TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Helen Tamm 193299IAIB

Hingamishelide andmeanalüüs maatriksprofiili meetodil

Bakalaureusetöö

Juhendaja: Ants Torim
PhD

Tallinn 2024

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Helen Tamm

10.01.2024

Annotatsioon

Käesolev töö keskendub kopsuhelide uurimuslikule eelanalüüsile. Töö eesmärk on hinnata maatriksprofiili meetodi abil leitud kopsuhelide anomaaliaid, motiive ja gruppe e klastreid ning leida nende seosed kroonilise obstruktiivse kopsuhaiguse jt kopsuhaiguste esinemisega ja läbi selle valideerida, kas maatriksprofiili meetod võib olla efektiivne kopsuhelide analüüsimisel.

Kopsuhelidest anomaaliate ja mustrite leidmine võib olla sisendiks masinõppemudelite treenimisel. Neid mudeleid saab kasutada näiteks kopsuhaiguste diagnoosimise ja monitoorimise seadmete täiendamisel ning suurendada seeläbi erinevate hingamisteede haiguste diagnoosimise ja ägenemise hindamise täpsust. Lisaks võib aidata selgete gruppide avastamine tuvastada kopsuhaiguste erinevaid alamtüüpe. Lõputöö eesmärk ei ole ennustava mudeli loomine.

Töö koosneb teoreetilisest osast, mis tugineb erialasele- ja teaduskirjandusele ning selgitab maatriksprofiili meetodi olemust ja praktilisest osast, kus antud meetodi abil ning kombineerides seda teiste andmeanalüüsimeetoditega otsitakse kopsuhelidest anomaaliaid ja motiive ning võimalikke klastreid.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 31 leheküljel, 8 peatükki, 16 joonist, 4 tabelit.

Abstract

Respiratory sound analysis with matrix profile

The thesis focuses on preliminary analysis of respiratory sounds. The aim of the thesis is to detect potential anomalies, motifs and distinct groups or clusters within a data set and find their relation with COPD and other respiratory diseases. Through this, the objective is to validate whether the matrix profile method can be an effective approach in analyzing lung sounds.

Finding anomalies and patterns in lung sounds can serve as input for training machine learning models. These models can be used, for example, to enhance diagnostic and monitoring devices for respiratory diseases, thereby increasing the precision of diagnosing and assessing exacerbations of various respiratory conditions. Additionally, the discovery of clear groups can help identify different subtypes of lung diseases. The goal of the thesis is not to create a predictive model.

The work consists of a theoretical part, based on professional and scientific literature, explaining the nature of the matrix profile method. The practical part utilizes this method, combined with other data analysis methods, to search for anomalies, patterns, and possible clusters in lung sounds.

The thesis is in Estonian and contains 31 pages of text, 8 chapters, 16 figures, 4 tables.

Lühendite ja mõistete sõnastik

Maatriksprofiil	Aegridade analüüsimise meetod
PCA	Peakomponentide analüüs
KOK	Krooniline obstruktiivne kopsuhaigus
URTI	Ülemiste hingamisteede haigusseisundid
LRTI	Alumiste hingamisteede haigusseisundid
DBSCAN	Tiheduspõhine klasterdamisalgoritm
Motiiv	Korduv alamjada pikemas aegreas
Anomaalia	Alamrida, mis erineb oluliselt aegrea teistest alamridadest
Vilin	Hingamisel tekkiv pidev kopsuheli, mis viitab obstruktsioonile
Rägin	Hingamisel tekkiv ragisev katkendlik kopsuheli
MFCC	Mel-sageduse kepstri kordaja

Sisukord

Jooniste loetelu

Tabelite loetelu

1 Sissejuhatus

1.1 Taust ja probleem

Maatriksprofiili meetodit tutvustasid esmakordselt 2016. aastal Eamonn Keogh ja Abdullah Mueen luues sellega uue võimaluse aegridade analüüsimiseks ning sealt anomaaliate ja motiivide tuvastamiseks.

Antud meetod on näidanud lootustandvaid tulemusi erinevates valdkondades. Teaduskirjanduses on mainitud edukust näiteks muusikaliste helifailide sarnasuste tuvastamisel [20] ja müratasteme analüüsimisel [12].

Maatriksprofiil on andnud häid tulemusi ka piiratud aegridade esitusel (nt analüüsides ainult ühte aegrea dimensiooni) olukordades, kus teised meetodid on vajanud aegrea mitmedimensioonilist esitust ning näidanud häid tulemusi meditsiiniliste andmete, nt DNA, analüüsimisel [20] Sellest lähtuvalt võib arvata, et maatriksprofiili meetod on edukas ka teiste keerukate aegridade analüüsimisel ning tervishoiusüsteemis esinevate probleemide lahendamisel.

Üheks selliseks probleemiks on kopsuhaiguste esinemine. Probleemi aktuaalsust näitab, et krooniline obstruktiivne kopsuhaigus ehk KOK, on maailmas surmemuse põhjuste seas 3. kohal põhjustades ca 3 miljonit surmajuhtu aastas ning tuues endaga kaasa suure majandusliku ja sotsiaalse koormuse. [19]

Kopsuhaiguste õigeaegne ja varajane diagnoosimine hoiab ära haiguse süvenemise ja aitab parandada patsiendi elukvaliteeti. Samuti võivad haigusest tingitud suuremad muutused kopsudes olla tagasipöördumatud, mis tähendab, et varane diagnoosmine, haiguse ägenemise kiire tuvastamine ja raviga õigeaegselt alustamine võib olla eluliselt tähtis [14].

Paljud KOK-i diagnoosiga inimesed suudavad edukalt tuvastada enda haigusseisundi ägenemist köhimise, rögaerituse ja enesetunde põhjal. Osa patsientidest suudab iseseisvalt koduse raviskeemiga sümptomeid leevendada või kontakteeruda vajadusel

perearstiga. Siiski esineb juhte, kus inimene satub haiguse ootamatu ägenemisega erakorralise meditsiini osakonda [18].

Selliste olukordade ennetamisel võiks abi olla tehisintellekti kaasamisest raviprotsessi. Tehnoloogiliste lahenduste abil on võimalik monitoorida inimest reaalajas. Kasutuses olevatest monitooringuseadmetest võib näitena tuua reaalajas veresuhkru taseme jälgimise lahenduse diabeetikutele, kus kogu sensori teave on nähtav mobiiliäpis ning laetakse üles ka pilve. Tänapäeva meditsiinis võiks aga pideva monitoorimise võimalusi olla oluliselt rohkem.

Antud lõputöö eesmärgiks on maatriksprofiili meetodi abil tuvastada kopsuhelides esinevaid anomaaliaid ja motiive ning gruppe e klastreid. Kopsuhelidest anomaaliate ja mustrite leidmine võib olla sisendiks masinõppemudelite treenimisel. Neid mudeleid saab kasutada kopsuhaiguste diagnoosimise ja monitoorimise seadmete täiendamisel ning suurendada seeläbi erinevate hingamisteede haiguste diagnoosimise ja ägenemise hindamise täpsust ja vähendada koormust tervishoiusüsteemile.

1.2 Töö eesmärk

Lõputöö esimeseks eesmärgiks on maatriksprofiili meetodi abil kopsuhelides esinevate anomaaliate ja motiivide tuvastamine. Antud eesmärgist lähtuvalt püstitati järgnevad uurimisküsimused:

1) Kas vilinad ja räginad liigituvad selgelt anomaaliateks või motiivideks
2) Kas motiivide ja anomaaliate abil on võimalik vilinaid ja räginaid liigitada

Lõputöö teine eesmärk on kopsuhelides esinevate võimalike gruppide e klastrite tuvastamine.

Klasterdamisel püstitati järgnevad uurimisküsimused:

1) Kuidas kopsuhelid rühmituvad ehk klasterduvad?
2) Kas tekivad selgelt eristuvad klastrid, mis on seotud erinevate diagnoosidega?
    - Kui klastrid ei ole seotud diagnoosidega, siis milline on loogika tekkinud klastrite taga?

1.3 Metoodika

Lõputöö metoodika on uurimuslik andmeanalüüs.

1.3.1 Anomaaliate ja motiivide tuvastamine

Kopsuhelidest anomaaliate ja motiivide tuvastamiseks kasutatakse maatriksprofiili meetodit, mis tugineb aegrea alamridade võrdlemisele ning nendevahelise kauguse arvutamisele.

1.3.2 Klasterdamine

Klasterdamisel kasutatakse maatriksprofiili põhist kaugusmõõtu kopsuhelide omavahelisel võrdlemisel, mida rakendatakse DBSCAN klasterdamisalgoritmi hüperparameetrina. Klasterdamise tulemuste valideerimiseks kasutatakse hii-ruut testi.

1.4 Töö struktuur

Lõputöö teoreetiline osa annab teaduskirjanduse põhjal ülevaate maatriksprofiili meetodi olemusest, aegridade vahelise kauguse arvutamisest ning kirjeldab antud meetodi abil aegreas esinevate anomaaliate ja motiivide tuvastamist.

Töö praktiline osa jaguneb kaheks – esimeses osas rakendatakse maatriksprofiili meetodit kopsuhelidest anomaaliate ja motiivide tuvastamiseks. Praktilise töö teine osa keskendub kopsuhelidest võimalike klastrite leidmisele.

2 Senised uuringud

Kroonilise obstruktiivse kopsuhaiguse ägenemise tuvastamisel on mitmetes uuringutes kasutatud masin- ja süvaõppe meetodeid [6], [17]. Ennustusmudelite loomisel on kasutatud näiteks juhusliku metsa, tugivektor-masina, logistilise regressiooni, K-lähima naabri ja Bayes'i algoritme.

Süvaõpet on kasutatud ka kopsuhelide klassifitseerimiseks. Näiteks on kopsuhelisid klassifitseeritud vilinate esinemise [2] ja ka teiste kopsuhaigustele omaste patoloogiliste helide põhjal. [8] Kopsuhelidest anomaaliate tuvastamisel on kasutatud süvanärvivõrke. [4].

Enim kasutatud meetoditeks kopsuhaiguste tuvastamisel ja klassifitseerimisel on närvivõrkude ja otsustuspuude treenimine.[15].

Maatriksprofiili meetod on küllaltki uus aegridade analüüsimise viis, mida on testitud erinevate aegridade, sh meditsiiniliste andmete, analüüsimisel, kus antud meetod on andnud lubavaid tulemusi. [20]. Hetkel ei leidu kirjanduses ühtegi viidet, kus maatriksprofiili meetodit oleks kasutatud kopsuhaiguste analüüsimisel või tuvastamisel.

3 Andmestik

Käesolevas töös on kasutatud avalikult kättesaadavat kopsuhelide andmebaasi [21]. Andmebaas sisaldab 5,5 tunni tunni ulatuses helifaile, milles on käsitletud 6898 hingamistsüklit. Neist 1864 sisaldab räginaid, 886 vilinaid ja 506 nii välinaid kui räginaid. Üks helifail koosneb mitmest hingamistsüklist ning iga helifailiga on kaasas samanimeline tekstifail, mis sisaldab informatsiooni vilinate ja räginate esinemise kohta failis sisalduvates hingamistsüklites. Lisaks on olemas iga helifaili jaoks eraldi fail, kus on täpselt kirjeldatud ajavahemikud, millal räginad või vilinad helifailis esinesid.

Andmebaasis on kokku 920 helifaili, mis on kogutud 126 erinevalt patsiendilt erinevates Portugali ja Kreeka meditsiiniasutustes. Helifailide hulgas esineb nii tervete inimeste kui erinevate kopsuhaigustega patsientide hingamisheli. Tabelis 1 on välja toodud andmestikus esinevad erinevad diagnoosid koos patsientide ja failide arvuga ja võimalikele patoloogiatele viitavate helidega hingamistsüklite arv failide hulgas.

Tabel 1Andmestiku kirjeldus

Diagnoos	Patsientide arv	Failide arv	Ainult vilinad	Ainult raginad	Vilinad ja raginad
KOK	64	793	752	1779	490
Terve	26	35	2	16	1
Kopsupõletik	6	37	24	21	0
Bronhioliit	6	13	64	15	5
Bronhiektaas	7	16	31	12	10
Ülemiste hingamisteede infektsioonid	14	23	8	21	0
Alumiste hingamisteede infektsioonid	2	2	1	0	0
Astma	1	1	4	0	0

3.1 Andmete eeltöötlemine

Helifailide esituseks valiti Mel-sageduse kepstri kordajad, sest neid on kopsuhelide analüüsimisel edukalt kasutatud [2]. Mel-sageduse kepstri kordaja abil on tuvastatud helifailide vahel sarnasusi ka maatriksprofiili meetodil [9].

Üldiselt ei vaja maatriksprofiili arvutamise algoritmid suuremat andmete eeltöötlemist. [10], kuid lähtuvalt kopsuhelide omadustest, optimeeriti vähesel määral ka helifailide omadusi.

Südamehelid võivad segada kopsuhelide analüüsimist. Südamehelide sagedus jääb üldiselt vahemikku 20-100 Hz. Kopsuhelid jäävad üldjuhul vahemikku 50-2500 Hz, kuid võivad kõri piirkonnas ulatuda ka kuni 4000 Hz [1].

Andmestikus esineb ka trahhea ehk kõri piirkonnas mõõdetud kopsuhelisid, seega tuleb ülempiiri seadmisel sellega arvestada. Ka ei saa alampiiri seadistada nii, et südamehelid analüüsitavast helist täielikult välja jääksid, sest see võib endaga kaasa tuua ka olulise kao kopsuhelide informatsioonis.

Lähtudes sellest keskenduti käesolevas töös anomaaliate ja motiivide leidmisel sagedusvahemikule 50-4000 Hz.

4 Tehnoloogilised valikud

Käesolevas töös on kasutatud programmeerimiskeelt Python 3 ning selle andme-, sh helianalüüsiteeke (LibROSA, STUMPY) ning Jupyter Notebooki.

Python 3 on interpreteeritud interaktiivne objektorienteeritud programmeerimiskeel, milles on kombineeritud märkimisväärne võimsus ja puhas süntaks [23].

Jupyteri töövihik on interaktiivne arenduskeskkond, mis võimaldab koodi jooksutada ja dokumenteerida. Lisaks pakub keskkond laialdasi võimalusi andmete visualiseerimiseks [22].

Programmeerimiskeele valikul sai määravaks Pythoni laildane masinõppe- ja helianalüüsiteekide olemasolu. Arenduskeskkonna valik tugines võimalusele tulemusi lihtsalt intepreteerida ning jagada.

5 Maatriksprofiil

Maatriksprofiili meetod on üks võimalikke aegridade analüüsimise ning sealt anomaaliate ja motiivide tuvastamise meetodeid. Aegridade võrdlemise üks viise on kalkuleerida nendevaheline otsene ehk eukleidiline kaugus. Maatriksprofiil ongi oma olemuselt vektor, mis hoiab seda informatsiooni z-normaliseeritud kujul mistahaes aegrea alamrea ja tema lähima naabri kohta [20].

Maatriksprofiili üheks eeliseks on arvutamise kiirus. Kui naiivsete algoritmidega aegridade võrdlemine võib pikematel aegridadel võtta aega kuid või päevi, suudavad maatriksprofiili algoritmid seda teha kordades kiiremini (Joonis 1). Samuti ei anna meetod valepositiivseid ega -negatiivseid tulemusi anomaaliate ja motiivide tuvastamisel. Maatriksprofiil sobib erisuguste aegridade võrdlemiseks ning ei nõua ulatuslikke parameetreid [20].

TABLE VII. TIME FOR A SELF-JOIN WITH $M = 256$, VARYING ALGORITHMS

Algorithm	$TSFR_{DAA}$	$HDSJ_{I\text{-}SAX}$	ONL	STAMP
Time Required	51.7 hours	19.6 min	28.1 hours	1.17 hours

Joonis 1 Erinevate algoritmide kiirus [20]

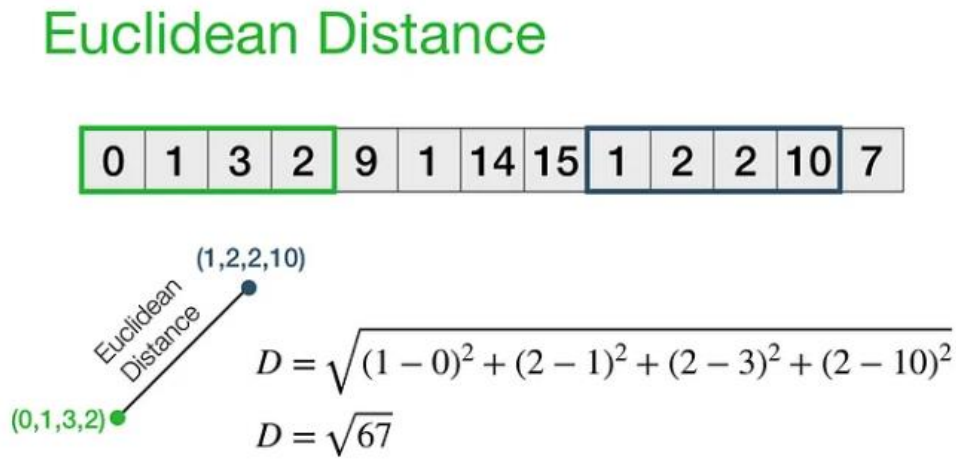
5.1 Maatriksprofiili arvutamine

Maatriksprofiili arvutamisel aegreal võrreldakse omavahel kõiki selle aegrea alamridasid. Algset või esmast alamrida pikkusega m võrreldakse kõigi järgnevate sama pikkusega alamridadega. Seejärel nihutatakse algset alamrida ühe positsiooni võrra edasi ja korratakse protseduur [16].

Ignoreeritakse võrdlust alamjada endaga ning kõigi võrdluste lõpus säilitatakse informatsioon vaid lähima kaugusega alamrea kohta. [16]

5.1.1 Üks aegrida

Maatriksprofiili arvutamine ühel aegreal on tuntud kui *self-join*, sest sel juhul võrreldakse aegrea alamjadasid ainult selle sama aegrea teiste sama pikkusega alamridadega ja arvutatakse nende vaheline kaugus.[16]



Joonis 2 Aegridade vahelise otsese kauguse arvutamine [16]

5.1.2 Kaks aegrida

Kahe aegrea võrdlemist tuntakse kui *AB-join*. Kahe aegrea A ja B võrdlemine maatriksprofiili abil võimaldab tuvastada, kui sarnased on aegrea A alamread aegrea B alamridadele. Võrreldavad aegread ei pea olema ühtlase pikkusega, kuid oluline, et võrreldavate alamridade pikkus oleks maksimaalselt sama pikk, kui on lühem aegrida [20].

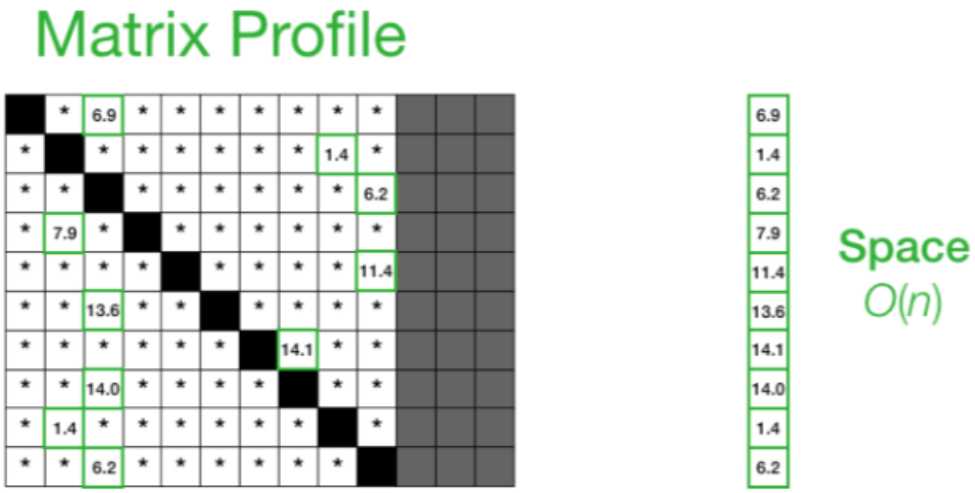
Faili A iga m pikkuse alamrea ($P(x_1, x_2, ..., x_n)$) kaugus igast faili B m pikkusega alamrreast ($Q(y_1, y_2, ..., y_n)$) arvutatakse valemiga:

$$d(P,Q) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \ldots + (y_n - x_n)^2}$$

5.1.3 Maatriksprofiil

Alamridade vaheliste kauguste arvutamise tulemusena tekib kaugusmaatriks, kus on informatsioon ainult iga elemendi lähima naabri kohta.Sellise kaugusmaatriksi

lihtsustatud esitus ongi maatriksprofiil ehk vektor, mis talletab z-normaliseeritud kujul informatsiooni iga alamrea ja selle lähima naabri kohta.



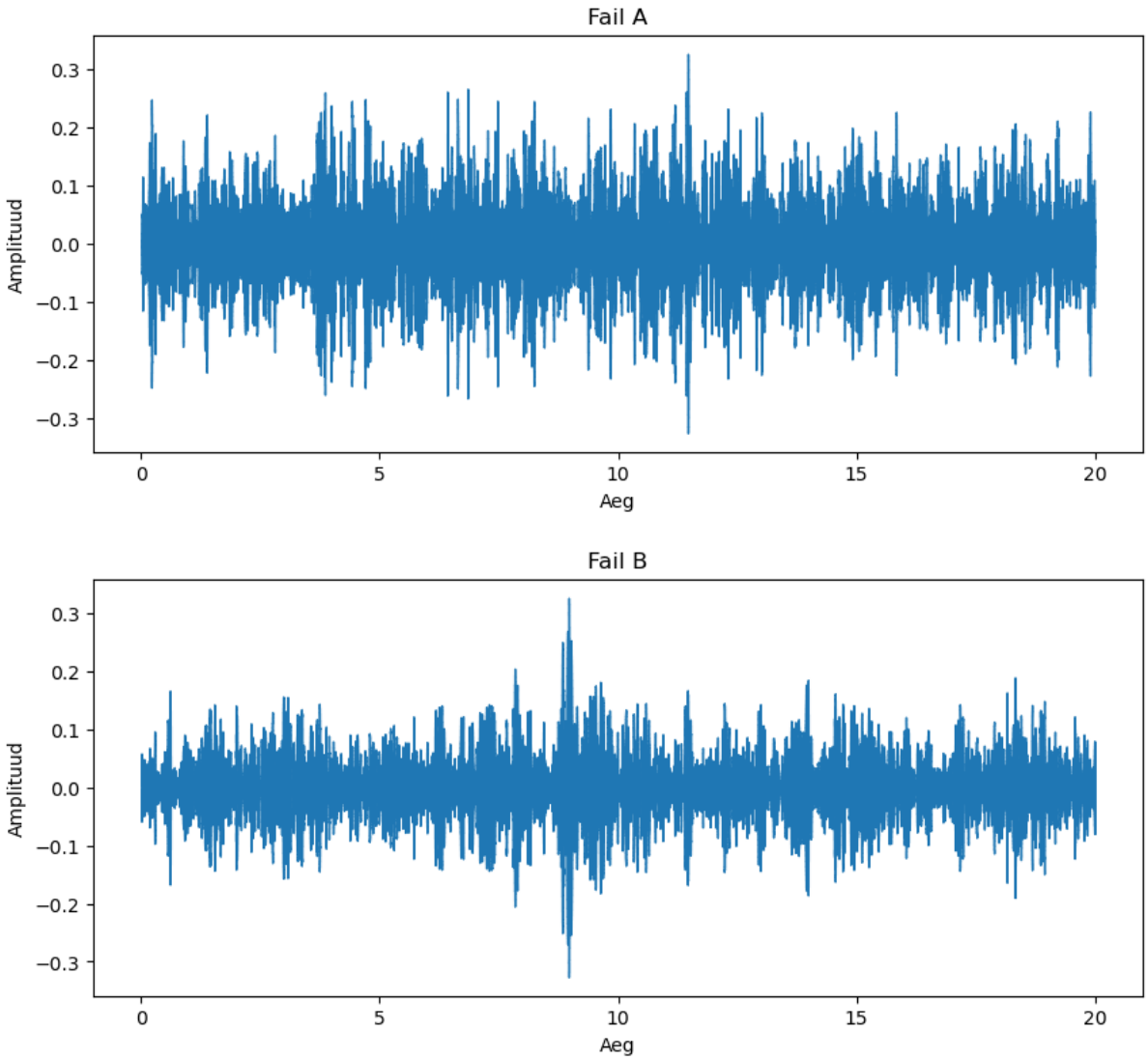
Joonis 3 Maatriksprofiili vektoresitus [16]

5.1.4 Motiivide ja anomaaliate tuvastamine

Motiiv on pikemas aegreas korduvalt esinev sarnane alamrida. Aegridade vahelise kauguse hindamisel viitab motiivi olemasolule see, kui aegrea lähim naaber asub lähedal. Vastupidiselt motiivile, annab informatsiooni aegridade suure erunevuse kohta kahe aegrea vaheline suur kaugus. Pikemas aegreas esinev alamrida, mille lähim naaber asub väga kaugel, võib viidata, et tegemist on aegreas esineva anomaaliaga. [20]

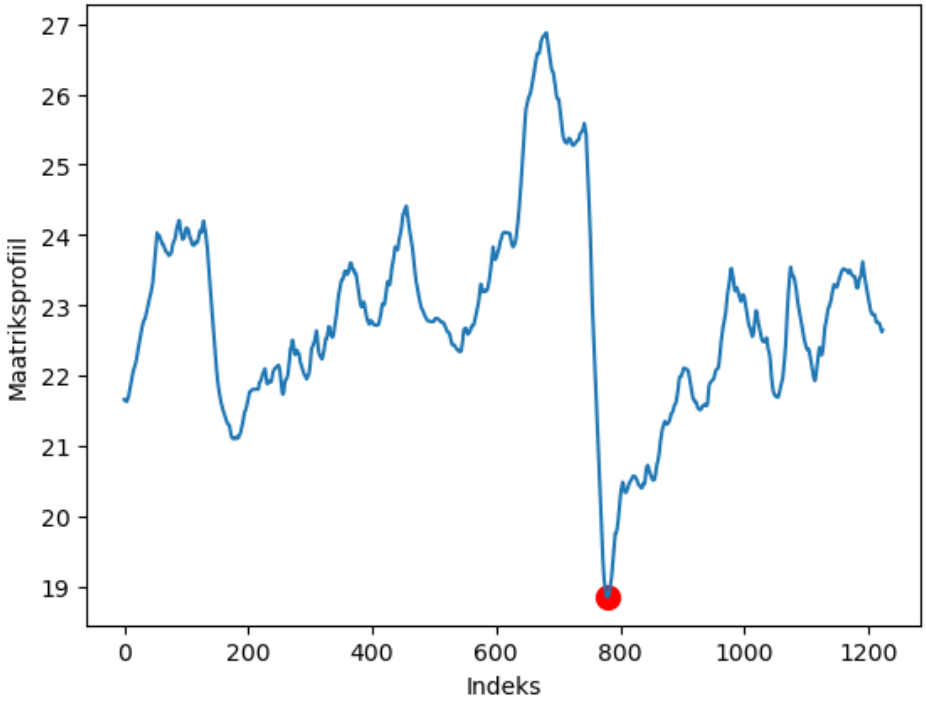
Maatriksprofiil pakub motiivide ja anomaaliate tuvastamiseks erinevaid võimalusi. Näiteks on võimalus motiive ja anomaaliaid otsida etteantud pikkuse järgi nii ühe aegrea alamridasid võrreldes kui ka arvutades kahe aegrea omavahelist kaugust. Antud meetod annab võimaluse avastada seni teadmata või algselt silmale nähtamatuid seoseid aegridades.

Alljärgnevalt on illustreeritud kahe aegrea A ja B kõige sarnasema alamrea ehk motiivi leidmist maatriksprofiili meetodil. Kuigi tegemist on sama patsiendi kopsuhelidega, on helilainetele peale vaadates raske öelda, kas antud helifailid on teineteisega sarnased või mitte.



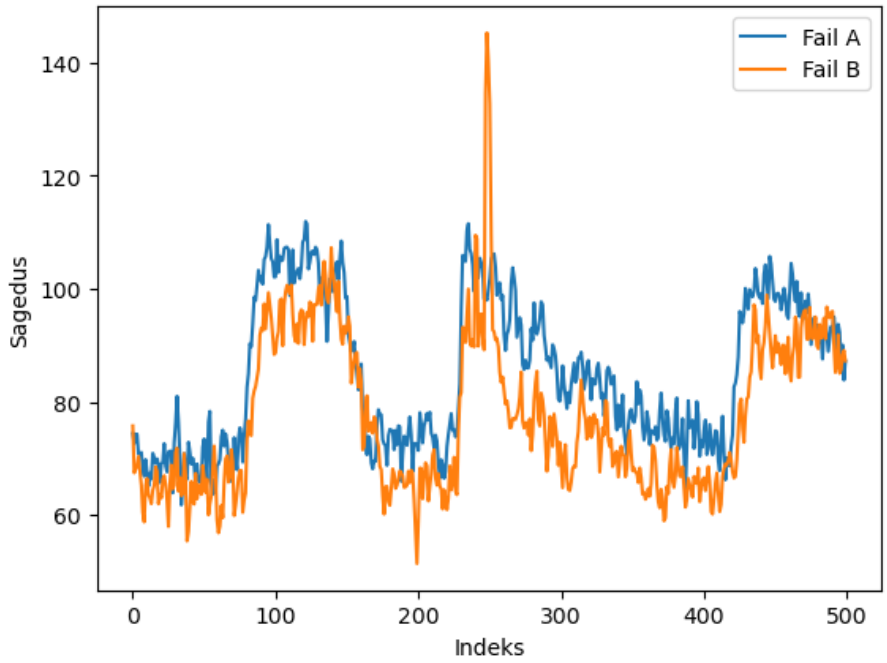
Joonis 4 Kopsuhelide esitus helilainena

Nende failide sarnasuse leidmiseks on arvutatud Mel-sageduse kepstri kordajad (kasutades Pythoni LibROSA teeki) ja arvutatud maatriksprofiil teise Mel-sageduse kepstri kordaja põhjal. Joonisel 5 on punasega märgitud faili A alamrea indeks, mille kaugus lähima naabrini failis B, oli kõige väiksem.



Joonis 5 Kahe helifaili ühise motiivi tuvastamine

Seejärel on kuvatud antud alamlõik failis A ja tema lähim alamlõik failis B joonisel 6, kust näeb juba selgemini nende kahe alamrea sarnasust.



Joonis 6 Kahe helifaili ühine motiiv

6 Anomaaliate ja motiivide tuvastamine

Antud peatükk annab ülevaate maatriksprofiili meetodi rakendamisest erinevate kopsuhaiguste helifailidel eesmärgiga leida kopsuhelides esinevad anomaaliad ja motiivid.

Näiteks on võimalikeks anomaaliateks ja motiivideks kopsuhelides esinevad üldjuhul patoloogiale viitavad helid – vilinad ja räginad, mida kirjeldakse alampeatükis 6.1. Maatriksprofiili meetodi rakendamisel kasutati Pythoni teeki STUMPY, mille ülevaade asub alampeatükis 6.2.

6.1 Hingamishelid

6.1.1 Vilinad

Vilin on muusikaline kõrge heli, mis tekib obstruktsiooni korral sisse- või väljahingamise ajal. Seda tüüpi heli on omane mistahes hingamisteede haigusseisundile, mis põhjustab hingamisteede obstruktsiooni, sh astma, bronhioliit, bronhiektaasia tõbi, kopsupõletik, alumiste hingamisteede haigusseisundid ning krooniline obstruktiivne kopsuhaigus. [24]

6.1.2 Räginad

Rägin on katkendlik mittemuusikaline ragisev heli, mis tekib enamasti sissehingamisel, kuid vahel ka väljahingamisel. Sõltuvalt rägina tüübist võib see olla tuleneda kas õhu liikumisest läbi lima väikestes bronhides või tuleneda bronhiektaasist [13].

Räginad võivad esineda nii bronhiektaasiatõve, krooniline obstruktiivse kopsuhaiguse kui kopsupõletiku korral [11].

6.2 STUMPY

Aegridade analüüsimist maatriksprofiili meetodil võimaldab Pythoni teek STUMPY. [25].

Antud lõputöös kasutatud keskseks meetodiks on selles sisalduv STOMP algoritmil baseeruv *stumpy.stump()* meetod, mis võimaldab arvutada maatriksprofiili nii ühel aegreal kui võrrelda omavahel kahte erinevat aegrida.

STOMP algoritm on üks kiiremaid ja väikseima mäluresursi vajadusega maatriksprofiili arvutamise algoritme vajamata sealjuures keerulist parameetrite tuunimist. [26]. Samuti ei mõjuta m parameetri väärtuse muutmine algoritmi kiirust [20].

stump() meetodi peamisteks parameetriteks on:

- T_A – aegrida, mille põhjal maatriksprofiili arvutama hakatakse
- m – akna suurus
- ignore_trivial – tõeväärtus, mis määrab, kas teostatakse self-join ühel aegreal või võrreldakse omavahel kahte aegrida. Vaikimisi väärtus *true*
- T_B – juhul kui soovitakse võrrelda kahte aegrida

Meetod tagastab iga võrreldava alamrea kohta nelja elemendiga listi, kus on info tema kauguse kohta lähimast naabrist, alamrea enda indeks, lähima vasakpoolse naabri indeksi, lähima parempoolse naabri indeksi kohta.

```
[0.43134577125812795 1294 46 1294]
```

6.3 Parameetrite valimine

Erinevat tüüpi räginate tüüpiline pikkus on vahemikus 5-15 ms ning vilinad esinevad harilikult pikemalt kui 100ms [1]. Et vilinate keskmist pikkust pole kirjanduses täpselt välja toodud, arvutati keskmine andmestikus esinev vilina pikkus andmestikuga kaasas olnud tekstifailide põhjal, kus iga vilina ja rägina kohta oli toodud ajavahemik, millal nad esinesid. Keskmine vilina pikkus andmestikus varieerus diagnoositi vahemikus 180 ms – 650 ms.

M parameetri väärtustamisel testiti erinevaid suurusi. Vilinate otsimisel anomaaliatest ja motiividest kasutati akna suurust, mis vastas antud diagnoosi keskmise vilina pikkusele. Räginate tuvastamisel toimiti sama moodi.

6.4 Töö käik ja tulemused

6.4.1 Motiivide ja anomaaliate seostamine vilinate ja räginatega

Aegrea esituseks valiti Mel-sageduse kepstri kordajad, sest maatriksprofiili meetod on varasemalt antud esituse korral tuvastanud edukalt helifailide sarnasust [20] ning antud esituse korral on võimalik kordajaid vaadata ka teineteisest sõltumatult.

Multidimensionaalse esituse korral oleks analüüs olnud liialt ajakulukas ning maatriksprofiili meetod on näidanud häid tulemusi aegridadel, mille analüüsimisel piirduti ühe dimensiooniga [20].

Motiivide ja anomaaliate leidmiseks erineva diagnoosiga patsientide kopsuhelist toimiti järgnevalt:

1. Võeti kõik antud diagnoosi kuuluvad helifailid ja ühendati need üheks pikaks tervikuks. Erandiks olid KOK diagnoosiga patsiendid, kelle helifaile oli andmebaasis nii ulatuslikul määral, et nende kokku liitmine ja analüüsimine oleks olnud väga ajakulukas. Seetõttu valiti juhuslikkuse alusem 35 (võrdselt tervete patsientide helifailidega) KOK diagnoosi sisaldavat helifaili ja liideti need omavahel kokku. Lisaks välistati astma (1 fail) ja alumiste hingamisteede haigustega patsientide helifailid (2 faili), sest neid esines andmestikus väga vähe. Selle sammu tulemusena tekkis 6 diagnoosipõhist pikka aegrida.

2. Arvutati igal aegreal maatriksprofiil kasutades *stumpy.stump()* funktsiooni. Testiti erinevaid m parameetri väärtuseid vastavalt antud diagnoosi helifailide keskmise vilina ja rägina pikkusele. Lisaks jooksutati algoritmi eraldi iga Mel-sageduse kepstri kordaja jaoks.

3. Leiti iga Mel-sageduse kepstri kordaja ja m parameetri väärtuse kombinatsiooni peal tehtud arvutustest 5 parimat motiivi (alamread, mille lähim naaber asus lähedal) ja 5 kõige potentsiaalsemat anomaaliat (alamread, mille lähim naaber asus kaugel). Anomaaliate ja motiivide leidmisel välistati failide üleminekukohad 1 sekundi ulatuses. Samuti seati kriteerium, et kõik tuvastatud anomaaliad paikneksid kõigist teistest anomaaliatest vähemalt 3 sekundi kaugusel (ligikaudne ühe hingamistsükli pikkus). Sama kehtis ka motiivide

korral. Sellega välditi olukorda, kus tuvastatud mustrid või anomaaliad asuvad väga lähestikku või kattuvad eelmise tuvastatud anomaalia või motiiviga.

4. Leiti antud alamridade asukoht failis ning kontrolliti, kas antud alamrida langes kokku vilina või rägina esinemisega.

Tabelis 2 on toodud anomaaliate ja motiivide tuvastamisel kasutatud helifailide hulk diagnoosi kohta ja vilinate ja räginate arv ühes ajaühikus antud diagnooside failide kogupikkusest.

Tabel 2 Anomaaliate ja motiivide tuvastamisel kasutatud helifailide karakteristikud

Diagnoos	Analüüsitavate failide arv	Vilinate arv ajaühikus (s)	Räginate arv ajaühikus (s)
KOK	35	0.14	0.62
Terve	35	0.004	0.04
Kopsupõletik	37	0.03	0.05
URTI	23	0.02	0.09
Bronhiektaasiatõbi	16	0.215	0.175
Bronhioliit	13	0.33	0.125

Räginate esinemist tuvastatud motiivide ja anomaaliate hulgas on esitatud Joonisel 7, kus A_ eesliide tähistab anomaalset räginat ja M_ eesliide motiivide hulgas esinenud räginat.

MFCC	M_KOK	A_KOK	M_Terve	A_Terve	M_Kopsupõletik	A_Kopsupõletik	M_URTI	A_URTI	M_Bronhiekt	A_Bronhiekt	M_Bronhioliit	A_Bronhioliit
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0
16	1	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	1	0
18	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	1	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0

Joonis 7 Räginate esinemine 5 parima motiivi ja 5 kõige tõenäolisema anomaalia hulgas

Räginate otsimisel minimaalse akna pikkuse korral (m=3) ei näidanud algoritm edu mitte ühegi diagnoosi ega Mel-sageduse kepstri kordaja korral, kuigi võrreldava aegrea pikkusele vastab ~35 ms pikkune lõik helifailist (ligikaudne keskmine rägina pikkus) ja tegemist võiks seega olla räginate tuvastamisel optimaalseks parameetri m väärtusega.

Kuigi räginaid esines helifailides peaaegu kõigis diagnoosides vilinatest rohkem, langes üksikutel juhtudel mõni tuvastatud anomaalia või motiiv räginate esinemise ajahetkele. See võib viidata, et maatriksprofiil käsitles helifailides rohkelt esinenud räginaid ühtlase taustamürana ning ei pidanud neid olulisteks motiivideks ega anomaaliateks. Lisaks võiks tulevikus täiendavalt uurida, kas mitmedimensioonilise helifaili esituse kasutamine parandaks algoritmi suutlikkust räginate tuvastamisel.

Vilinate otsimisel väärtustati m vastavalt diagnoosi keskmisele vilina pikkusele. Tulemused on toodud Joonisel 8, kus A_ eesliide tähistab anomaalset vilinat ja M_ eesliide motiivide hulgas esinenud vilinat.

MFCC	M_KOK	A_KOK	M_Terve	A_Terve	M_Kopsupõletik	A_Kopsupõletik	M_URTI	A_URTI	M_Bronhiekt	A_Bronhiekt	M_Bronhioliit	A_Bronhioliit
2	0	1	0	0	0	0	0	0	0	0	0	1
3	0	2	0	0	0	0	0	0	0	0	3	0
4	0	1	0	0	0	1	0	0	0	1	0	0
5	0	1	0	0	0	0	0	0	0	1	1	0
6	0	0	0	0	0	0	0	0	0	0	1	2
7	0	0	0	0	0	0	0	0	2	0	1	0
8	1	1	0	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	1	0	0	0	0	0	1	0	1	1	1	0
11	0	1	0	0	0	0	0	0	1	0	0	0
12	2	0	0	0	0	0	0	0	0	0	0	0
13	1	1	0	0	1	0	0	0	0	0	0	1
14	4	0	0	0	0	0	0	0	0	1	0	0
15	2	1	0	0	0	0	0	0	0	1	1	0
16	5	0	0	0	0	0	0	0	0	0	1	1
17	4	0	0	0	1	0	1	0	3	1	1	1
18	3	1	0	0	0	0	1	0	1	1	2	1
19	4	1	0	0	2	0	1	0	0	0	0	0
20	5	0	0	0	1	0	1	0	0	1	0	1

Joonis 8 Vilinate esinemine 5 parima motiivi ja 5 kõige tõenäolisema anomaalia hulgas

Enim motiive langes kokku vilinatega KOK diagnoosi korral, kuid ka vilinate esinemise arv antud diagnoosis oli suurim. Tervete patsientide kopsuhelidest ei tuvastatud algoritm vilinaid ega räginaid ei anomaaliate ega motiividena. Lisaks on Joonisel 8 näha, et kõrgemad Mel-sageduse kepstri kordajad olid üldjuhul vilinate tuvastamisel edukamad.

Tulemused ei näita, et kindla pikkusega alamridade võrdlemine kasutades Mel-sageduse kepstri kordajaid, on edukas viis kopsuhelidest vilinate ja räginate tuvastamiseks. Üheks võimalikuks põhjuseks võib olla failides esinev müra. Kuigi lähtuvalt kopsuhelide karakteristikutest keskenduti sagedusvahemikule 50-4000 Hz, võivad kopsuhelid vajada enne analüüsimist täiendavat eeltöötlust müra eemaldamiseks. Tulevikus võiks testida, kas maatriksprofiili meetodi rakendamisele eelnevalt töödeldud kopsuhelidele parandab oluliselt meetodi tulemuslikkust vilinate ja räginate tuvastamisel.

Lisaks võib meetodi tulemuslikkust suurendada erinevate aegrea esituste katsetamine (nt kasutada Mel-sageduse kepstri kordajate asemel esitust spektrogrammina). Kui

otsitav motiiv on ette teada (nt kindlat tüüpi vilin või rägin) võib sobilik maatriksprofiili meetod olla ka motiivi otsimine etteantud alamrea (*query*) abil.

7 Klasterdamine

Järgnev peatükk keskendub kopsuhelidest klastrite e gruppide tuvastamisele tuginedes helifailide vahelisele kaugusmõõdule, mis on arvutatud maatriksprofiili meetodil.

7.1 DBSCAN

Klasterdamisel kasutati tiheduspõhist klasterdamisalgoritmi DBSCAN, sest antud algoritm tuleb hästi toime mürarikkate andmetega ning on suuteline tuvastama ka mittekorrapärase kujuga klastreid. Lisaks ei vaja DBSCAN parameetrina otsitavate klastrite hulka [27]. Lisaks osutus antud meetodi valikul määravaks ka DBSCANi võimekus töötada ka ainult kaugusmaatriksiga, erinevalt näiteks K-means meetodist, mis eeldab andmetabelit.

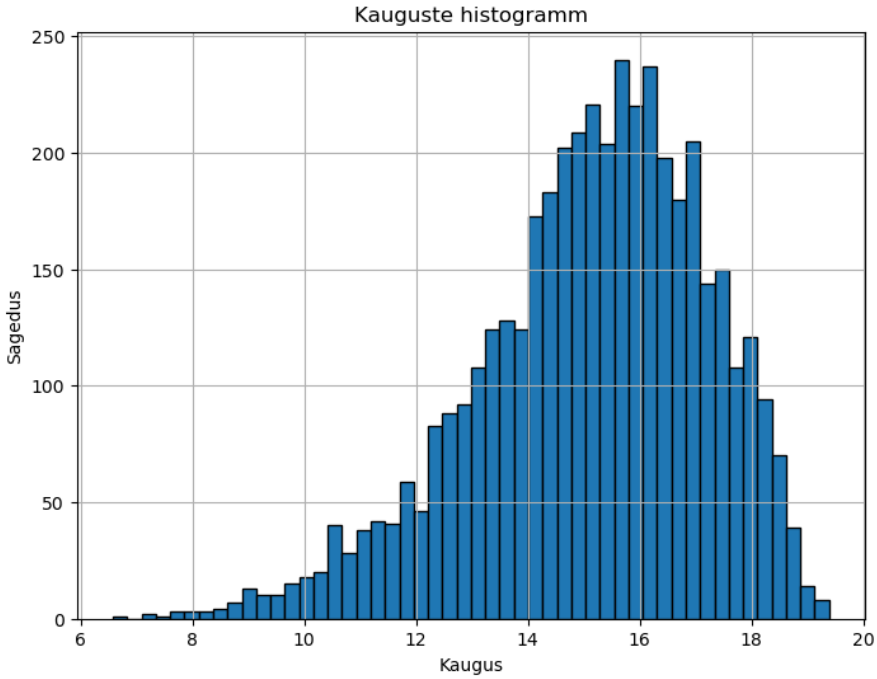
DBSCAN olulised parameetrid on:

- Eps – maksimaalne kaugus kahe elemendi vahel, et neid käsitletaks naabritena
- Min_samples – põhipunkti defineerimiseks vajalik naabrite arv. Kõrgema väärtuse korral leitakse tihedamad klastrid, madalamate väärtuste korral on klastrid hõredamad
- Metric – vaikimisi *euclidean*, aga *precomputed* võimaldab klasterdada kasutades sisendina kaugusmaatriksit

7.2 Optimaalsete hüperparameetrite leidmine

DBSCANi efektiivsus sõltub optimaalsete hüperparameetrite väärtuste leidmisest. Valesti valitud hüperparameetrid viivad tulemusteni, kus tekib vaid üks klaster või kus kõik punktid klassifitseeritakse kui müra [27].

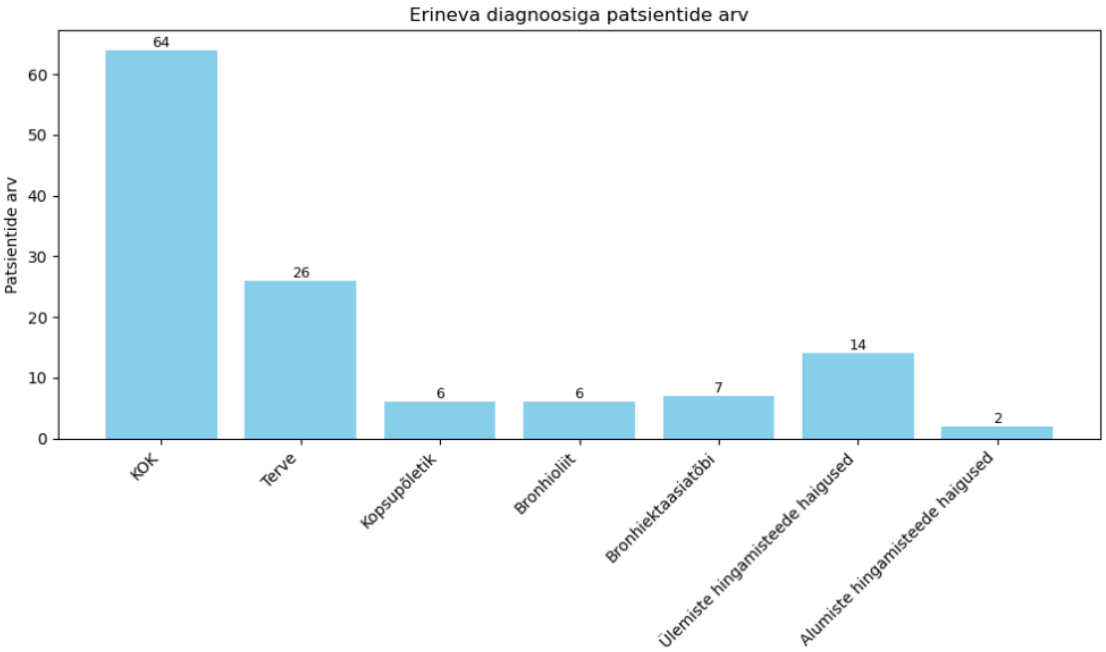
Optimaalsete hüperparameetrite väärtuste leidmisel tugineti domeeniteadmistele ja katsetati erinevate väärtustega, et leida parim tulemus. Epsiloni väärtuse valikul andis informatsiooni ka kauguste histogrammi analüüsimine (Joonis 9)



Joonis 9 Kauguste histogramm optimaalse eps väärtuse hindamiseks

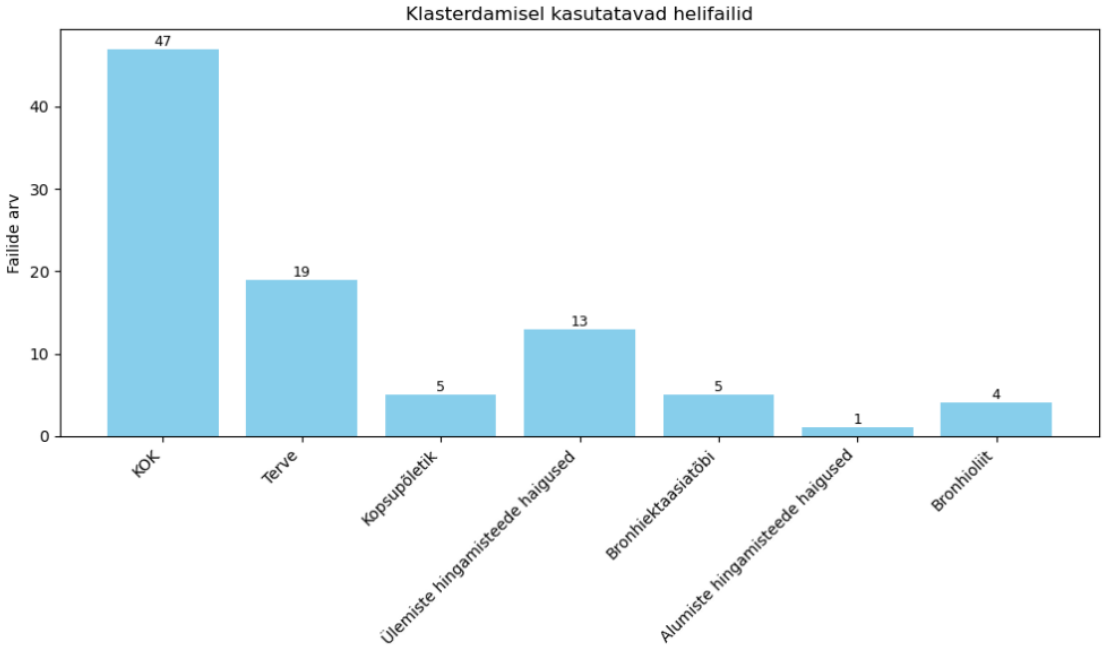
7.3 Kaugusmaatriksi koostamine

Andmestiku jaotus erineva diagnoosiga patsientide osas on väga ebaühtlane (Joonis 10)



Joonis 10 Diagnooside jaotus kasutatavas andmestikus

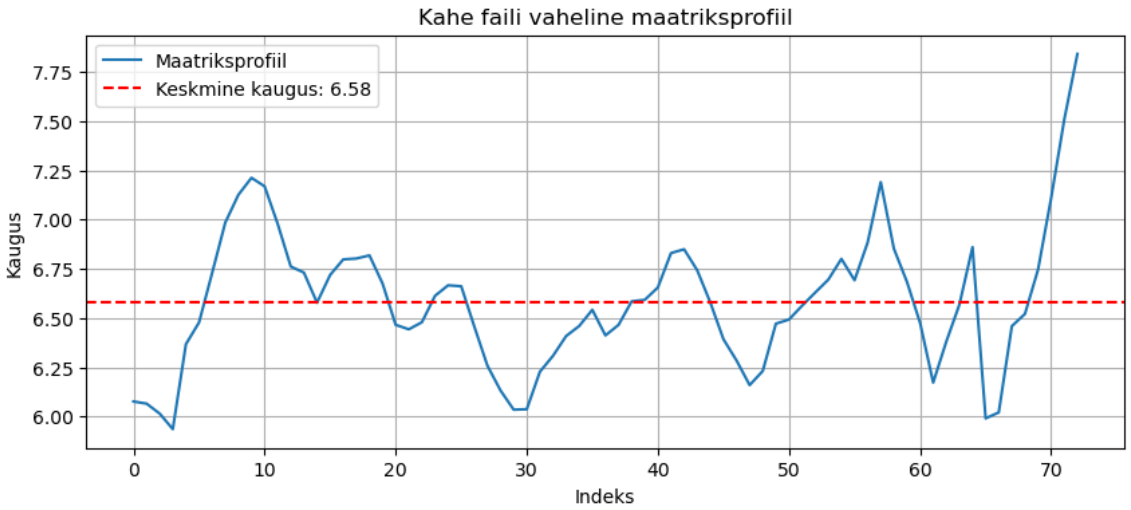
Seetõttu valiti patsiendid nii, et KOK diagnoosiga ja muu diagnoosiga (sh tervete) patsientide hulk andmestikus oleks võrdne. Samuti oli oluline kriteerium, et helifailide kvaliteet oleks sama. Nende kriteeriumite tõttu koosnes lõplik kasutatav andmestik 47 KOK diagnoosiga patsiendi helifailist ja 47 muu diagnoosiga patsiendi helifailist. Iga patsiendi kohta valiti juhuslikkuse alusel 1 helifail. Kokku oli klasterdatavas andmestikus 94 helifaili. Joonisel 11 on välja toodud klasterdamisel kasutatavate failide jaotus.



Joonis 11Erineva diagnoosiga kopsuhelide arv klasterdatavate andmete hulgas

Seejärel arvutati iga failipaari kohta nende vaheline kaugusmõõt kasutades selleks STOMP algoritmil baseeruvat *stumpy.stump()* meetodit, kus võrreldavate aegridadena kasutati võrreldavate audiofailide erinevaid Mel-sageduse kepstri kordajaid.

Lähtuvalt maatriksprofiili meetodi olemusest, arvutati faili A iga m pikkuse alamrea kaugus igast faili B m pikkusega alamrreast ja säilitati kaugusmõõt iga alamrea lähima naabri kohta. Neist kaugusmõõtudest võeti aritmeetiline keskmine ning saadud numbrit käsitleti kui nende failide vahelist kaugust.



Joonis 12 Kahe faili vaheline maatriksprofiil koos keskmise kaugusmõõduga

Kõigi failipaaride vaheliste kauguste põhjal komplekteeriti kaugusmaatriks, mida kasutati parameetrina *dbscan.fit_predict()* meetodis, kus metric väärtus seati *precomputed*. Parima tulemuse leidmiseks katsetati kaugusmaatriksi komplekteerimist erinevate maatriksprofiili arvutamise parameetrite korral. Testiti erinevaid Mel-sageduse kepstri kordajaid ja akna suurust.

7.4 Uurimisküsimused

Klasterdamisel püstitati järgnevad uurimisküsimused:

3) Kuidas kopsuhelid rühmituvad ehk klasterduvad?

4) Kas tekivad selgelt eristuvad klastrid, mis on seotud erinevate diagnoosidega?

    - Kui klastrid ei ole seotud diagnoosidega, siis milline on loogika tekkinud klastrite taga?

7.5 Tulemused ja analüüs

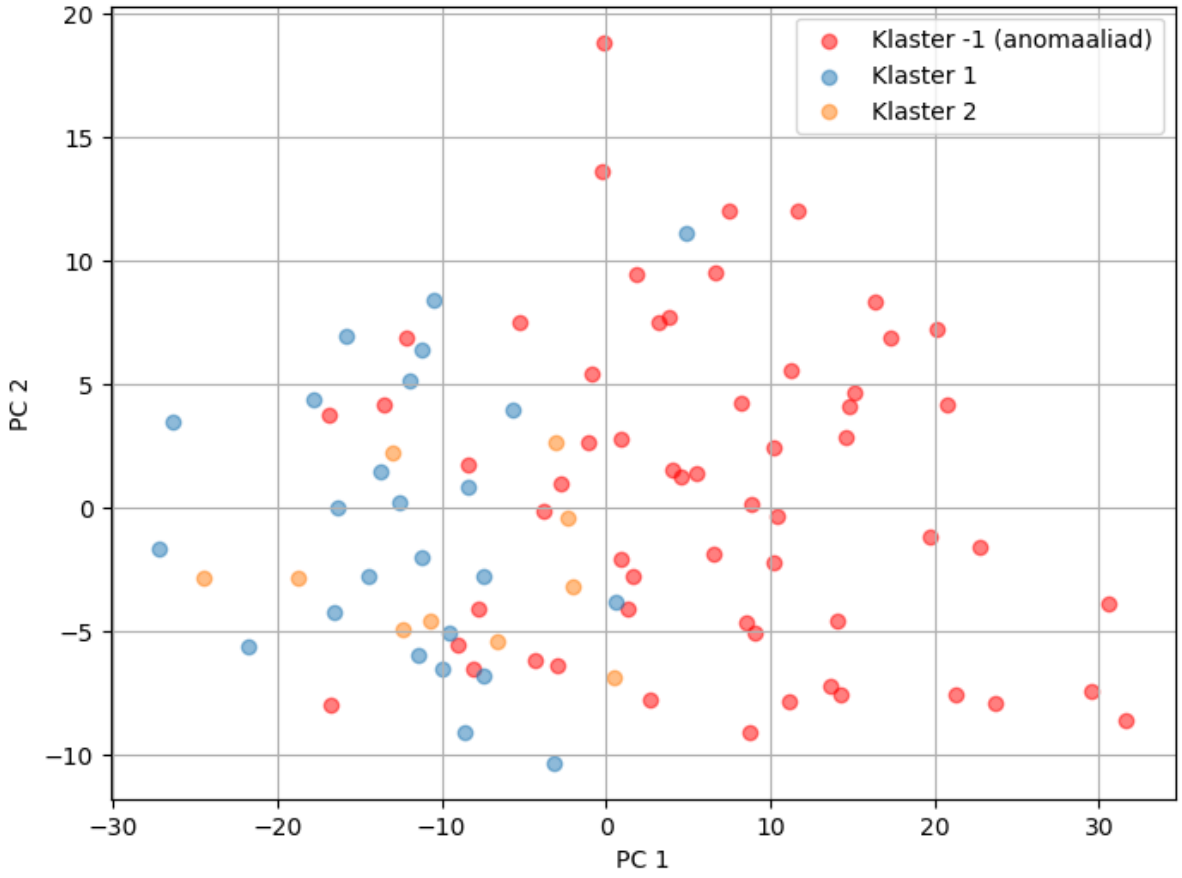
Siin peatükis kirjeldab töö autor klasterdamise tulemusi ja annab vastused eelnevas alampeatükis püstitatud uurimisküsimustele.

Selgeima klasterduse andsid järgmised hüperparameetrite kombinatsioonid

- Eps = 10.75 ja min_samples = 22
- Teisel Mel-sageduse kepstri kordajal teostatud maatriksprofiili arvutuste põhjal (m = 250, mis vastab ligikaudselt ühele hingamistsükli pikkusele helifailis) koostatud kaugusmaatriks

Järgnevalt on kirjeldatud ja visualiseeritud antud hüperparameetrite väärtustega teostatud klasterdamise tulemusi.

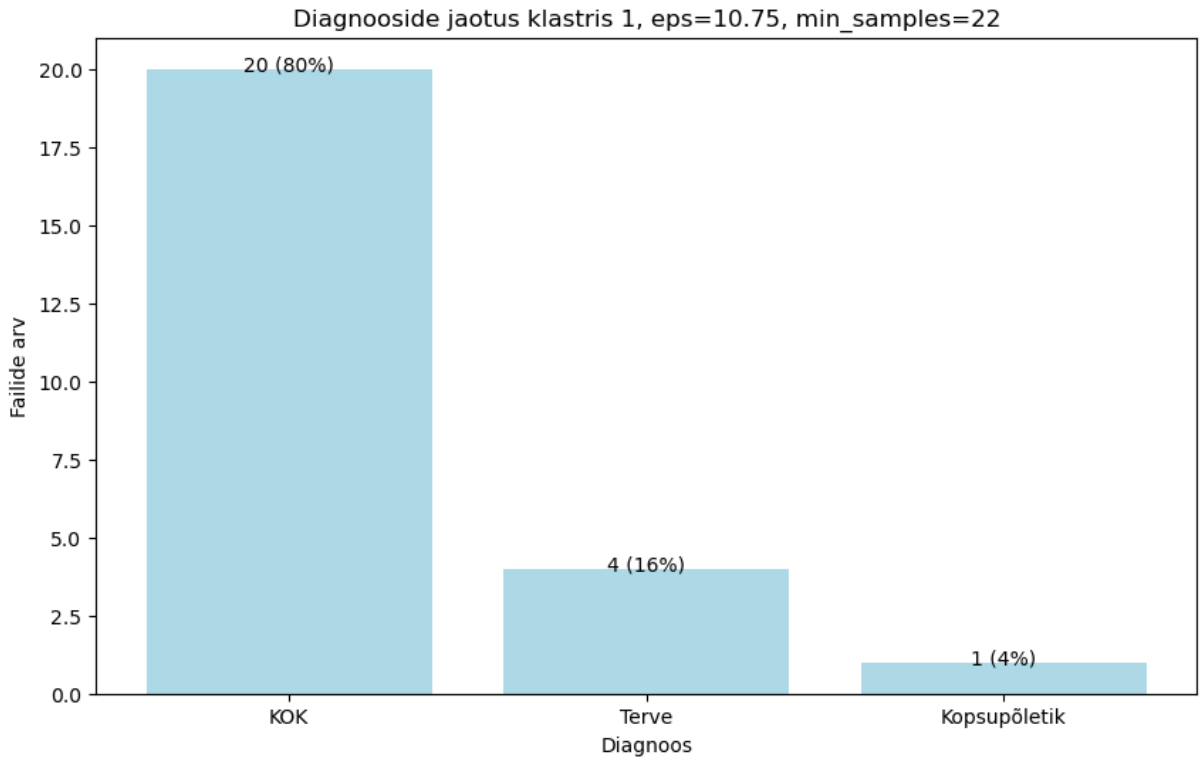
Klasterdamise tulemusel tekkis kaks klastrit ja üks grupp failidest, mis ei kuulunud ühtegi klastrisse.



Joonis 13 Klastrite visuaalne esitus

7.5.1 Klaster 1

Esimeses klastris oli 25 helifaili ning diagnooside jaotus on kujutatud joonisel 14.



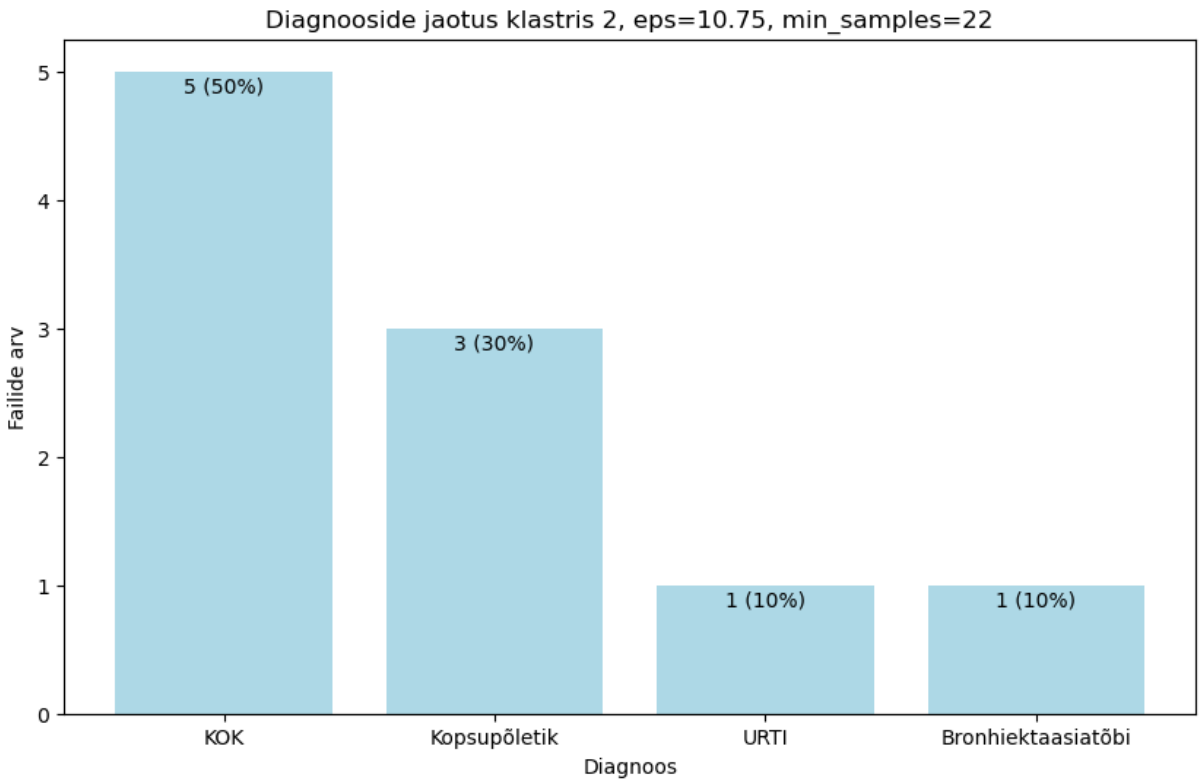
Joonis 14 Klaster 1 – diagnooside jaotus

Keskmine vilinate arv 1. klastris esinenud failides: 0.24

Keskmine räginate arv 1. klastris esinenud failides: 0.05

7.5.2 Klaster 2

Teine klaster koosnes 10 helifailist ning diagnooside jaotus klastris on esitatud Joonisel 15.



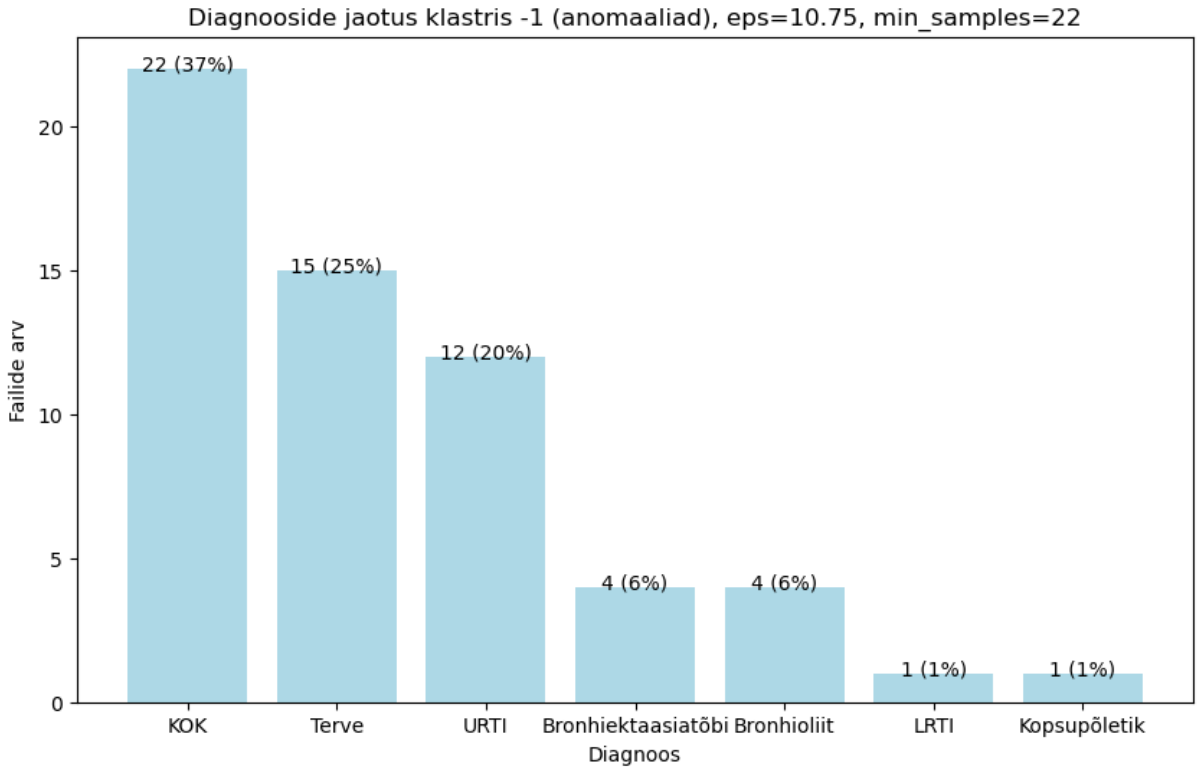
Joonis 15 Klaster 2 – diagnooside jaotus

Keskmine vilinate arv klastris esinenud failides: 0.23

Keskmine räginate arv klastris esinenud failides: 0.04

7.5.3 Ilma klastrita

Järgnevalt on Joonisel 16 välja toodud informatsioon helifailide kohta, mis ei kuulnud antud parameetrite korral ühtegi tuvastatud klastrisse.



Joonis 16 Klaster -1 (anomaaliad) – diagnooside jaotus

Keskmine vilinate arv klastris esinenud failides: 0.14

Keskmine räginate arv klastris esinenud failides: 0.11

7.5.4 Klastrite seos erinevate diagnoosidega ja loogika tekkinud klastrites

Esimesene klaster sisaldas 25 helifaili ja koosnes 80% ulatuses kroonilise obstruktiivse kopsuhaiguse helifailidest. Sellest tulenevalt võib antud klastrit seostada KOK diagnoosiga.

Teises klastris paiknes 10 faili, mille hulgas oli 5 KOK diagnoosiga patsiendi helifaili ja 5 muu kopsuhaigusega patsiendi helifaili. Kuigi siin on raske üheseid järeldusi teha, millise diagnoosiga antud klastrit seostada võib, väärib märkimist, et 60% kõigist kopsupõletiku helifailidest paiknes siin klastris.

7.5.4.1 Hii-ruut test

Hii-ruut testiga kontrolliti nullhüpoteesi, et muutujate klaster ja diagnoos vahel puudub statistiliselt oluline seos.

Hii-ruut testi rakendamiseks koostati sagedustabelid kõigi diagnooside esinemise kohta klastrites. Testi tulemused on näidatud Tabelis 3.

Tabel 3 P väärtus

Kaasatud klastrid	p-väärtus
1,2 ja -1	0,001
1,2	0,019

Mõlemal juhul näitab tulemus, et tekkinud klastrite ja klastrites esinenud diagnooside vahel on statistiliselt oluline seos, lükates sellega ümber nullhüpoteesi.

Väike p väärtus näitab, et diagnooside jaotumine klastrite vahel ei ole juhuslik, vaid toetub olulistele diagnooside vaheliste seoste tuvastamisele ning valideerib, et maatriksprofiili arvutustel põhinev kaugusmõõt on relevantne viis erinevate diagnooside tuvastamiseks.

7.6 Kokkuvõte

Antud alampeatükk võtab kokku klasterdamise tulemused ja annab soovitused edasisteks uurimusteks.

7.6.1 Tulemused

DBSCAN algoritmi kasutamisel on väga oluline hüperparameetrite väärtuste valimine. Valesti valitud väärtused mõjutavad oluliselt tulemusi.

Parimate hüperparameetrite väärtuste korral oli algoritm suuteline eristama kroonilist obstruktiivset kopsuhaigust teistest kopsuhaigustest. Statistiliselt olulist seost diagnooside ja tekkinud klastrite vahel näitab hii-ruut testi tulemus $p=0.001$ (< 0.05).

Antud tulemus näitab, et maatriksprofiili meetod mõõdab kopsuhaiguste tuvastamiseks olulisi aspekte ning antud meetod võib anda häid tulemusi erinevate kopsuhaiguste diagnoosimisel. Antud väidet tuleks siiski täiendavate uuringutega kinnitada.

7.6.2 Soovitused edasisteks uurimusteks

Ühtegi klastrisse ei kuulunud 78% tervete patsientide helifailidest ja 92% ülemiste hingamisteede haigusseisunditega patsientide kopsuhelidest, mis moodustasid klasterdamata failidest kokku 45%. Kuna ülemiste hingamisteede haigused ei mõjuta otseselt kopse ning ka klasterdamata helifailide grupi keskmine vilinate arv failis oli teistest tekkinud klastritest väiksem (Tabel 4), siis võiks tulevastes uurimustes täiendavalt analüüsida, kas algoritm võib jätta klasterdamata helifailid, mis pärinevad kas tervete patsientide või väheste haigussümptomitega kopsuhelidest.

Kuigi ka 46% kroonilise obstriivse kopsuhaiguse failidest kuulus antud gruppi, tuleb meeles pidada, et andmestikus ei ole kroonilise obstruktiivse kopsuhaigusega patsientide haiguse ägedust kirjeldatud ning ka KOK diagnoosiga patsientide helifailide hulgas esineb faile, kus ei ole räginaid ega vilinaid.

Tabel 4 Keskmine vilinate ja räginate arv klastri kohta

Klaster	Keskmine vilinate arv	Keskmine räginate arv
0	0.24	0.05
1	0.23	0.04
-1 (ilma klastrita)	0.14	0.11

Klasterdamistulemuste parandamisel võiks olla kasu võimalusest kaasata iga diagnoosi kohta rohkem helifaile.Samas tuleb ka arvestada, et iga faili lisandumisega suureneb aeg ja mälu vajadus, mis kulub maatriksprofiili arvutamisele. Näiteks tuleb 94 failist koosneval andmestikul teha 4371 *AB-joini*. Kui andmestik oleks kaks korda suurem, tuleks võrdlusi teha juba 17578.

8 Kokkuvõte

Lõputöö raames otsiti kopsuhelidest maatriksprofiili meetodit kasutades anomaaliaid ja motiive ning võimalikke klastreid ning nende seoseid kroonilise obstruktiivse kopsuhaiguse jt kopsuhaiguste esinemisega.

Lõputöö praktiline osa keskendus maatriksprofiili meetodi rakendamisele kopsuhelide analüüsimisel. Praktilise töö esimeses osas otsiti antud meetodi abil tuvastatud motiivide ja anomaaliate seoseid kopsuhelides esinevate patoloogiale viitavate helide – vilinate ja räginatega. Helifailide esitusena kasutati Mel-sageduse kepstri kordajaid ning võrreldavate aegridade pikkust kohandati andmestikus esinenud vilinate ja räginate keskmise pikkuse järgi.

Analüüsiti, kas tuvastatud motiivid ja anomaaliad langevad kokku kopsuhelides esinevate patoloogiale viitavate helide – vilinate ja räginatega. Väga selget seost motiivide ja anomaaliate ja huvipakkuvate helide vahel leida ei õnnestunud.

Teine osa praktilisest tööst keskendus kopsuhelide klasterdamisele. Klasterdamiseks kasutati DBSCAN klasterdamisalgoritmi, hüperparameetrite valikul katsetati erinevate väärtustega, mis valiti nii andmestiku karakteristikute kui kirjanduses antud soovituste põhjal.

Maatriksprofiili meetodit kasutati kopsuhelide omavaheliste kauguste leidmiseks. Leitud kauguste põhjal koostati kaugusmaatriks, mida kasutati DBSCANi hüperparameetrina. Helifailide klasterdamisel tekkis 2 klastrit ning 1 klaster anoomaliatega. Klastreid eristas KOK diagnoosiga helifailide erinev esinemissagedus, mida kinnitati ka hii-ruut testiga ($p = 0{,}001$), mis lükkab ümber nullhüpoteesi, et klastrite ja klastrites paiknevate diagnooside vahel puudub statistiliselt oluline seos.

P väärtus $0{,}001 << 0.05$ näitab, et diagnooside jaotumine klastrite vahel ei ole juhuslik, vaid toetub olulistele diagnooside vaheliste seoste tuvastamisele ja valideerib, et maatriksprofiili arvutustel põhinev kaugusmõõt on relevantne viis kopsuhelidest erinevate diagnoose tuvastamiseks. Maatriksprofiili põhise keskmise kaugusmõõdu ja

DBSCANi kombineerimise meetod võib tulevikus huvi pakkuda mistahes aegridade analüüsimisel.

Antud klasterdamismetoodikat võiks tulevikus testida ka suuremal kopsuhelide andmestikul. Samas tuleb meeles pidada, et iga faili lisandumisega suureneb ka aja- ja mäluressursi vajadus, mis kulub maatriksprofiili arvutamisele.

Anomaaliate ja motiivide tuvastamist võiks testida ka eelnevalt töödeldud helifailidel, kus välise müra taset on vähendatud. Lisaks võiks testida, kas mõni alternatiivne aegrea esitus, nt spektrogramm, parandab vilinate ja räginate tuvastamise sagedust.

Kasutatud kirjandus

[1] Andrès, E., Gass, R., Charloux, A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. Journal of Medicine and Life 1 April 2018.

[2] Bahoura, M. Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. Computers in Biology and Medicine, 39(9), 824-843. DOI: 10.1016/j.compbiomed.2009.06.011.

[3] Bennet, S., Bruton, A., Barney, A. The Relationship Between Crackle Characteristics and Airway Morphology in COPD. PMID: 25425707. DOI: 10.4187/respcare.03543.

[4] Dar, J.A., Srivastava, K.K., Mishra, A. Lung anomaly detection from respiratory sound database (sound signals). Computers in Biology and Medicine, 164, 107311.

[5] Eesti Kopsuliit [Võrgumaterjal] KOK. [Kasutatud 23.09.2023]

[6] Guerra, B., Gaveikaite, V., Bianchi, C., Puhan, M.A. Prediction models for exacerbations in patients with COPD. European Respiratory Review, 26(143), 160061. DOI: 10.1183/16000617.0061-2016.

[7] Kim, H., Koh, D., Jung, Y. Breathing sounds analysis system for early detection of airway problems in patients with a tracheostomy tube. Scientific Reports volume 13, Article number: 21029 (2023).

[8] Kim, Y., Hyon, Y., Jung, S.S. Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning. Scientific Reports, 11, 17186.

[9] Law, S. Part 8: AB-Joins with STUMPY. Finding Conserved Patterns Across Two Time Series(https://towardsdatascience.com/part-8-ab-joins-with-stumpy-af985e12e391).

[10] Lu, Y. How to Preprocess Your Time Series. Handling missing data and singular continuous values(https://matrixprofile.org/posts/how-to-preprocess-your-time-series/).

[11] Piirila, P., Sovijärvi, A.R.A. Crackles: recording, analysis and clinical significance. December 1995 European Respiratory Journal 8(12).

[12] Pita, A., Rodriguez, F. ja navarro, J.M. Cluster Analysis of Urban Acoustic Environments on Barcelona Sensor Network Data. Int J Environ Res Public Health, 2021.

[13] Sarkar, M., Madabhavi, I., Niranjan, N. Auscultation of the respiratory system. Annals of Thoracic Medicine, 1 July 2015.

[14] Soriano, J.B., Polverino, F. ja Cosio B.G. What is early copd and why is it important? European Respiratory Journal, 52(6), 2018.

[15] Srivastava, A., Jain, S., Miranda, R., Patil, S. Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. PeerJ Computer Science, 11 February 2021.

[16] The Matrix Profile [Võrgumaterjal] (https://stumpy.readthedocs.io/en/latest/Tutorial_The_Matrix_Profile.html). [Kasutatud 14.10.2023]

[17] Wang, C., Chen, X., Du, L. Comparison of machine learning algorithms for the identification of acute exacerbations in chronic obstructive pulmonary disease. Computer Methods and Programs in Biomedicine, 188, 105267.

[18] Williams, V., Hardinge, M. , Ryan, S ja Farmer, A. Patient’s experience of identifying and managing exacerbations in chronic obstructive pulmonary disease. PeerJ Computer Science, 2021.

[19] World Health Organization. [Võrgumaterjal] The top 10 causes of death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. [Kasutatud 12.09.2023]

[20] Yeh, C.M., Zhu, Y., Ulanova, L. Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. 2016 IEEE 16th International Conference on Data Mining (ICDM).

[21] ICBHI 2017 Challenge. Respiratory sound database [Võrgumaterjal]. [Kasutatud 24.12.2023]

[22] Jupyter Notebook Documentation [Võrgumaterjal] (https://docs.jupyter.org/en/latest/#what-is-a-notebook) [Kasutatud 23.09.2023]

[23] Python Documentation [Võrgumaterjal] (https://docs.python.org/3/faq/general.html#what-is-python) [Kasutatud 23.09.2023]

[24] Patel, P. H., Mirabile, V.S., Sharma, S. Wheezing. StatPearls, 2023.

[25] Law, S.M. STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining. Journal of Open Software, 2019

[26] Zhu, Y., Zimmermann, Z. et al. Matrix profile II: Exploiting a Novel Algorithm and GPUs to Break the One Hundred Million Barrier for Time Series Motifs And Joins. 2016 IEEE 16th International Conference on Data Mining (ICDM).

[27] Ester, M., Kriegel, H.-P. et al. A Density-Based Algorithm for Discovering Clusters in Large Spacial Databases with Noise. 2nd International Conference on Knowledge Discovery and Data Mining

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks[1]

Mina, Helen Tamm

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Hingamishelide andmeanalüüs maatriksprofiili meetodil", mille juhendaja on Ants Torim
   1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
   1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

10.01.2024

1 Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.