méretével kapcsolatos eltérések kiszűrése, 3. eltérő alakú hurkok felismerése. A detektált kívülálló hurkok zölddel vannak szemléltetve, az impedancia hurok szokásos módon vannak megjelenítve (kék: Xrs és piros: Rrs).

4. OSZTÁLYOZÁS

A célom itt egy megfelelő osztályozót tanítani tEFL mintázatok detektálására, amelyet majd fel tudok használni az OHS-ben szenvedők mérési anyagának kiértékelésére. Az alkalmazási feladat a tEFL hiányának, illetve meglétének megítélése e mérési anyagban, melyhez *tanítóanyagként* egyrészt egészséges (kontroll) személyek méréseit, másrészt a tEFL jelenséget tipikusan (bár nem kizárólagosan) felmutató COPD-s páciensek méréseit használom fel. Ez egy felügyelt tanítási feladat, ám ahhoz, hogy a tEFL jelenléte mellett a mértékére is tudjak becslést

adni, olyan osztályozókat alkalmaztam, amelyek képesek folytonos kimenetet adni és ennek a jelentése összefüggésbe hozható az alapfeladattal.

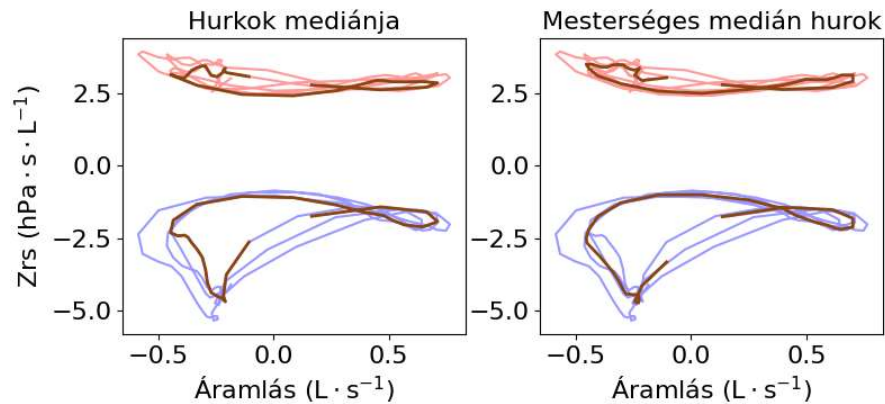
A kontroll mérési anyag a torontói University Health Network légzésfunkciós laboratóriumából származik. Itt a vizsgált személyek akár több különböző, követéses vizsgálatban is részt vettek, emellett minden vizsgálat során legalább három, egyenként 30 s hosszú intra-breath oszcillometriás felvételt készítettek. A tanulmányom során minden páciensről egy mérési alkalom adatait használtam fel, abból viszont az összes adatot, mivel ez ugyanazon időpont béli állapot leírása. A COPD-s beteganyag egy montreáli pulmonológiai klinika méréseiből származik, ahol viszont egy vizsgálatban egyetlen intra-breath mérést végeztek. Az adathalmazok származási különbségei miatt egy beteghez jelentősen eltérő számú hurok tartozhat, ennek a kezelését a 4.1. alfejezetben fejtem ki. Mindkét helyen egy kereskedelmi forgalomban lévő oszcillométerben (tremoflo®, Thorasys Ltd., Montreal, Kanada) a csoportunk által kidolgozott kutatási célú intra-breath modalitást használták.

4.1. Jellemzők kinyerése

A vizsgálódásom alapja az impedancia hurok, ennek alakjára és nevezetes pontjaira definiáltam a tEFL kimutatására használható jellemzőket. Problémaként megjelent, hogy a páciensek egyes mérési alkalmaihoz változó mennyiségű hurokok tartoztak. Szükség van egy módszerre, amivel egységesen lehet kezelni az adathalmazt, így azt is, akinek három légvétele került rögzítésre, és azt is, akinél ez a szám harminc fölött van. Ez azért lehetséges, mert a mérés hossza nem standardizált, a páciensek légzési frekvenciája változó és a készült felvételek száma is eltérhet. A probléma áthidalására négy opciót dolgoztam ki, a módszer teszteléskor paraméteresen változtatható.

- **Átlag értékek:** Ebben az esetben az összes légvételre kiszámolom az összes jellemzőt és a végén jellemzőnként veszem az átlagértéket. Ez a módszer megfelelően jellemzi a sokaságot, de kevés hurok esetén érzékeny a kiugró értékekre.
- **Medián értékek:** Az előző módszer, csak medián érték használatával. Ez a módszer kevésbé érzékeny a kiugró adatokra.
- **Hurkok mediánja:** Egy kiválasztott jellemző alapján (ami általában az AXV) sorba rendezett hurkok közül a medián hurok. Erre a kiválasztott egyedi hurokra számolom ki a jellemzőket. Hatékonyan jellemzi a sokaságot, ha a hurkok nagyfokú hasonlóságot mutatnak. Amennyiben nagy a variabilitás, akkor egyes szakaszokon a többi hurok pályájától eltérhet.

- **Mesterséges medián hurok:** A hurkokat újramintavételezem, annak érdekében, hogy mindegyik ugyanannyi pontot tartalmazzon és minden adatsor esetén pontonként veszem a medián értéket. Ezzel olyan hurok keletkezik, amely a mérésben nem létezik, mégis minden szakasza az átlag viselkedést mutatja.



4.1. ábra: Hurkok mediánja és mesterséges medián összehasonlítása

A 4.1. ábra két féle reprezentatív hurok kiválasztási módot hasonlítja össze, első esetben biztos, hogy a hurok valós és megtörtént viselkedést mutat, míg a második eset jobban mutatja a hurkok sokaságának pályáját. Ahogy látható, a kilégzés végi Xrs pálya (a csúcsáramlástól a limit áramlásig tartó szakasz) jelentősen különbözik, ami az XeE értékében meg fog mutatkozni. A további elemzés érdekében bemutatom a hurkokra definiált jellemzőket, amelyeket az előbb számított négy módszerrel lehet egy teljes mérési alkalomra értelmezni az egyéni hurkok helyett. A jellemzők között szerepelnek az 3.3 fejezetben felsorolt értékek, emellett igyekeztem többféle megközelítéssel is különböző információt kinyerni, a jellemzők listája a 4.1. táblázatban szerepel.

TI	Belégzés hossza
TE	Kilégzés hossza
Ttot	Teljes légvétel hossza
TE/Ttot	A kilégzés és a teljes légvétel hosszának aránya
fbr	Légzés frekvenciája
VT	Légzési térfogat
ReE	Kilégzés végén a Rrs értéke
ReI	Belégzés végén a Rrs értéke
XeE	Kilégzés végén a Xrs értéke
XeI	Belégzés végén a Xrs értéke
ΔR	ReI-ReE
ΔX	XeI-XeE
AVR *	Hurok területe
PVR *	Hurok kerülete
asp_VR *	<i>aspect ratio</i> : befoglaló négyzet hosszának és szélességének aránya
circ_VR *	<i>circularity</i> : $4 \cdot AVR / PVR^2$
frate_VR *	<i>filling ratio</i> : A hurokterület aránya a befoglaló négyszög területéhez a
$\dot{V}_{MINX}$	Minimum Xrs helye az áramlás tengely mentén
meanRE	Kilégzés során a Rrs értékek átlaga
meanRI	Kilégzés során a Rrs értékek átlaga
meanXE	Kilégzés során a Xrs értékek átlaga
meanXI	Kilégzés során a Xrs értékek átlaga
statisztikai	R és X értékekre: minimum, maximum, átlag, szórás
E_A $\dot{V}$ X	A reaktancia-áramlás- hurokban a kilégzési ág területének aránya a teljes területhez

4.1. táblázat: Jellemzők jelölése és jelentése. *-gal jelölt cellák esetén R-V hurok mellett az X-V, R- $\dot{V}$ és X- $\dot{V}$ hurkokra is vonatkozik

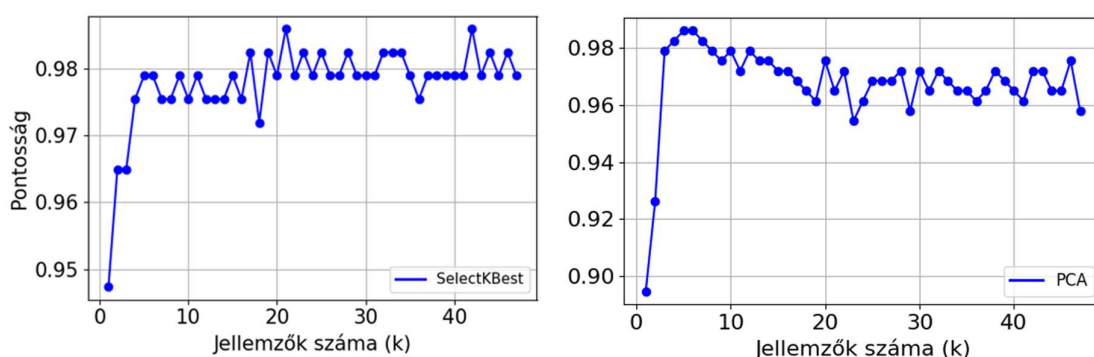
Egy *baseline* modellt választottam ki, amely alapján kiértékeltem az összes jellemző teljesítményét, majd különféle jellemzőkiválasztási és előfeldolgozási módszereket alkalmaztam. A *baseline* modellként a *RandomForest* osztályozót használtam. Az eredmények összehasonlítását az f1-score alapján végeztem, mivel ez a pontosság és érzékenység harmonikus átlaga, és összetettebb képet ad a modell teljesítményéről, mivel a két csoport egyedszáma eltérő. A vizsgálataim során a scikit-learn Python csomag [16] függvényeit és modelljeit használtam.

Elsőként a jellemző számítási módokat vizsgáltam meg, hogy az előbb említett négy módszer (átlag értékek, medián értékek, hurokok mediánja, mesterséges medián hurok) kimenete mennyire tér el egymástól. A vizsgálathoz a tanítóhalmaz minden egyedére kiszámoltam az összes jellemzőt külön-külön minden módszerrel, majd jellemzőnként elemeztem a módszerek kimenetei közti eltérést. Az elemzés során az adott jellemző egy-egy módszerrel kiszámított értéksorozatát hasonlítottam össze a többi számítási módszerrel, korreláció alapján; így 41 darab, négyszer négyes korrelációs mátrixot kaptam (M.1. ábra). Az ábráról leolvasható, hogy a *filling ratio* típusú jellemzők kivételével az értékek átlaga, mediánja és a mesterséges medián hurok technikák kimenetei 0.9 fölötti egyezést mutatnak. Ezek alapján a mesterséges mediánhurok használatát választottam, mivel szakaszai egy általános pályát követnek, és a kiszámolt jellemzők egy hurok valós adatait adják vissza, megőrizve az egymásból következő jellemzők közötti viszonyokat.

A jellemzők kiválasztási technikáit ezt követően a *baseline* modell segítségével értékelttem ki. Első esetben minden jellemzőt felhasználtam előfeldolgozás nélkül, ezzel 95%-ot értem el. A második esetben a jellemzők normalizálása után eltávolítottam a nulla varianciájú jellemzőket, majd a *SelectKBest* [17] módszerrel optimalizáltam a jellemzők számát. Az *f-classif scoring* [18] függvényt választottam, mivel ez alkalmas folytonos jellemzők kezelésére diszkrét osztályozási feladatok esetén, az ANOVA teszt segítségével vizsgálja a jellemzők és az osztályok közötti szignifikáns különbséget és varianciát. A jellemzők optimális számának meghatározására rács keresést (*grid search*) alkalmaztam 5-szörös keresztvalidációval, amely során a teszhalmaz 5 részre van osztva és az eredmény 5-ször van kiértékelve különböző részhalmazokra, így a validációs halmaz folyamatosan változik. A keresés eredményét a 4.2. ábra szemlélteti. Az ábrából látható, hogy a legjobb jellemző önmagában 96,5%-os pontosságot biztosít a tanítóhalmazon (amely jelenleg az adathalmaz 50%-át teszi ki), azonban a legmagasabb pontosság eléréséhez 21 jellemzőre van szükség. Az eredmények alapján kevés jellemző is elegendő a jelenség leírásához, így dimenziócsökkentéssel folytattam a vizsgálataimat. Alkalmaztam a *principal component analysis* (PCA [19]) és *linear discriminant analysis* (LDA [20]) módszereket. Az LDA alkalmazása indokolt, ugyanis címkézett adathalmazon dolgozom, és reálisnak tartom, hogy egyetlen jellemző is elegendő lehet - mivel két osztály esetén az LDA egy dimenzióra vetít - az adathalmaz leírásához, amit a 4.2-es ábrán látható *SelectKBest* esetén egy jellemzővel elért eredmény is alátámaszt. A PCA-t is kipróbáltam, abban az esetben, ha egy dimenzióban mégsem lenne elegendő a maximális pontosság eléréséhez. A *baseline* modell tesztelésekor az LDA 96% pontosságot ért el. A PCA dimenziócsökkentő módszer, megkeresi azokat az irányokat, amelyek mentén az adatok

legjobban szóródnak (főkomponensek) és új koordináta rendszert hoz ez alapján létre. Ebben az esetben is a megmaradt jellemzők számát a *baseline* modell segítségével állapítottam meg, mint a *SelectKBest* esetében. A 4.2. ábra jobb paneljén látható, hogy kevesebb jellemzővel, akár már négy komponenssel ugyanazt a pontosságot tudja elérni, mint a *SelectKBest* 21 jellemző esetén.

A négy előfeldolgozási lehetőséggel elért eredményt a 4.2. táblázat hasonlítja össze, ahol az eredmények a két osztályon elért mutatók súlyozott átlaga; a súly az egyes csoportok mintaszáma. Az f1-score alapján a PCA érte el a legjobb eredményt. A jellemzők tehát olyan jól leírják az adathalmazt, hogy a kiválasztási módszerek fontossága elhalványul. Minthogy a PCA kevés jellemzővel nagyon jó eredményt ért el, innentől ezt alkalmazom.



4.2. ábra: SelectKBest és PCA esetén az ideális paraméterszám vizsgálata

Jellemzőkiválasztás típusa	Pontosság	Recall	f1-score	Mintaszám
Nincs	0.96	0.95	0.95	151
SelectKBest	0.97	0.97	0.97	151
PCA	0.98	0.98	0.98	151
LDA	0.96	0.96	0.96	151

4.2. táblázat: A különböző jellemző kiválasztási technikákkal elért eredmény a *baseline* modell esetében

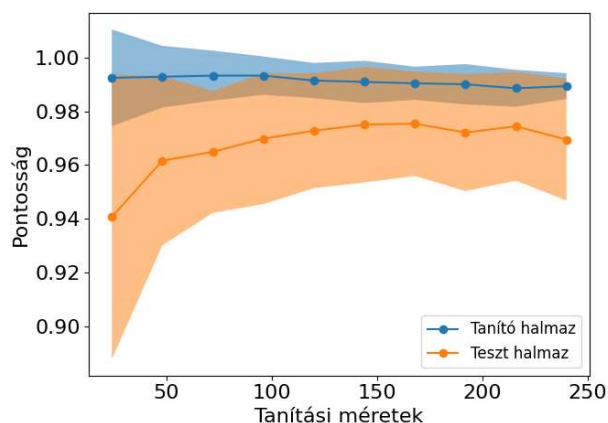
4.2. Osztályozó algoritmusok

A vizsgálataim során hét különböző modell teljesítményét hasonlítottam össze. Minden esetben a metaparaméterek optimalizálása rácskereséssel történt és a keresési folyamat egyes lépéseiben a legjobb kimenetet keresztvalidációval választottam ki. Az összehasonlítás alapjául az f1 mutatót (*f1-score*) választottam. A kiválasztott modellek mindegyikére igaz, hogy bináris osztályozást végez, azonban képes folytonos kimenetet generálni, ami fontos lesz az OHS-es alanyok méréseinek kiértékelésénél. Legtöbb együttes (*ensemble*) módszer alkalmas ilyen

kimenetet előállítására. Az alábbi hét algoritmust azért választottam, mert megfelelnek az előzetes feltételeimnek és széles spektrumot képviselnek az osztályozó algoritmusok terén. Az algoritmusokat szintén a scikit-learn könyvtár megvalósításában alkalmaztam, mivel egységes interfészt nyújt a vizsgálódásokhoz. Az alábbi modelleket alkalmaztam:

- Véletlen erdő - RandomForestClassifier (baseline): Több különálló döntési fa végezi el az osztályozási feladatot, a végén a különálló fák szavazata számít. Így a döntési fák szavazatainak átlaga megfelelő folytonos kimenet. [21]
- Gradiens növelő osztályozó - GradientBoostingClassifier: Egy olyan együttes módszer, amely egyszerűbb osztályozók szekvenciális alkalmazását jelenti, ahol minden új tag az előző hibáját csökkenti. Az osztályba tartozás mellett egy valószínűséget is ad minden osztályhoz, amely leírja mennyire biztos a döntésében. [22]
- Support Vector Machine - SVC: Olyan hipersíkokat keres, amelyek a különböző osztályok legközelebbi pontjai közötti különbséget maximalizálják. Különösen alkalmas magas dimenziójú adatok nemlineáris szétválasztására. A scikit-learn megvalósítás lehetővé teszi a döntési függvény értékének kinyerését, ez az adott adatpont és a döntési felület távolsága, értelmezhető az osztályba tartozás bizalom mértékeként, tehát minél nagyobb a távolság, annál távolabb vagyunk a döntési felülettől, az osztályozás annál biztosabb. Ebben az esetben a kimenet előjele adja meg az osztályt és az abszolútértéke adja meg azt, hogy a modell mennyire biztos az osztályozásban. [23]
- MultiLayer Perceptron classifier- MLPClassifier: Egy olyan neurális háló modell, amely a bemenetet több rejtett neuronrétegen át továbbítja, ezzel nemlineáris kapcsolatokat tanul meg. Itt a kimeneti rétegben *softmax* aktivációs függvényt alkalmazva a nyers kimenetből egy valószínűségi változó lesz, amely az osztályba tartozás valószínűségét adja meg. [24]
- K legközelebbi szomszéd - KNeighboursClassifier: Egy adatponthoz a K legközelebbi szomszédja alapján rendel osztály címkét. A folytonos kimenet ebben az esetben egyfajta súlyozott szavazás, ahol a K legközelebbi szomszéd szavazatát a távolságuk súlyozza, a legközelebbi kapja a legnagyobb súlyt. [25]
- ExtraTreesClassifier: A véletlen erdő algoritmus egy variánsa. Véletlen küszöbszinteket alkalmaz jellemző kiválasztásra és csomópontok felosztására. A folytonos kimenet szintén a döntési fák szavazatának átlaga. [26]
- LogisticRegression: Egy lineáris modell, amely egy logisztikai függvény segítségével jósolja meg az adott osztályba tartozás valószínűségét. [27]

A vizsgált modellekhez a beállított paramétereket, a keresési teret a legjobb eredménnyel a melléklet 1. táblázata tartalmazza. Első lépésként a *baseline* modell segítségével megvizsgáltam azt, hogy miként szeparáljam az adathalmazt. Ehhez előállítottam a tanulási görbét (4.5. ábra), a görbe előállításához is kereszt validációt alkalmaztam. A teszt halmazon elért valószínűség hamar tetőzik, a 301 adatpontból már 150 adatpont felhasználásával megfelelő eredményeket lehet elérni. A teszt halmaz mellett validációs halmazt is alkalmazok a hiperparaméterek beállítására, ennek 50 adatpontot hagyok és a maradék 100 pontból áll a teszhalmazom. Így összességében a halmaz kicsit több, mint felét használom tanítási célokra.



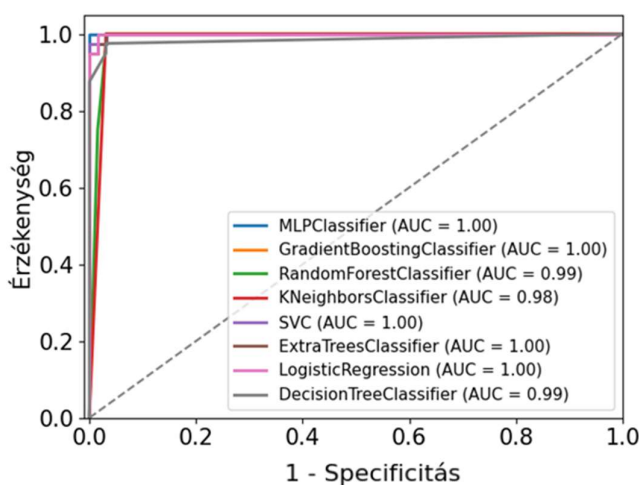
4.5. ábra: Tanuló görbe *baseline* modellel létrehozva

Végül minden osztályozó esetében megkerestem a legjobb paramétereket rácskereséssel a tanító halmazon keresztvalidációval, majd minden algoritmusból a legjobban teljesítőt kiértékeltem a teszt halmazra, ahol a kimeneteket a 4.3. táblázat mutatja be.

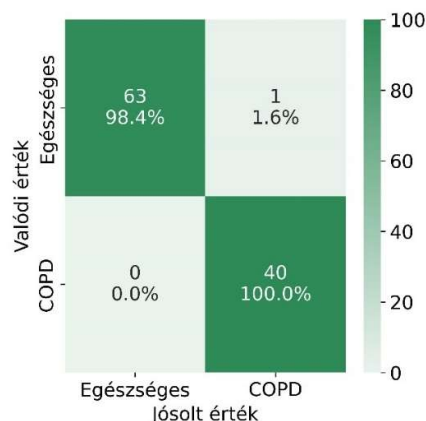
Modell neve	Osztály	Pontosság	Érzékenység	F1-Score	Mintaszám
SVC	Egészséges	0.98	0.97	0.98	64
	COPD	0.95	0.97	0.96	40
RandomForestClassifier	Egészséges	1	0.97	0.98	64
	COPD	0.95	1	0.98	40
LogisticRegression	Egészséges	1	0.97	0.98	64
	COPD	0.95	1	0.98	40
KNeighborsClassifier	Egészséges	1	0.97	0.98	64
	COPD	0.95	1	0.98	40
GradientBoostingClassifier	Egészséges	1	0.97	0.98	64
	COPD	0.95	1	0.98	40
ExtraTreesClassifier	Egészséges	1	0.97	0.98	64
	COPD	0.95	1	0.98	40
MLPClassifier	Egészséges	1	0.98	0.99	64
	COPD	0.98	1	0.99	40

4.3. táblázat: A modellekkel elért eredmények összehasonlítása

Az eredményeket osztályonként szerepeltetem, ezzel teret adva pontosabb elemzéseknek. A legjobb eredményt az *f1-score* alapján az *MLPClassifier* érte el. A modellek teljesítményének értékeléséhez és szemléltetéséhez *Receiver Operating Characteristic* (ROC) görbéket is alkalmaztam, ami a 4.6. ábrán látható. A ROC görbéről szintén leolvasható, hogy a legjobb eredményt az *MLPClassifier* érte el. Ennél a modellnél a kimeneti réteg softmax aktivációs függvényt használ, amely valószínűségi értékeket rendel az egyes osztályokhoz, jelezve, hogy a bemenet milyen valószínűséggel tartozik az adott osztályba. Ehhez a valószínűségi jelentéshez társítom a tEFL mértékének meghatározását. Amennyiben a modell kimenete 1, az azt jelenti, hogy a tEFL mintázata markánsan megjelenik az adott mérésben; 0 kimenet esetében pedig nem található erre utaló tényező. A küszöbszintet 0,5-nél húzom meg, mivel az osztályozó tanításánál ez jelentette az osztálycímkek közötti határt, ez magában hordozza azt az információt, hogy egy 0,5 körüli kimenet bizonytalan, tehát a tEFL gyengén jelen van vagy éppen kialakulóban van. A pontos határ meghatározása ezen adathalmaz alapján nehezített, mivel csak egyértelműen egészséges egyének és COPD-s betegek méréseit tartalmazza, így ennek a meghatározása a modellre van bízva. *MLPClassifier* esetén a teljesítmény részletesebb elemzésére tévesztési mátrixot is készítettem (4.7. ábra).



4.6. ábra: Paraméterezett modellek összehasonlítása



4.8. ábra: *MLPClassifier*: tévesztési mátrix a tesztalmazon

5. ALKALMAZÁS

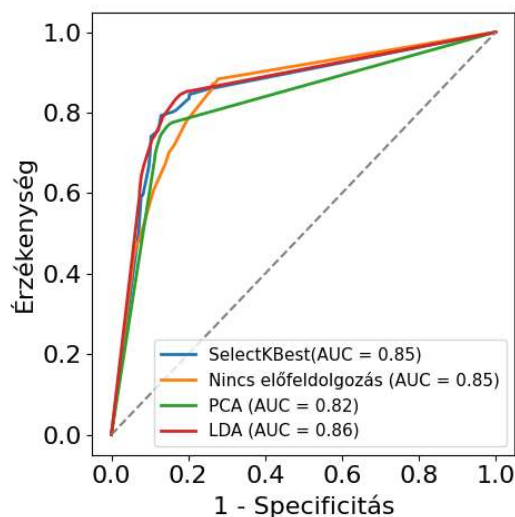
Ebben a fejezetben az előzőleg betanított modell kiértékelését végzem más adathalmazokon. Az első esetben egy nagy méretű címkézett adathalmazt használok, amelyben ILDBen (intersticiális tüdőbetegség) szenvedő alanyok 10 Hz-es intra-breath mérései vannak. Ezen betegséggel rendelkezők intra-breath mintázatai nagyfokú átfedésben vannak az egészséges páciensekével, így a teszt arra irányul, hogy ezen jellemző halmaz mennyire képes általánosan leírni az intra-breath hurok mintázatokat. A második esetben OHS-ben szenvedő alanyok 10 Hz-es méréseit értékelem ki, ezek mindössze 33 alany méréseiből származnak, és nincsenek címkézve. Az OHS egy tünete a tEFL (amelynek eddig a detektálását végeztem), amely csökkenthető CPAP segítségével. Ezt a folyamatot számszerűsítem a 4.2-es fejezet legjobb osztályozójával való kiértékeléssel.

5.1 ILD tesztanyag

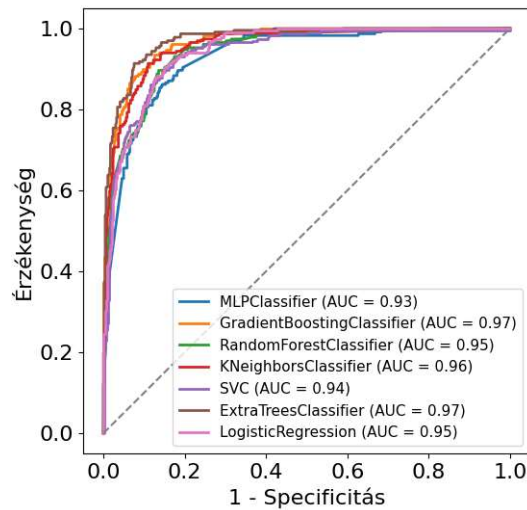
Az adathalmaz 384 alany méréseit tartalmazza, a mérések a torontói University Health Network légzésfunkciós laboratóriumából származnak. Az adathalmaz felépítése az egészséges anyag felépítésével megegyező: lehet több mérési alkalom is és minden alkalommal több egymás utáni mérési fájl. Itt minden páciens esetén egy mérési alkalom adatait használom fel és abból az összes fájlból kinyert hurkokat egy egységként kezelem. Ahhoz, hogy a tanítás és a kiértékelés során kiegyensúlyozott legyen a csoportok száma, az ILDBen adathalmazból véletlenszerűen választok az egészségesek számával megegyező adatot. Ezt követően a tanítás menete megegyezik a 4.2. alfejezetben szereplő folyamattal. Első lépésként a legmegfelelőbb jellemző szelekciós technikát választottam ki. Mivel az ILDBen és az egészséges alanyok hurokmintázata ránézésre azonos is lehet, ebben az osztályozási feladatban jobban számít minden jelfeldolgozási

döntés. A tanítás három osztállyal történt: egészséges, COPD, ILD. A ROC görbék előállításához több osztályos esetben *micro-average* technikát alkalmaztam. Ahogy az 5.1. ábrán látható, a legjobb eredményt a LDA érte el, tehát ezzel dolgoztam tovább.

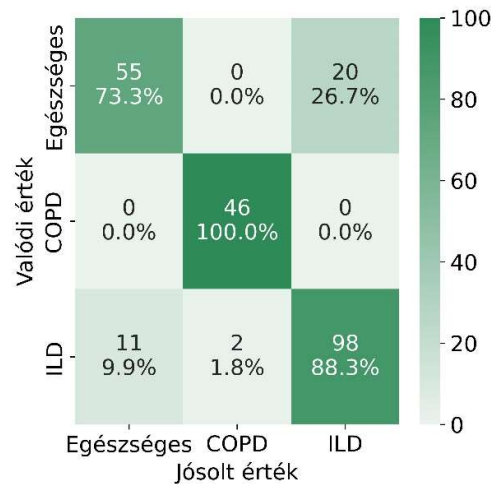
Ezt követően minden modell paramétereit rács keresés segítségével optimalizáltam és összehasonlítottam a teljesítményüket a teszt halmazon. Az eredményt az 5.2. ábra mutatja, ahonnan leolvasható, hogy jelen esetben a *ExtraTreesClassifier* teljesített legjobban. Az *ExtraTreesClassifier* tévesztési mátrixota látható az 5.3. ábrán. A félreosztályzások oka egészséges és ILD-s esetben a kettő csoport intra-breath hasonlósága, illetve az ILD diagnosztizálás nem egységessége okozzák, ezért az osztályozás itt bonyolultabb, mint a tEFL és nem tEFL esetben. Ez abban is megmutatkozik, hogy a megfelelő jellemző kiválasztási technika és jól paraméterezett modell választása érdemi javulást okozott. Összességében az 5.3. ábra mutatja, hogy a COPD-seket jól el tudja különíteni, tehát a modell más beteganyag bevezetése mellett is képes a tEFL hatékony felismerésére. Ezt alátámasztja az M.2. táblázat, amely a modellek részletes összehasonlítását végzi. Látható, hogy az *f1-score* COPD-s csoport esetén minden osztályozónál 92% fölötti. Az ILD-s páciensek, akik COPD kategóriába lettek sorolva, valójában tEFL-el rendelkező ILD-sek, ami ritka, de létező kondíció, az ILD-s adathalmazban 14 ilyen ismert eset van, így a besorolás megalapozott.



5.1. ábra: Jellemző szelekciós technikák összehasonlítása



5.2. ábra: ILD-s és egészségesek összehasonlítása különböző modellekkel



5.3. ábra: Tévesztési mátrix három osztály esetén *ExtraTreesClassifier* modellel

5.2 OHS beteganyag

A dolgozat elsődleges célja az OHS-es beteganyag kiértékelése és az egyes CPAP szinteken a tEFL jelenlétének számszerűsítése, amihez a 4.2. alfejezetben a legjobbnak talált *MLPClassifier* modellt alkalmaztam. Az adathalmaz származási helye a Semmelweis Egyetem Aneszteziológiai és Intenzív Terápiás Klinikája, az alkalmazott intra-breath oszcillometriás mérő és értékelő rendszert az SZTE TTIK Műszaki Informatika Tanszék kutatócsoportja dolgozta ki. Ellentétben a kereskedelembe megtalálható konvencionális eszközökkel (pl. a Tremoflo), ez az oszcillométer képes megemelt nyomásszint (gépi légzéstámogatás) mellett üzemelni [3]. Itt a kiértékelést a mérőszoftver segítségével minden páciens minden mérésénél kézzel végezték, amíg az általam készített program minden lépést automatizáltan végez el, illetve hibaszűrést is alkalmaz információ maximalizálás céljából. Harminchárom túlsúlyos

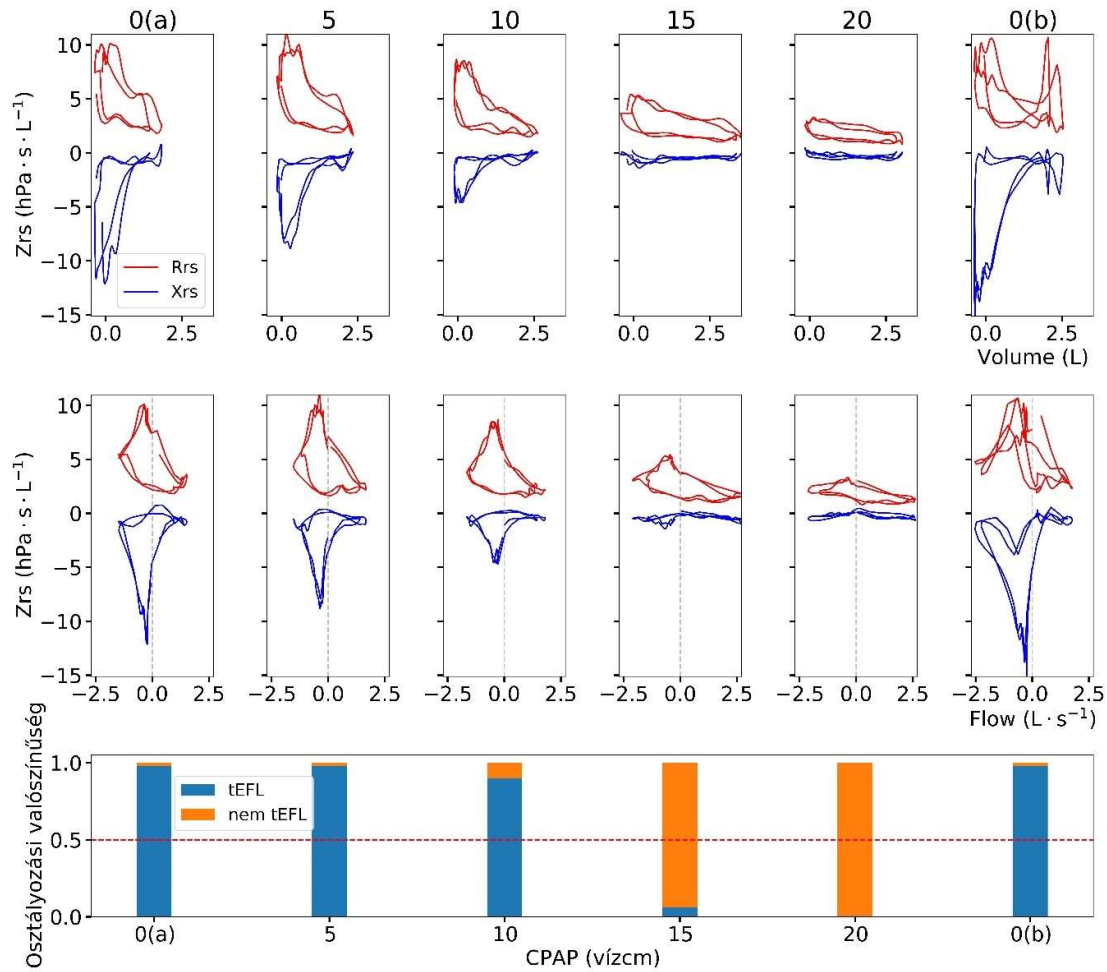
pácienszt vett részt a vizsgálatban, akik többségében az OHS súlyos stádiumában vannak. A demográfiai adatok releváns részét az 5.1-es táblázat írja le; fontos megfigyelni, hogy a BMI skála egészen 91,1-ig terjed, ez már extrém kóros elhízásnak számít. Az extrém testtömeg jelentős terhet ró a mellkasfalra, valamint a hasüreg révén nyomást gyakorol a rekeszizomra. Ez a tüdő kapacitásának csökkenéséhez vezet, miközben a légutak szűkülnek, elzáródási küszöbükhez közelednek, ami elősegíti a tEFL kialakulását. Ezen páciensek légzőszervi szempontból nem rendelkeznek más rendellenességgel, azonban náluk nem található kontroll anyag, hiszen a testtömegükkel jár a kórkép.

	Átlag	SD	Skála
Férfi	23 (65,7%)		
Kor (év)	56,3	8,6	39-78
BMI (kg/m²)	49,3	10,9	32,3-91,1
OHS súlyossági szint			
I	3 (8,6%)		
II	1 (2,9%)		
III	19 (54,3%)		
IV	12 (34,3%)		

5.1. táblázat: A vizsgálatban szereplő betegek legfontosabb demográfiai adatai

Gyermekek és felnőttek oszcillometriás méréseit ülő testhelyzetben végzik, de kimutatott, hogy fekvő helyzetben nagyobb nyomás nehezedik a tüdőre, így a légzőrendszer impedanciája nagyobb. Ez különösen igaz az OHS-es betegekre, akiknek emiatt a fekvő helyzetben az éjszakai lélegzés különösen nehezített. Ezen segít a CPAP terápia. A mérést fekvő testhelyzetben 5 különböző, egyre növekvő CPAP szinten végezték minden beteg esetén, a célom egyénenként minden CPAP szinthez egy számot rendelni, amely leírja az adott mérésben a tEFL súlyosságát, ezzel segítve a megfelelő nyomásszint kiválasztását. (Eddig a nyomást SI mértékegységben hPa-ban mérve számoltam minden esetben, viszont a légzéstámogató eszközökben a vízcm használata elterjedt, így a különböző CPAP szintek nyomását így adom meg, 1 vízcm = 0,98 hPa.)

A kiértékelés során betegenként minden CPAP szinten mért adatokból meghatároztam az impedancia hurkokat, ezekből kinyertem a 4.1-ben definiált jellemzőket és kiértékeltem a *MLPClassifier* segítségével. Egy mérés kiértékelésének eredményét mutatja az 5.4. ábra.

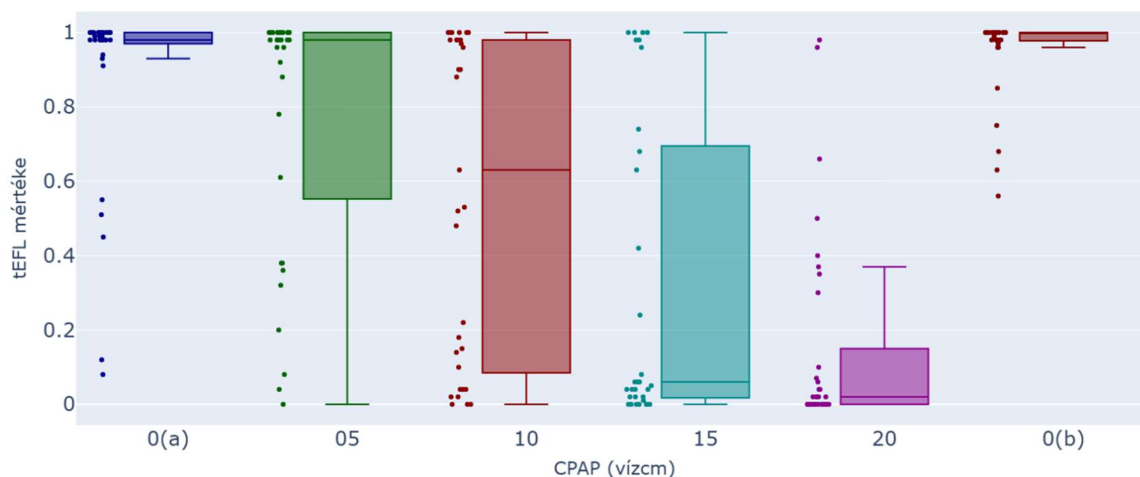


5.4. ábra: Intra-breath hurkok és az osztályozási eredmények minden CPAP szinten

Ahogy látható, CPAP nélkül a kilégzés végén súlyos áramláskorlátozás lép fel, mely az Xrs leesésével társul: a minimum Xrs értéke $-10 \text{ hPa} \cdot \text{s} / \text{L}$ alatt van. A CPAP emelésével a minimum Xrs érték egyre feljebb kerül, végül 15 vízcmm-nél a hurokterület drámaian lecsökken és szinte eltűnik a tEFL mintázata. A modell döntése a mérések alatt látható, ahogy megítéli a tEFL mértékét: eleinte az 1 körüli érték egyértelműen tEFL-t jelez, majd ez szintről szintre csökken, amikor is 15 vízcmm-nél a küszöbszint alá esik az értéke (de a küszöbszinthez közel marad, ami helyes, hiszen a mintázat alakja követi a tEFL-re jellemző alakot, csak már az egészségesekre jellemző tartományban). 20 vízcmm esetén már a mintázat alakján se lehet tEFL-hez hasonló lefolyást látni, a modell kimenete is nulla közelében van. A 0(b) mérés a 20 vízcmm-es mérés után következik, azt vizsgálja, hogy a CPAP hatása mennyire maradandó, de látható, hogy a 0(a)-ban látott alapmintázat egyből visszatér.

A tEFL jósló mértékét vizsgáltam minden páciens esetén, amiből kiderült, hogy a növekvő CPAP szinteken egyénenként változó mértékben, de csökken a tEFL. Ez azért történik,

mert a megemelt légúti nyomás nyitva tartja a tüdőt és segíti a normál működésben. A mellékletben bemutatok két szélsőséges esetet és két *nem* pulmonális tEFL típusú jelenséget. Az első szélsőséges esetben (M.2.) a tEFL mintázata már 10 vízcm támogatásnál azonnal eltűnik. A második esetben (M.3.) még a legmagasabb CPAP szint sem elegendő a tEFL megszüntetésére, annak ellenére, hogy a javuló tendencia végig fennáll, 20 vízcm-nél is még megjelenik a negatív csúcs az Xrs értékeiben. A harmadik példában (M.4.) viszont egy általános lefolyású mérés található, amit meg lehet figyelni, hogy 20 vízcm CPAP szinten zavarosabbá válnak az impedancia hurkok, ami jelen esetben annak a jele, hogy a páciens ellenáll a légzéstámogatásnak, ez a szint már nem komfortos számára. Az Xrs-ben látható tEFL mintázatokhoz az Rrs viselkedése eltérő módon társulhat. A M.3. ábrán látványosan jelenik meg a Xrs minimumához tartozó beesés az Rrs értékében, amely úgy értelmezhető, hogy a légúti összellenállásban kis hányadot képező legkisebb légutak fékezik a légzőhólyagocskák kiürülését, az alacsony áramlás limit nem tudja érvényesíteni a nagyobb légutak áramlásfüggő nemlineáris viselkedését. Amikor viszont a kilégzési áramláskorlátozásban a nagyobb rezisztenciájú centrálisabb hörgők is érintettek, az Rrs is kicsúcsosodik az Xrs minimumánál, ahogy az 5.4. ábrán is látható. Az Rrs mintázatának változása arra adhat tehát tájékoztatást, hogy az OHS kórképében az egyes hörgőgenerációk mennyire érintettek. Az M.5. ábrán egy olyan mérés látható, ahol 0 vízcm CPAP szinten sem látható tEFL. Itt nem pulmonális tEFL lép fel, ilyenkor a páciens a hangrés szűkítésével mesterségesen fenntartja a tüdő térfogatát a könnyebb légzés érdekében, tehát ez egy nem pulmonológiai eredetű felső légúti kilégzési fékezés.



5.5. ábra: Minden OHS mérés szintenkénti válogatásában a tEFL jósolt mértéke

Az 5.5. ábrán minden CPAP szinthez tartozó mérésnek a modell által jósolt tEFL mértékét ábrázolja, látható a pontok eloszlása is a dobozdiagram mellett. 0 vízcm CPAP esetén

három páciens kivételével mindenki rendelkezik tEFL-lel, de ahogy nő a légzéstámogatás, úgy egyre több páciens át lesz sorolva az egészségesekhez. Jól látható, hogy 20 vízcm-nél is van, akinél fennmaradt a tEFL, illetve a 0 vízcm-es újramérés (0b) pontosan követi a legelső mérések eredményeit.

6. ÖSSZEFOGLALÁS

A légzőrendszer mechanikájának oszcillometriás vizsgálata a légzőrendszer bemeneti impedanciájának (komplex Zrs) meghatározását teszi lehetővé. Amíg a hagyományos oszcillometriás méréstechnikák többfrekvenciás vizsgálójellel egy átlag impedancia értéket határoznak meg a kilégzési és belégzési szakaszokra, az általam vizsgált intra-breath technika mindösszesen egy 10 Hz frekvenciájú vizsgálójellel írja le a légzési ciklusokon belüli impedancia változást, amivel a különböző légzőszervi megbetegedések esetén különböző impedancia mintázatokat eredményez.

A kutatásom célja az OHS-ben (obezitás-hipoventiláció szindrómában) szenvedő alanyok intra-breath méréseinek feldolgozása és kiértékelése volt, amelyet öt különböző, egyre növekvő nagyságú folyamatos pozitív légúti nyomás (CPAP) mellett, fekvő testhelyzetben mértek. A kiértékelés fő szempontja a betegek mérési mintázataiban egy bizonyos légzésmechanikai abnormalitásnak, a nyugodt légzés közbeni kilégzési áramlás-korlátozásnak (tEFL) a kimutatása és mértékének megállapítása gépi tanuló módszerek segítségével volt. Tehát a kiértékelést egy olyan adathalmazon végeztem el, amely a mérési módszer és protokoll szempontjából is egyedi.

A nyers szenzoradatok feldolgozása magába foglalta a Zrs kiszámítását, légzési ciklusokra való darabolását, a Zrs adatok (rezisztencia és reaktancia) légzési térfogat és áramlás függvényében való megjelenítését. Ezt követően az így kapott, úgynevezett intra-breath impedancia hurkokból egy mérési hibáktól és légzési szabálytalanságoktól mentes alcsoportot alakítottam ki és ezen csoportokból reprezentatív hurkokat állítottam elő.

Mivel az oszcillometria területén eddig nem végeztek OHS-ben szenvedő betegekkel hasonló kutatásokat, nem állt rendelkezésre címkézett tanítóadatbázis. Ennek hiányában a tanításhoz egészséges légzőrendszerű egyének (kontrollcsoport) mellett a tEFL-t tipikusan mutató, krónikus obstruktív tüdőbetegségben (COPD) szenvedő alanyok vizsgálati anyagát használtam fel. A tEFL folyamatát megragadó jellemzők (mint az áramlás-reaktancia hurok előjeles területe) definiálását követően megfelelő jellemző kiválasztási technikákat alkalmaztam, majd hét gépi tanuló algoritmus segítségével végeztem el az osztályzást. A

választott algoritmusok a gépi tanuló módszerek széles skáláját lefedik. Az elemzés során a legjobban teljesítő algoritmus az MLPClassifier (Multi-layer Perceptron classifier) volt, amely ezen adathalmazon 99,02%-os pontosságot ért el. Az algoritmus teszteléséhez egy speciális betegcsoport (intersticiális tüdőbetegség: ILD) mérési anyagát alkalmaztam.

A tanítás során legjobbnak ítélt modell segítségével kiértékeltem harminchárom OHS-ben szenvedő, kóros testtömegindexszel (BMI tartomány: 32,3-91,1 kg/m²) rendelkező, egyéb légzőszervi megbetegedéssel (pl.: COPD, asztma) nem érintett alany intra-breath méréseit. Az eredmények alapján légzéstámogatás nélkül 30 alanynál mutatkozott a tEFL mintázata, a modell becslése szerint a tEFL mértékének az adathalmazon vett átlag értéke 0,887, ahol 1 jelenti a tEFL biztos, markáns jelenlétét, és 0 az ettől teljes mértékben eltérő hurokmintázatot. A tEFL detektálásához meghatározott küszöbszint 0,5. A folyamatos pozitív légúti nyomás (CPAP) értékének növelésével a tEFL mértéke egyénenként erősen eltérő mértékben csökkent. A legmagasabb vizsgált CPAP értéknél (20 vízcm) a modell csupán három alanynál detektált tEFL-t, itt a tEFL mértékének átlaga már csak 0,149. A CPAP emelésére adott egyéni válaszok ismerete az elsősorban éjszakai CPAP terápia optimális beállítását támogatja. Ehhez az újszerű méréstechnikához csatlakozó gépi döntéshozatal ígéretes eszköz lehet.

A kutatás eredményeit bemutattam a *8th Annual Montreal Oscillometry Summer Seminar* konferencián. A kutatócsoport jövőbeni céljai között szerepel az eredmények tudományos publikálása. További tervek közé tartozik az asztma bevonása az osztályozásba, valamint a tüdőtranszplantáción átesett betegek követéses méréseinek vizsgálata.

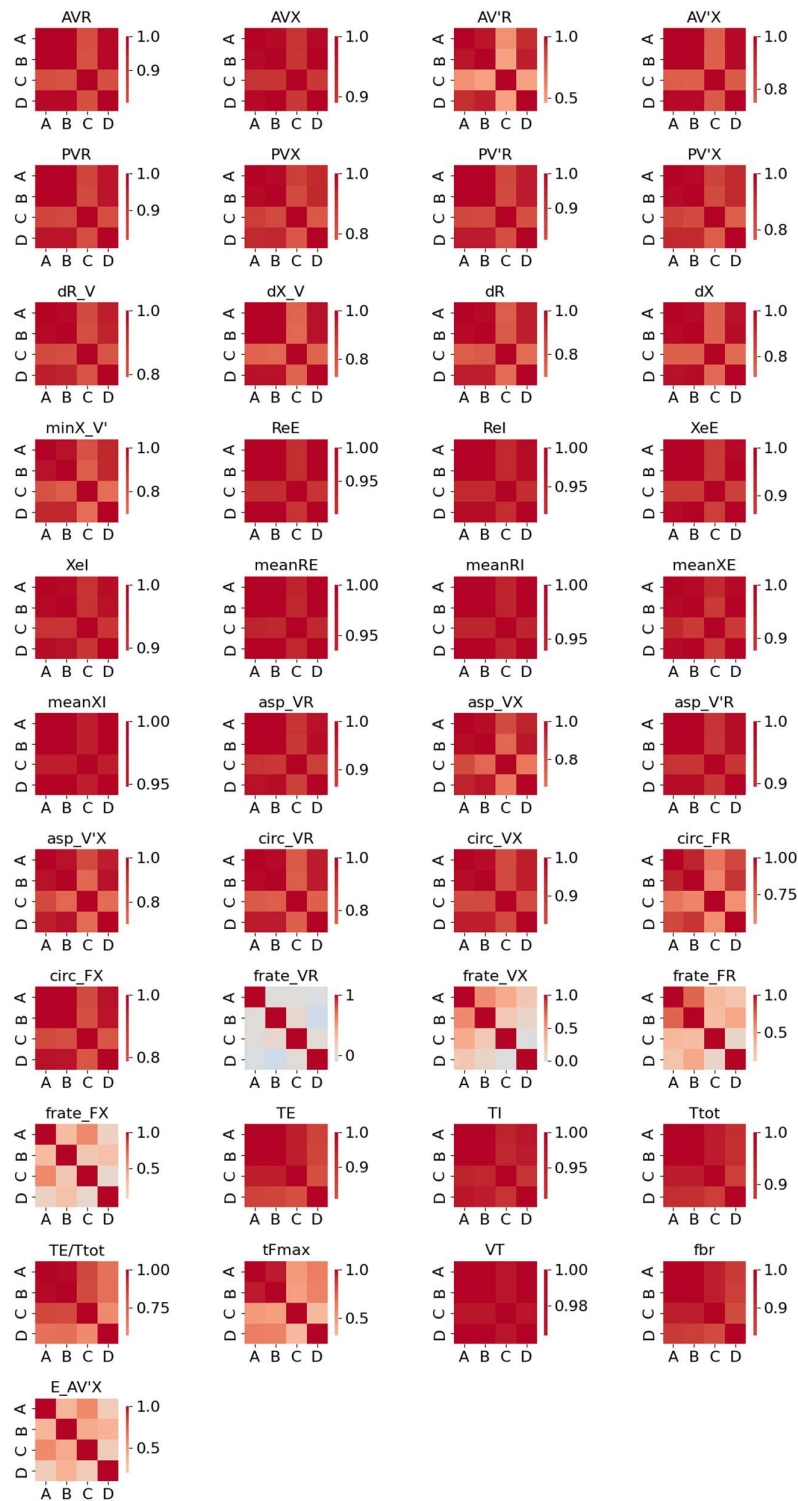
Irodalomjegyzék

1. O.F. Pedersen, J.P. Butler, *Expiratory flow limitation*. Comprehensive Physiology, (2011) 1. 1861–1882. <https://doi.org/10.1002/cphy.c100025>.
2. R.L. Dellacà, P. Santus, A. Aliverti, N. Stevenson, S. Centanni, P.T. Macklem, A. Pedotti, P.M.A. Calverley, *Detection of expiratory flow limitation in COPD using the forced oscillation technique*. The European Respiratory Journal, (2004) 23. 232–240. <https://doi.org/10.1183/09031936.04.00046804>.
3. S. Baglyas, L. Valkó, V. Móró, E. Podmaniczky, D. Czövek, G. Makan, Z. Gingl, J. Gál, Z. Hantos, A. Lox, *Using intra-breath oscillometry in obesity hypoventilation syndrome to detect tidal expiratory flow limitation: a potential marker to optimize CPAP therapy*. BMC Pulmonary Medicine, (2023) 23. 477. <https://doi.org/10.1186/s12890-023-02777-x>.
4. J.H.T. Bates, C.G. Irvin, R. Farré, Z. Hantos, *Oscillation Mechanics of the Respiratory System*, in: Comprehensive Physiology, John Wiley & Sons, Ltd, 2011 1233–1272. <https://doi.org/10.1002/cphy.c100058>.
5. G.G. King, J. Bates, K.I. Berger, P. Calverley, P.L. de Melo, R.L. Dellacà, R. Farré, G.L. Hall, I. Ioan, C.G. Irvin, D.W. Kaczka, D.A. Kaminsky, H. Kurosawa, E. Lombardi, G.N. Maksym, F. Marchal, B.W. Oppenheimer, S.J. Simpson, C. Thamrin, M. van den Berge, E. Oostveen, *Technical standards for respiratory oscillometry*. The European Respiratory Journal, (2020) 55. 1900753. <https://doi.org/10.1183/13993003.00753-2019>.
6. Z. Hantos, *Intra-breath oscillometry for assessing respiratory outcomes*. Current Opinion in Physiology, (2021) 22. 100441. <https://www.sciencedirect.com/science/article/pii/S2468867321000596> (Utoljára megtekintve: 2024. 10. 29.).
7. D.A. Kaminsky, S.J. Simpson, K.I. Berger, P. Calverley, P.L. de Melo, R. Dandurand, R.L. Dellacà, C.S. Farah, R. Farré, G.L. Hall, I. Ioan, C.G. Irvin, D.W. Kaczka, G.G. King, H. Kurosawa, E. Lombardi, G.N. Maksym, F. Marchal, E. Oostveen, B.W. Oppenheimer, P.D. Robinson, M. van den Berge, C. Thamrin, *Clinical significance and applications of oscillometry*. European Respiratory Review: An Official Journal of the European Respiratory Society, (2022) 31. 210208. <https://doi.org/10.1183/16000617.0208-2021>.
8. A.B. DuBois, A.W. Brody, D.H. Lewis, B.F. Burgess, *Oscillation Mechanics of Lungs and Chest in Man*. Journal of Applied Physiology, (1956) 8. 587–594. <https://doi.org/10.1152/jappl.1956.8.6.587>.

9. G.G. King, J. Bates, K.I. Berger, P. Calverley, P.L. de Melo, R.L. Dellacà, R. Farré, G.L. Hall, I. Ioan, C.G. Irvin, D.W. Kaczka, D.A. Kaminsky, H. Kurosawa, E. Lombardi, G.N. Maksym, F. Marchal, B.W. Oppenheimer, S.J. Simpson, C. Thamrin, M. van den Berge, E. Oostveen, *Technical standards for respiratory oscillometry*. European Respiratory Journal, (2020) 55. <https://doi.org/10.1183/13993003.00753-2019>.
10. R. Peslin, T. Hixon, J. Mead, *Variations of thoraco-pulmonary resistance during the respiratory cycle studied by the oscillation method*. Bulletin De Physio-Pathologie Respiratoire, (1971) 7. 173–188.
11. A. Lox, D. Czövek, Z. Gingl, G. Makan, B. Radics, D. Bartusek, S. Szigeti, J. Gál, G. Losonczy, P.D. Sly, Z. Hantos, *Airway dynamics in COPD patients by within-breath impedance tracking: effects of continuous positive airway pressure*. The European Respiratory Journal, (2017) 49. 1601270. <https://doi.org/10.1183/13993003.01270-2016>.
12. X. Shen, H. Liu, *Using machine learning for early detection of chronic obstructive pulmonary disease: a narrative review*. Respiratory Research, (2024) 25. 336. <https://doi.org/10.1186/s12931-024-02960-6>.
13. J.L.M. Amaral, A.J. Lopes, J.M. Jansen, A.C.D. Faria, P.L. Melo, *Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease*. Computer Methods and Programs in Biomedicine, (2012) 105. 183–193. <https://doi.org/10.1016/j.cmpb.2011.09.009>.
14. A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, K. Kotecha, *Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease*. PeerJ Computer Science, (2021) 7. e369. <https://doi.org/10.7717/peerj-cs.369>.
15. *LocalOutlierFactor*. Scikit-Learn,. <https://scikit-learn/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html> (Utoljára megtekintve: 2024. 11. 21.).
16. *scikit-learn: machine learning in Python — scikit-learn 1.5.2 documentation*. <https://scikit-learn.org/stable/index.html> (Utoljára megtekintve: 2024. 11. 18.).
17. *SelectKBest*. Scikit-Learn,. https://scikit-learn/stable/modules/generated/sklearn.feature_selection.SelectKBest.html (Utoljára megtekintve: 2024. 11. 18.).
18. *f_classif*. Scikit-Learn,. https://scikit-learn/stable/modules/generated/sklearn.feature_selection.f_classif.html (Utoljára megtekintve: 2024. 11. 18.).

19. *PCA*. Scikit-Learn,. <https://scikit-learn/stable/modules/generated/sklearn.decomposition.PCA.html> (Utoljára megtekintve: 2024. 11. 18.).
20. *1.2. Linear and Quadratic Discriminant Analysis*. Scikit-Learn,. https://scikit-learn/stable/modules/lda_qda.html (Utoljára megtekintve: 2024. 11. 18.).
21. *RandomForestClassifier*. Scikit-Learn,. <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html> (Utoljára megtekintve: 2024. 11. 18.).
22. *GradientBoostingClassifier*. Scikit-Learn,. <https://scikit-learn/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html> (Utoljára megtekintve: 2024. 11. 18.).
23. *SVC*. Scikit-Learn,. <https://scikit-learn/stable/modules/generated/sklearn.svm.SVC.html> (Utoljára megtekintve: 2024. 11. 18.).
24. *MLPClassifier*. Scikit-Learn,. https://scikit-learn/stable/modules/generated/sklearn.neural_network.MLPClassifier.html (Utoljára megtekintve: 2024. 11. 18.).
25. *KNeighborsClassifier*. Scikit-Learn,. <https://scikit-learn/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html> (Utoljára megtekintve: 2024. 11. 18.).
26. *ExtraTreesClassifier*. Scikit-Learn,. <https://scikit-learn/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html> (Utoljára megtekintve: 2024. 11. 18.).
27. *LogisticRegression*. Scikit-Learn,. https://scikit-learn/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (Utoljára megtekintve: 2024. 11. 18.).

MELLÉKLETEK



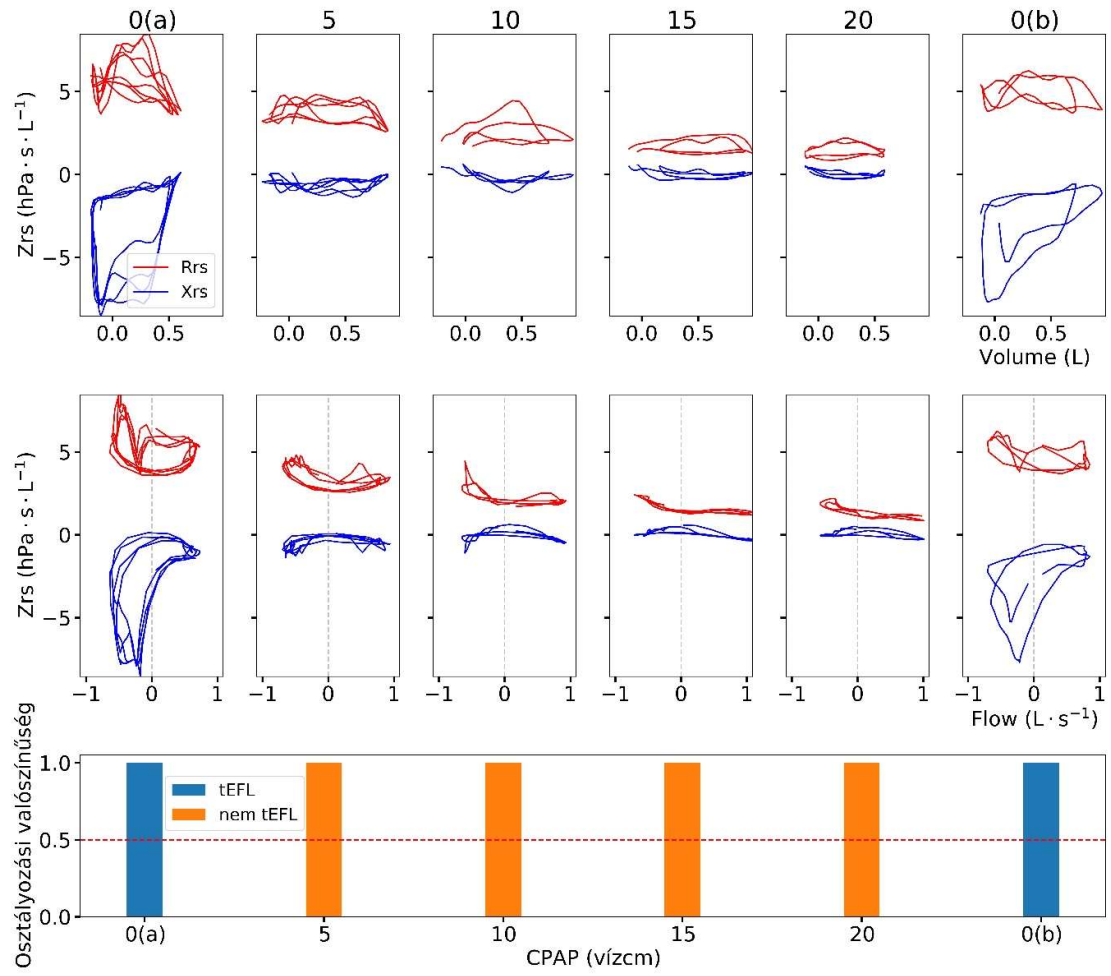
M.1. ábra: Jellemzőszámítási módszerek korrelációs összehasonlítása, az összes adatpontra számolva. A: átlag értékek, B: medián értékek, C: hurokok mediánja, D: mesterséges medián hurok

paraméter	Keresési tér
RandomForestClassifier	
n_estimators	2, 3, 4, 5
max_depth	2,4,8,16
GradientBoostingClassifier	
loss	log_loss, exponential
learning_rate	0.1, 0.05, 0.01
n_estimators	5, 10, 50
subsample	1.0, 0.8, 0.9
max_depth	3, 7, 10
MLPClassifier	
solver	adam, sgd
alpha	0.1, 0.01, 0.001, 0.0001
hidden_layer_sizes	(6,3), (35,10), (20,15,10,5)
learning_rate_init	0.1, 0.01, 0.001, 0.005
SVC	
C	0.1, 1, 10, 100
kernel	linear, rbf, poly, sigmoid
gamma	scale, auto, 0.001, 0.01, 0.1
degree	2, 3, 4, 5
shrinking	true, false
KNeighborsClassifier	
n_neighbors	3, 5, 7, 10
weights	uniform, distance
algorithm	ball_tree, kd_tree, brute, auto
p	1, 2
ExtraTreesClassifier	
n_estimators	10, 50, 100, 200, 500
criterion	gini, entropy, log_loss
max_depth	10, 20, 30, 50
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4
max_features	sqrt, log2
bootstrap	true, false
LogisticRegression	
penalty	l1, l2, elasticnet, none
C	0.1, 1, 10, 100
solver	lbfgs, liblinear, saga, newton-cg
max_iter	100, 200, 500

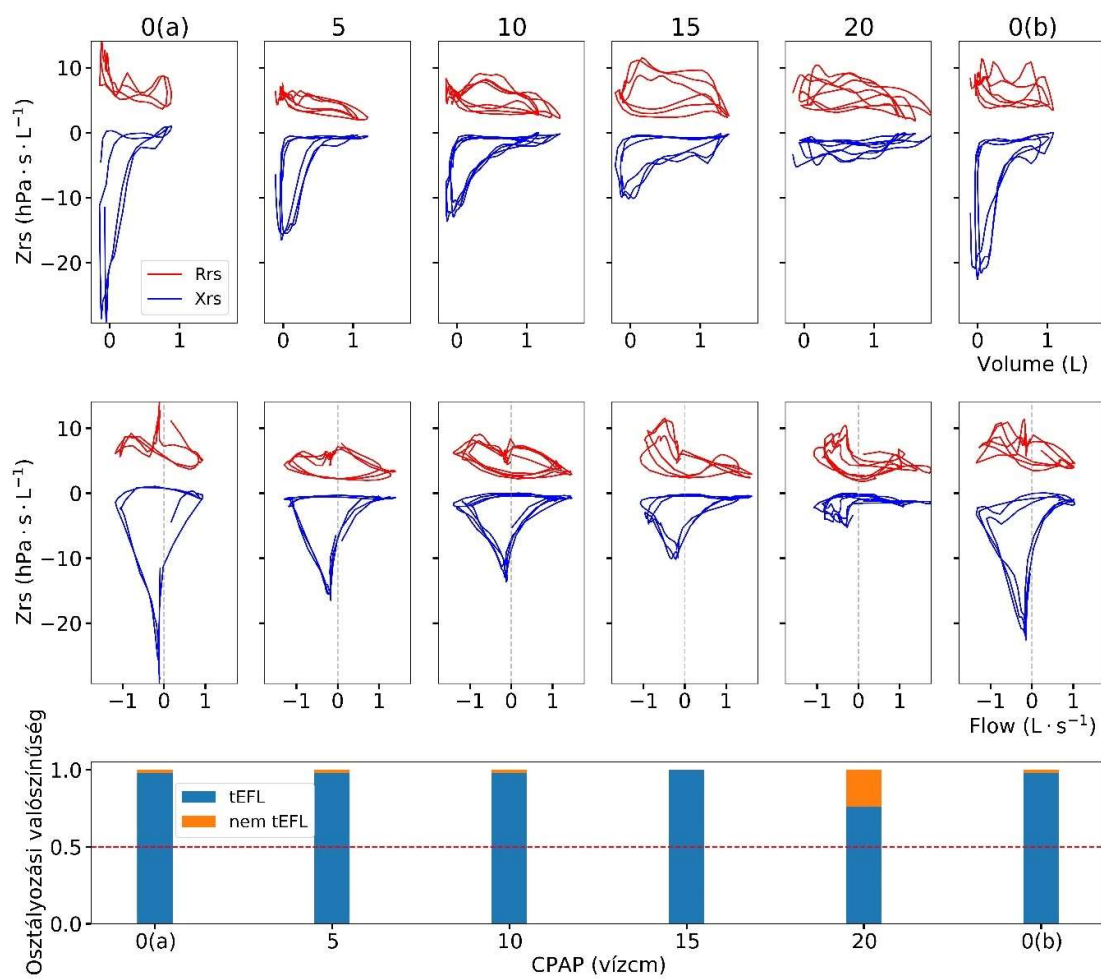
M.1. táblázat: Vizsgált metaparaméterek, és a legjobb értékük pirossal

Modell neve	Találati arány	F1-Score (Egészséges)	F1-Score (COPD)	F1-Score (ILD)	Súlyozott átlag találati arány
MLPClassifier	0.79	0.68	0.93	0.8	0.79
GradientBoostingClassifier	0.87	0.79	0.99	0.86	0.87
RandomForestClassifier	0.79	0.64	0.97	0.8	0.8
KNeighborsClassifier	0.85	0.78	0.96	0.85	0.85
SVC	0.8	0.7	0.92	0.81	0.8
ExtraTreesClassifier	0.86	0.78	0.98	0.86	0.86
LogisticRegression	0.8	0.72	0.97	0.78	0.8

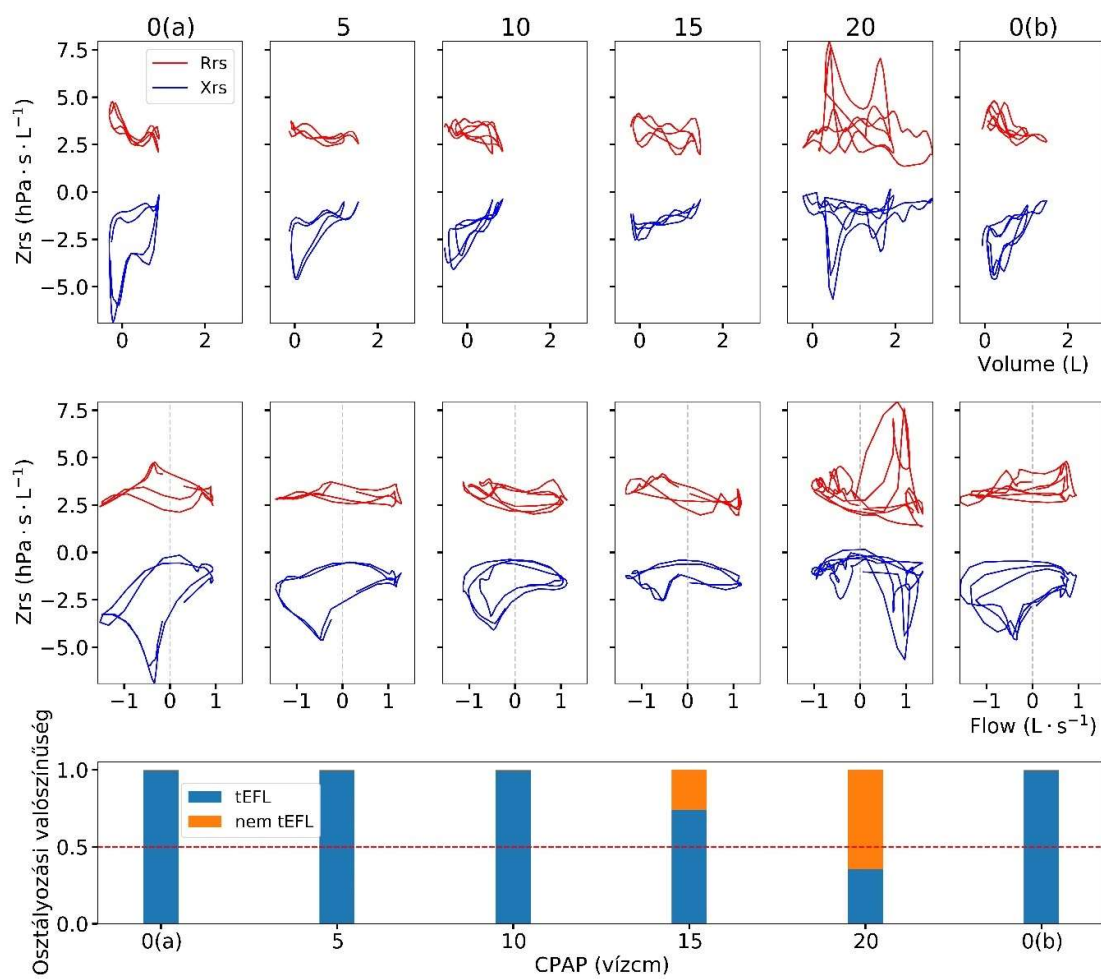
M.2. táblázat: ILD, COPD, és egészséges beteganyag osztályzásának részletes kimenete



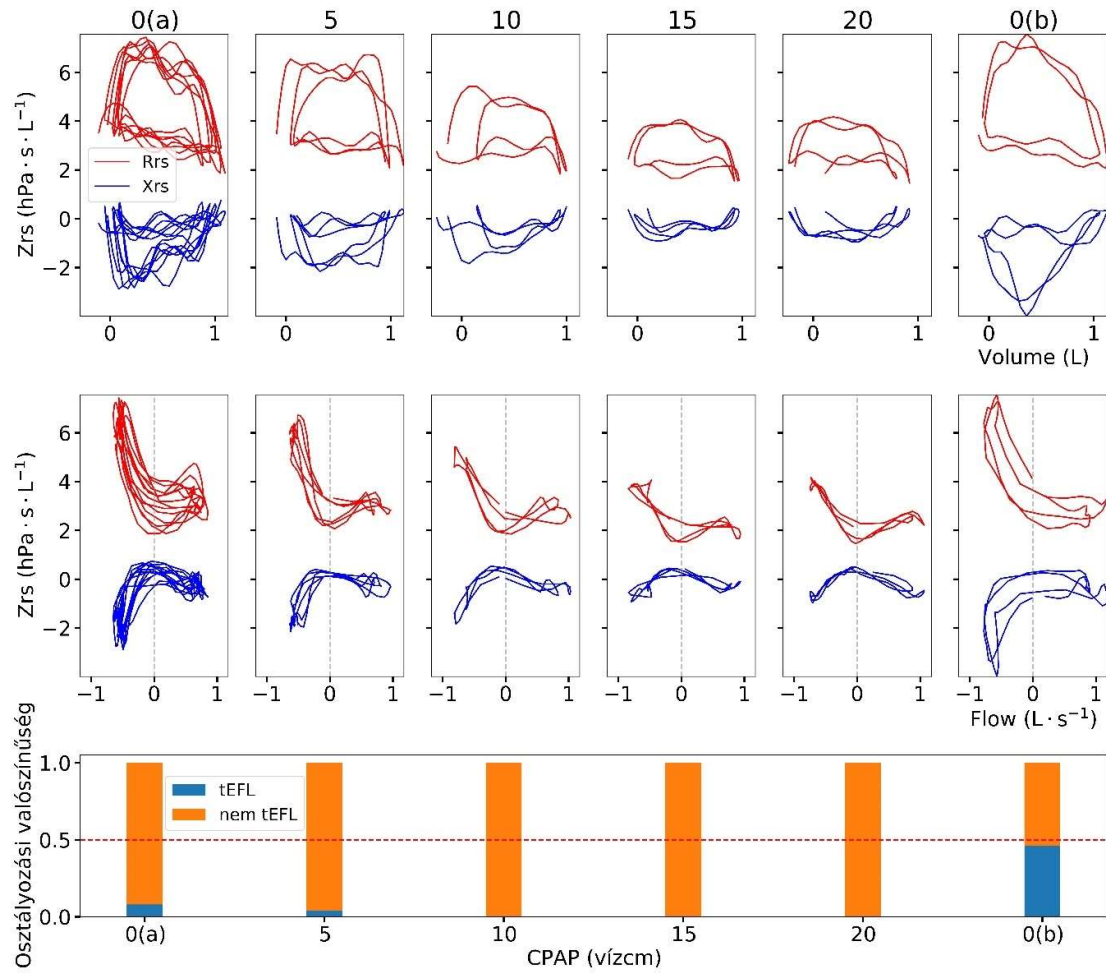
M.2. ábra: OHS-ben szenvedő alany mérésének kiértékelése minden CPAP szinten



M.3. ábra: OHS-ben szenvedő alany mérésének kiértékelése minden CPAP szinten



M.4. ábra: OHS-ben szenvedő alany mérésének kiértékelése minden CPAP szinten



M.5. ábra: OHS-ben szenvedő alany mérésének kiértékelése minden CPAP szinten