

TARTU ÜLIKOOL  
Loodus- ja täppisteaduste valdkond  
Arvutiteaduse instituut  
Informaatika õppekava

Aleksis Krass

Parkinsoni tõve tuvastamine eestikeelsete  
hääleklippide analüüsi abil kasutades  
masinõppe meetodeid

Bakalaureusetöö (9 EAP)

Juhendajad: Joonas Järve, MSc  
Pille Taba, PhD

Tartu 2025

## **Parkinsoni tõve tuvastamine eestikeelsete hääleklippide analüüsi abil kasutades masinõppe meetodeid**

### **Lühikokkuvõte:**

Käesolev bakalaureusetöö uurib masinõppemeetodite rakendatavust Parkinsoni tõve (PT) tuvastamisel eestikeelsete hääleklippide põhjal. Töö keskendub kolmele uurimisküsimusele: esiteks hinnatakse hispaaniakeelsel PC-GITA andmestikul treenitud akustilistel tunnustel põhineva mudeli üldistusvõimet eestikeelsetele andmetele; teiseks uuritakse, kas hispaania- ja eestikeelsete andmete kombineerimine treeningfaasis parandab mudeli jõudlust; kolmandaks testitakse mujal peenhäälestatud tippasemel isejuhendatud õppel (SSL) põhineva WavLM Base mudeli otserakendatavust eestikeelsetel andmetel. Tulemused näitavad, et akustilistel tunnustel põhinevate mudelite otsene ristkeelne ülekantavus on piiratud, kuid andmestike kombineerimine parandab oluliselt jõudlust kuni 0.7893. Eelnevalt peenhäälestatud SSL-mudeli otserakendamine eestikeelsetele lühikestele helilõikudele ilma täiendava kohandamiseta ei andnud edukaid tulemusi. Töö rõhutab vajadust keespetsiifilise kohandamise ja mitmekeelsete andmekogude kasutamise järele PT häälepõhises tuvastamises.

### **Võtmesõnad:**

Parkinsoni tõbi, masinõpe, audio, hääleanalüüs, süvaõpe, siirdeõpe, akustilised tunnused, eesti keel

**CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)**

# **Parkinson's Disease Detection from Estonian Voice Clips using Machine Learning Methods**

## **Abstract:**

This Bachelor's thesis investigates the applicability of machine learning methods for Parkinson's Disease (PD) detection using Estonian voice clips. The research focuses on three main questions: firstly, evaluating the generalizability of an acoustic feature-based model trained on the Spanish PC-GITA dataset to Estonian data; secondly, examining whether combining Spanish and Estonian data during training improves model performance; and thirdly, testing the direct applicability of a state-of-the-art self-supervised learning (SSL) based WavLM Base model, fine-tuned elsewhere, on Estonian data. The results indicate that the direct cross-lingual transferability of acoustic feature-based models is limited, but combining datasets significantly improves performance up to 0.7893. The direct application of a pre-fine-tuned SSL model on short Estonian audio segments without further adaptation was not successful. The thesis highlights the need for language-specific adaptation and the use of multilingual datasets in voice-based PD detection.

## **Keywords:**

Parkinson's Disease, Machine Learning, Audio, Voice Analysis, Deep Learning, Transfer Learning, Acoustic Features, Estonian Language

**CERCS: P170 Computer science, numerical analysis, systems, control**

# Sisukord

<b>1</b>	<b>Sissejuhatus</b>	<b>6</b>
1.1	Vajalikud eelteadmised . . . . .	7
1.1.1	Parkinsoni tõbi . . . . .	7
1.1.2	Masinõpe (ML) . . . . .	8
1.1.3	Heli ja kõne akustiline analüüs . . . . .	8
<b>2</b>	<b>Metoodilised Suunad Kirjanduses</b>	<b>9</b>
2.1	Klassikalised meetodid: akustilised tunnused ja klassikaline masinõpe .	10
2.2	Süvaõppe meetodid: CNNidest isejuhendatud mudeliteni . . . . .	11
2.3	Andmestikud, keelteülene rakendatavus ja robustsus . . . . .	12
2.4	Hindamismetoodikad . . . . .	13
2.5	Süntees ja tehnilised uurimislüngad . . . . .	14
<b>3</b>	<b>Andmed</b>	<b>15</b>
3.1	Kasutatud andmestikud . . . . .	15
3.1.1	Eestikeelne andmestik . . . . .	15
3.1.2	PC-GITA andmestik . . . . .	18
3.2	Andmete esmane analüüs . . . . .	18
3.3	Andmete eeltöötlus masinõppe jaoks . . . . .	20
<b>4</b>	<b>Metodoloogia</b>	<b>21</b>
4.1	Kasutatud andmestikud ja eeltöötlus . . . . .	21
4.2	Masinõppemudelid . . . . .	21
4.2.1	Lihtne masinõppemudel (U1 ja U2 jaoks) . . . . .	21
4.2.2	Eeltreenitud mudel (U3 jaoks) . . . . .	21
4.3	Eksperimentide läbiviimine ja tunnuste skaleerimine . . . . .	22
4.3.1	U1: treenimine PC-GITA-l, testimine mõlemal . . . . .	22
4.3.2	U2: Treenimine ja testimine liidetud andmestikul . . . . .	22
4.4	Hindamismõõdikud ja statistiline analüüs . . . . .	23
<b>5</b>	<b>Eksperimendid</b>	<b>24</b>
5.1	U1: Treenimine PC-GITA andmestikul, testimine mõlemal . . . . .	24
5.1.1	Stsenaarium 1: Kõikidel harjutustel põhinev klassifitseerimine .	24
5.1.2	Stsenaarium 2: harjutuste kaupa klassifitseerimine (vokaalid ja monoloog) . . . . .	25
5.1.3	Stsenaarium 3: Erinevate harjutuste andmepunktide kombineerimine . . . . .	26
5.2	U2: Treenimine ja testimine liidetud andmestikul . . . . .	27
5.2.1	Stsenaarium 1: Kõik ühised tunnused liidetud andmestikul . . .	27

5.2.2	Stsenaarium 2: Harjutuste kaupa klassifitseerimine liidetud and- mestikul (vokaalid ja monoloog) . . . . .	28
5.2.3	Stsenaarium 3: Kombineeritud harjutused liidetud andmestikul .	30
5.3	U3: Mujal peenhäälestatud SSL-mudeli otserakendatavus Eesti andmetel	32
<b>6</b>	<b>Järeldused</b>	<b>33</b>
	<b>Viited</b>	<b>37</b>
	<b>Lisad</b>	<b>41</b>
	I. Sõnastik . . . . .	41
	II. Kasutatud akustilised tunnused . . . . .	44
	II. Litsents . . . . .	46

# 1 Sissejuhatus

Parkinsoni tõbi (PT) on progresseeruv neurodegeneratiivne haigus, mis mõjutab hinnanguliselt 7–10 miljonit inimest maailmas (Feigin et al. 2019). Eestis tehtud uuringute alusel on PT levimus 100 000 elaniku seas 152:100000 (Kadastik-Eerme et al. 2018), kusjuures haigestumus suureneb märkimisväärselt vanusega. Haigust iseloomustavad mitmesugused motoorsed sümptomid nagu bradükineesia (liigutuste aeglus), treemor (värin) ja rigiidsus (lihasjäikus), kuid oluliselt mõjutatud on ka kõne ja neelamise motorika (Logemann et al. 1978). Kuna rahvastik vananeb ja PT levimus kasvab, suureneb ka vajadus täpsete, mitteinvasiivsete ja kulutõhusate diagnoosimeetodite järele.

Viimastel aastatel on hääleanalüüs kujunenud paljulubavaks suunaks Parkinsoni tõve varaseks tuvastamiseks ja haiguskulu jälgimiseks (Shen et al. 2025). Kõne teke on keeruline motoorne protsess, mida PT neurodegeneratiivsed muutused mõjutavad, põhjustades tuvastatavaid muutusi hääle akustilistes omadustes. Masinõppe (*Machine learning*, ML) ja süvaõppe (*Deep learning*, DL) meetodite rakendamine hääleklippide analüüsimiseks võimaldab neid peeneid muutusi objektiivselt hinnata (Suppa et al. 2022). Erinevad uuringud on näidanud, et analüüsides hääle omadusi nagu põhitooni varieeruvust, sagedushälvet, intensiivsushälvet, formante ja kepstraalkoefitsiente, suudavad masinõppemudelid eristada PT patsiente tervetest kontrollisikutest sageli kõrge õigsusega (Suppa et al. 2022; Favaro et al. 2023). Eriti paljulubavad on olnud kaasaegsed süvaõppemudelid, sealhulgas isejuhendatud õppemeetodid (*Self Supervised Learning* (SSL), mis suudavad õppida informatiivseid tunnuseid otse helisignaalist ja näitavad häid tulemusi ka erinevates keeltes (Favaro et al. 2023; La Quatra et al. 2024; Tougui et al. 2024).

Kuigi rahvusvaheliselt on tehtud mitmeid uuringuid PT häälepõhiseks tuvastamiseks, kasutades sageli standardseid andmestikke nagu hispaaniakeelne PC-GITA (Orozco-Aroyave et al. 2014; Favaro et al. 2023; La Quatra et al. 2024), puuduvad sarnased põhjalikud uuringud, mis keskenduksid spetsiifiliselt eestikeelsele kõnele ja rakendaksid kaasaegseid ML/DL meetodeid. See loob uurimislünga, kuna pole teada, kui hästi rahvusvahelised mudelid ja meetodid kanduvad üle eesti keelele ning milline on parim lähenemine PT tuvastamiseks just Eestis kogutud andmetega.

Oluline osa käesolevast tööst on olnud ka spetsiaalse eestikeelse hääleklippide andmestiku kogumine, mis võimaldab järgnevalt püstitatud uurimisküsimusi käsitleda. Uurimistöö eesmärk on uurida erinevate masinõppepõhiste meetodite rakendatavust ja jõudlust Parkinsoni tõve tuvastamisel eestikeelsete hääleklippide põhjal. Töö keskendub kolmele peamisele uurimisküsimusele ja nendega seotud eksperimentaalsele stsenaariumile:

- **Uurimisküsimus 1 (U1):** Hinnatakse standardse, akustilistel tunnustel põhineva mudeli üldistusvõimet. Selleks treenitakse mudelit ainult hispaaniakeelsel PC-GITA andmestikul ja rakendatakse seejärel otse eestikeelsele andmestikule.
- **Uurimisküsimus 2 (U2):** Uuritakse, kas hispaania- ja eestikeelsete andmete

kombineerimine treeningfaasis parandab akustilistel tunnustel põhineva mudeli jõudlust võrreldes mudeliga, mis on treenitud ainult ühel keelel (st U1 tulemustega).

- **Uurimisküsimus 3 (U3):** Hinnatakse tipptasemel, isejuhendatud õppel põhineva *WavLM Base* mudeli (La Quatra et al. 2024) jõudlust otse eestikeelsel andmestikul. Eesmärk on testida mujal (antud juhul PC-GITA andmestikul) peenhäälestatud mudeli ülekantavust ja võimekust võrreldes lihtsamate, akustilistel tunnustel põhinevate mudelitega.

Nende uurimisküsimuste süstemaatiline uurimine aitab vastata küsimusele, milline lähenemine on kõige paljulubavam Parkinsoni tõve tuvastamiseks eestikeelsete hääleandmete põhjal.

Töö on struktureeritud järgmiselt: Peatükk 2 annab ülevaate varasemast kirjandusest Parkinsoni tõve häälepõhise tuvastamise meetodite kohta. Peatükk 3 sünteesib kirjanduse peamised leiud ja toob välja uurimislüngad. Peatükk 4 kirjeldab detailselt töös kasutatud eestikeelset ja PC-GITA andmestikku ning nende eeltöötlust. Peatükk 5 esitab üksikasjaliku metodoloogia uurimisküsimuste U1, U2 ja U3 vastamiseks, sh kasutatud masinõppemudelid ja hindamismõõdikud. Peatükk 6 esitleb eksperimentide tulemused. Peatükk 7 võtab töö kokku, teeb järeldused vastavalt saadud tulemustele ja arutleb edasiste uurimisvõimaluste üle.

Lõputöö keelelise korrektsuse tagamiseks ja sõnastuse parandamiseks ning osaliseks tõlkimiseks kasutati abistava vahendina tehisintellekti mudelit Google Gemini <https://gemini.google.com/>.

## 1.1 Vajalikud eelteadmised

Metodoloogia ja tulemuste mõistmiseks on kasulikud teadmised järgmistest valdkondadest: Parkinsoni tõbi ja selle mõju kõnele, masinõppe põhimõisted ning helisignaali analüüsi alused.

### 1.1.1 Parkinsoni tõbi

Nagu sissejuhatuses mainitud, on Parkinsoni tõbi progresseeruv neurodegeneratiivne haigus, mis mõjutab peamiselt motoorikat. Lisaks üldistele liigutuste aeglusele (bradükineesia), lihaskõvadusele (rigiidsus) ja värinale (treemor), mõjutab PT sageli ka kõneaparaadi lihaseid, põhjustades seisundit, mida nimetatakse düsartriaks (Logemann et al. 1978). See võib avalduda mitmesuguste hääle ja kõne muutustena, sealhulgas monotoonus (vähenenud kõnemeloodia), vähenenud helitugevus, ebatäpne artikulatsioon, hääle kähedus või värisemine ning muutused kõne tempos (Logemann et al. 1978).

Parkinsoni tõve raskusastme ja progresseerumise hindamiseks kasutatakse kliinilises praktikas mitmeid standardiseeritud skaalasid. Üks levinumaid on Hoehn & Yahi (H&Y)

skaala (Orozco-Arroyave et al. 2014), mis klassifitseerib haiguse staadiumid (tavaliselt 1-st 5-ni) peamiselt mootorsete sümptomite ulatuse ja tasakaaluhäirete olemasolu põhjal, kus kõrgem staadium tähistab raskemat haigusvormi. Patsiendi funktsionaalset võimekust ja toimetulekut igapäevaelu tegevustega hinnatakse sageli *Schwab-Englandi* skaalal (Orozco-Arroyave et al. 2014). See on protsentuaalne skaala (0–100%), kus 100% tähendab täielikku iseseisvust igapäevatoimingutes ja 0% täielikku sõltuvust teistest ning voodihaigust. Madalam *Schwab-Englandi* skoor viitab suuremale haigusest tingitud piiratusel. Nende skaalade abil saadud hinnangud on olulised nii ravi planeerimisel kui ka haiguse kulu objektiivsel jälgimisel.

Just need PT poolt esile kutsutud ja kliiniliselt hinnatavad hääles avalduvad muutused on aluseks uurimisküsimusele, et hääle akustilise analüüsi abil on võimalik haigust tuvastada.

### 1.1.2 Masinõpe (ML)

Masinõpe on tehisintellekti haru, mis tegeleb algoritmide loomisega, mis suudavad õppida andmetest mustreid ja teha ennustusi või otsuseid ilma otsese programmeerimiseta (Kotsiantis et al. 2006). Antud töös kasutatakse juhendatud õpet (*supervised learning*), täpsemalt **klassifitseerimist**. Klassifitseerimise eesmärk on määrata igale andmepunktile (antud juhul hääleklipile või sellest eraldatud tunnustele) eelnevalt defineeritud kategooria ehk klass antud juhul 'Parkinsoni tõbi' või 'Terve'. Selleks eraldatakse või õpitakse andmetest **tunnused** (*features*) – mõõdetavad omadused, mis iseloomustavad andmepunkti ja aitavad klasse eristada. Traditsioonilises masinõppes (kasutatud uurimisküsimustes U1 ja U2) eraldatakse tunnused käsitsi (nt akustilised parameetrid). Süvaõppe (Deep Learning, DL) ja isejuhendatud õppe (Self-Supervised Learning, SSL) meetodid (kasutatud uurimisküsimuses U3) suudavad seevastu õppida olulisi tunnuseid otse töötlemata andmetest (nt helisignaali) närvivõrkude abil (LeCun et al. 2015). Tuntud klassifitseerimisalgoritmid, mida ka siin töös mainitakse, on tugivektormasinad (SVM) (Cortes et al. 2009) ja otsustusmets (*Random Forest*) (Cutler et al. 2011).

### 1.1.3 Heli ja kõne akustiline analüüs

Inimese hääle tekib kopsudest tuleva õhuvoolu toimel häälekurdude võnkumisel ehk foneerimisel ning seejärel heli resonanceerimisel ja moduleerimisel häälekulglas (kõri, suu- ja ninaõõs) (Titze 2008). Seda komplekset helisignaali saab analüüsida erinevate akustiliste parameetrite abil, et kvantifitseerida hääle omadusi. Relevantsed parameetrid hõlmavad esiteks **põhisagedust** ( $F_0$ ), mis vastab häälekurdude võnkumise sagedusele ja määrab tajutava helikõrguse; selle keskväärtus ja varieeruvus (nt standardhälve) on olulised tunnused. Hääle stabiilsust iseloomustavad **sagedushälve**, mis mõõdab  $F_0$  lühiajalist ebastabiilsust ehk perioodide kestuse kõikumist järjestikuste võngete vahel, ning **intensiivsushälve**, mis mõõdab helisignaali amplituudi (helitugevuse) lühiajal-

ist ebastabiilsust. Suurenenud sagedushälve ja intensiivsushälve võivad viidata hääle värisemisele, kähedusele või kvaliteedi halvenemisele (Teixeira et al. 2013). Teine oluline häälekvaliteedi mõõdik on **harmonia-müra suhe hääles (HNR)**, mis kvantifitseerib häälesignaali harmoonia (perioodilise) osa ja mürakomponendi (aperioodilise) osa suhet, kusjuures madalam HNR viitab tavaliselt suuremale mürale või kähedusele hääles (Yumoto et al. 1982). Artikulsiooni õigsust võivad peegeldada **formandid (F1, F2, F3, F4)**, mis on häälekulglu resonantsidest tingitud sageduspiigid helisignaali spektris ja määravad peamiselt vokaalide kõla (Suppa et al. 2022). Lisaks kasutatakse laialdaselt **Mel-sageduse kepstraalkoefitsiente (MFCC)**, mis kirjeldavad helisignaali spektraalkuju lühiajalistes lõikudes, võttes arvesse inimese kuulmistaju eripärasid (Mel-skaala); need on levinud tunnused nii kõnetuvastuses kui ka häälepatoloogiate analüüsis (Davis et al. 1980). Nende ja teiste sarnaste tunnuste väärtused võivad PT patsientidel erineda tervetest kontrollisikutest, mis võimaldab masinõppemudelitel neid klasse eristada. Lisaks kasutatakse helisignaali visualiseerimiseks sageli **spektrogramme**, mis kujutavad heli sageduskomponentide intensiivsust ajaperioodi jooksul, kuid peamiselt keskendutakse arvulistele tunnustele ja SSL-mudelite otse helist õpitud representatsioonidele.

## 2 Metoodilised Suunad Kirjanduses

Parkinsoni tõve (PT) häälepõhise tuvastamise uurimisel on viimastel aastatel toimunud märkimisväärne areng, mis tuleneb peamiselt masinõppe (ML) ja süvaõppe (DL) meetodite laialdasemast rakendamisest (Iyer et al. 2023; Favaro et al. 2023; Suppa et al. 2022; Tougui et al. 2024; Orozco-Arroyave et al. 2014; Jeancolas et al. 2021; Akila et al. 2024; Carrón et al. 2021; Valarmathi et al. 2025).

Kirjanduses võib eristada peamiselt kahte metodoloogilist suunda. Esimene neist tugineb klassikalistele tehnikatele, kus kasutatakse käsitsi konstrueeritud akustilisi tunnuseid koos traditsiooniliste masinõppemudelitega. Teine, uuem suund, kasutab süvaõpet, et õppida tunnuseid otse andmetest. See hõlmab laia spektrit mudeleid alates konvolutsioonilistest närvivõrkudest (CNN) kuni uuemate transformeritel (*Transformers*) ja isejuhendatud õppel (SSL) põhinevate arhitektuurideni (Favaro et al. 2023; La Quatra et al. 2024; Tougui et al. 2024; Mallela et al. 2020).

Avaldatud võrdlused näitavad, et sügavate närvivõrkudega automaatselt õpitud representatsioonid ületavad sageli oma jõudluselt interpreteeritavaid, käsitsi loodud tunnuseid. Eriti tugevaid tulemusi, eriti mitme- ja ristkeelsetes stsenaariumides, on demonstreerinud isejuhendatud õppe (SSL) mudelitest saadud representatsioonid (Favaro et al. 2023; La Quatra et al. 2024).

Nende arengute valguses on aktuaalseteks uurimistemadeks kujunenud mitmed väljakutsed. Üheks neist on optimaalse tasakaalu leidmine mudelite ennustusjõudluse ja nende otsuste interpreteeritavuse vahel (Favaro et al. 2023; Iyer et al. 2023). Teine jätkuvalt oluline suund on mudelite robustsuse tagamine erinevates salvestustingimustes

ja keelekeskkondades (Carrón et al. 2021; Vásquez-Correa et al. 2021; La Quatra et al. 2024; Favaro et al. 2023).

Kirjanduse ülevaade keskendub järgnevalt nendele peamistele arvutuslikele meetoditele ja tehnoloogiatele, mida PT häälepõhises tuvastamises on kasutatud. Ülevaade käsitleb arenguid alates klassikalistest tunnuspõhistest masinõppemudelitest kuni kaasaegsete süvaõppearhitektuurideni, pöörates erilist tähelepanu siirdõppe rakendamisele, SSL-mudelite kasutamisele, andmekvaliteedi mõjule ja valideerimismetoodikatele. Eesmärk on kaardistada praegused tehnoloogilised suundumused ja väljakutsed kirjanduse põhjal, mis annavad konteksti lõputöö raames teostatavale analüüsile eestikeelsetele hääleandmetel.

## 2.1 Klassikalised meetodid: akustilised tunnused ja klassikaline masinõpe

Varasemad ja ka mitmed tänapäevased uuringud Parkinsoni tõve häälepõhiseks tuvastamiseks, näiteks Narendra jt (Narendra et al. 2021) töö, tuginevad traditsioonilisele lähenemisele. Selles eraldatakse helisignaalist esmalt suur hulk eeldefineeritud akustilisi ja prosoodilisi tunnuseid, mida seejärel kasutatakse sisendina klassikalistele masinõppe klassifikaatoritele. Levinumad eraldatavad tunnused püüavad kvantifitseerida erinevaid hääle aspekte, mida PT teadaolevalt mõjutab. Nagu kirjeldavad mitmed autorid (Iyer et al. 2023; Suppa et al. 2022; Orozco-Arroyave et al. 2014; Narendra et al. 2021), kuuluvad nende hulka hääle põhisageduse ( $F_0$ ) ja selle stabiilsuse mõõdikud, hääle amplituudi stabiilsuse mõõdikud, hääle kvaliteeti iseloomustavad tunnused nagu harmooniamüra suhe hääles (HNR), formantsageduste karakteristikud ning kõne artikulatsiooni ja prosoodiat kirjeldavad parameetrid. Lisaks kasutatakse sageli spektraalseid tunnuseid, nagu Mel-sageduse kepsraalkoefitsiente (MFCC) või nende statistilisi momente, mis on laialt levinud ka üldises kõnetöötluses (Suppa et al. 2022; Nagasubramanian et al. 2021; Boualoulou et al. 2023; Akila et al. 2024; Valarmathi et al. 2025).

Kuna sellisel viisil eraldatud tunnuste arv võib ulatuda tuhandetesse (nt kasutades standardseid komplekte nagu OpenSMILE ComParE (Suppa et al. 2022)), rakendatakse sageli tunnuste valiku (*feature selection*) või dimensioonide vähendamise (*dimensionality reduction*) meetodeid, et leida kõige informatiivsem alamhulk ja vältida mudelite üleõppimist (Akila et al. 2024). Levinud on näiteks korrelatsioonipõhine tunnuste valik (CFS) (Suppa et al. 2022) või põhikomponentide analüüs (PCA) (Favaro et al. 2023).

Nende valitud või vähendatud tunnuste põhjal treenitakse erinevaid klassifitseerimismudeleid. Tugivektormasinaid (SVM), nii lineaarse kui ka mittelineaarse tuumaga (Suppa et al. 2022; Vásquez-Correa et al. 2021; Orozco-Arroyave et al. 2014; Narendra et al. 2021). Ansamblimeetoditest on populaarne otsustusmets (*Random Forest*, RF), mille kasutamist Parkinsoni tõve kontekstis kirjeldavad teiste hulgas (Iyer et al. 2023; Favaro et al. 2023; Carrón et al. 2021; Akila et al. 2024).

Lisaks mainitud meetoditele on uuringutes rakendust leidnud ka logistiline regressioon (*logistic regression*, LR) (Iyer et al. 2023; Carrón et al. 2021), K-lähima naabri meetod (KNN) ning XGBoost (Favaro et al. 2023; Akila et al. 2024; Valarmathi et al. 2025). Carrón jt (Carrón et al. 2021) annavad ülevaate ka teistest klassikalistest algoritmidest selles valdkonnas. Tulemuste optimeerimiseks on oluline mudelite hüperparameetrite hoolikas häälestamine, milleks sageli kasutatakse ruudustikotsingut (*grid search*) (Vásquez-Correa et al. 2021; Carrón et al. 2021).

Selle lähenemisega saavutatud tulemused varieeruvad oluliselt sõltuvalt kasutatud andmestikust, tunnuste komplektist, valideerimismetoodikast ja klassifikaatorist. Kontrollitud laboritingimustes kogutud andmetel (nagu PC-GITA) on saavutatud kõrgeid õigsuseid, sageli üle 80-90% (Orozco-Aroyave et al. 2014; Suppa et al. 2022), eriti kui kasutatakse hoolikalt valitud tunnuseid ja optimeeritud mudeleid. Samas on tulemused tagasihoidlikumad keerulisemate või mürarikkamate andmete puhul ning sõltuvad tugevalt valitud tunnuste võimest haiguspetsiifilisi muutusi tabada (Iyer et al. 2023; Narendra et al. 2021; Vásquez-Correa et al. 2021). klassikaliste meetodite peamiseks eeliseks peetakse sageli nende paremat interpreteeritavust võrreldes süvaõppe "musta kasti" mudelitega (Favaro et al. 2023), kuid nende piiranguks on vajadus käsitsi tunnuseid konstrueerida ja potentsiaalsed probleemid tunnustevaheliste lineaarsete sõltuvustega (Iyer et al. 2023).

## 2.2 Süvaõppe meetodid: CNNidest isejuhendatud mudeliteni

Vastukaaluks klassikalistele meetoditele, mis tuginevad käsitsi konstrueeritud tunnustele, on süvaõppe (DL) meetodid toonud kaasa paradigma muutuse, võimaldades õppida hierarhilisi ja informatiivseid tunnuseid otse andmetest. Üks varasemaid ja levinumaid DL lähenemisi PT tuvastamisel oli konvolutsiooniliste närvivõrkude rakendamine kõne spektrogramm-esitustel (Iyer et al. 2023; Vásquez-Correa et al. 2021; Boualoulou et al. 2023). Analüüsides spektrogrammi kui kahemõõtmelist pilti, suudavad CNNid tabada aja-sageduslikke mustreid, mis võivad olla seotud hääle kvaliteedi ja artikulatsiooni muutustega (Boualoulou et al. 2023). Uuringud on näidanud, et CNN-põhised meetodid, eriti kombineerituna siirdõppega suurtes pildandmestikes eeltreenitud mudelilt, võivad ületada traditsioonilisi tunnuspõhiseid meetodeid, eriti keerukamate või mürarikkamate andmete puhul (Iyer et al. 2023; Boualoulou et al. 2023). Siiski jäi CNNide interpreteeritavus sageli piiratuks (Iyer et al. 2023; Favaro et al. 2023).

Fookus on liikunud veelgi võimsamate süvaõppearhitektuuride poole, eriti Transformer-põhiste mudelitele ja isejuhendatud õppe meetoditele (Favaro et al. 2023; La Quatra et al. 2024). Tuntud SSL mudelid, nagu *Wav2Vec 2.0*, *HuBERT* ja *WavLM* (Alexei Baevski et al. 2020; Wei-Ning Hsu et al. 2021; Wang et al. 2022), on näidanud paljulubavaid tulemusi (Favaro et al. 2023; La Quatra et al. 2024). Need mudelid õpivad representatsioone otse suurest hulgast märgistamata heliandmetest andmestikult, nagu on kirjeldatud samades allikates (Favaro et al. 2023; La Quatra et al. 2024). Need mudelid on näidanud

muljetavaldavat võimekust erinevates kõnetöötlusülesannetes ning ka PT tuvastamisel. Favaro jt (Favaro et al. 2023) ja La Quatra jt (La Quatra et al. 2024) leidsid, et SSL mudelitest eraldatud esitus või nende mudelite peenhäälestamine (*fine-tuning*) spetsiifiliselt PT tuvastamise ülesandele annab sageli paremaid ja robustsamaid tulemusi kui klassikalised tunnused või varasemad DL meetodid, eriti mitmekeelsetes ja ristkeelsetes stsenaariumides (Favaro et al. 2023). Ka teised Transformer-põhised arhitektuurid, nagu *Audio Spectrogram Transformer* (AST), mis on spetsiaalselt kohandatud audio spektrogrammide töötlemiseks, on näidanud siirdõppe ja peenhäälestamise abil kõrgeid tulemusi PT tuvastamisel (Tougui et al. 2024).

Nende suurte, sageli kümneid või sadu miljoneid parameetreid omavate mudelite rakendamisel on siirdeõpe ja peenhäälestamine kriitilise tähtsusega, kuna meditsiinilised andmestikud on tavaliselt liiga väikesed, et selliseid mudeleid nullist trennida (La Quatra et al. 2024; Tougui et al. 2024; Vásquez-Correa et al. 2021). Eeltreening suurtes üldistes heli- või kõneandmestikes võimaldab mudelil õppida üldisi akustilisi ja keelelisi mustreid, mida saab seejärel kohandada spetsiifilisele PT tuvastamise ülesandele suhteliselt väikese hulga märgistatud andmetega (La Quatra et al. 2024; Tougui et al. 2024). Edukas siirdeõpe ja peenhäälestamine on näidanud potentsiaali märkimisväärselt parandada PT tuvastamise õigsust võrreldes nullist trennimisega (La Quatra et al. 2024; Tougui et al. 2024). Samas võib siirdõppe efektiivsus sõltuda eeltreeningu ja sihtülesande sarnasusest ning kasutatud mudeli arhitektuurist (Vásquez-Correa et al. 2021).

### 2.3 Andmestikud, keelteülene rakendatavus ja robustsus

Parkinsoni tõve häälepõhise tuvastamise uuringutes kasutatakse mitmeid avalikke ja privaatseid andmestikke. Laialdaselt levinud *benchmark*-andmestikuks, mida kasutatakse ka käesolevas töös, on hispaaniakeelne PC-GITA andmestik (Orozco-Arroyave et al. 2014; Vásquez-Correa et al. 2021; La Quatra et al. 2024; Narendra et al. 2021; Boualoulou et al. 2023). Orozco-Arroyave jt (Orozco-Arroyave et al. 2014) kirjeldavad seda kui kontrollitud laboritingimustes kogutud andmestikku. Sellised laboris kogutud andmestikud (nagu PC-GITA ja ka käesolevas töös loodud eestikeelne andmestik) võimaldavad küll täpsemalt uurida haiguse mõju häälele, minimeerides segavaid faktoreid, kuid tõstatavad küsimuse mudelite robustsusest ja üldistusvõimest reaalsetes, mürarikkamates tingimustes (Carrón et al. 2021; La Quatra et al. 2024). Uuringud on järjepidevalt näidanud, et mudelite jõudlus võib märkimisväärselt langeda, kui neid rakendada kontrollimata keskkonnas nutitelefoniga kogutud andmetel (Carrón et al. 2021; La Quatra et al. 2024; Iyer et al. 2023; Tougui et al. 2024). Selle leevendamiseks uuritakse küll kõneparanduse (*speech enhancement*, SE) tehnikaid (La Quatra et al. 2024), kuid robustsus muutlikes tingimustes jääb oluliseks väljakutseks (Carrón et al. 2021; La Quatra et al. 2024).

Teine keskne uurimissuund on mudelite rakendatavus ja siirdatavus erinevate keelte vahel, arvestades, et PT motoorsed sümptomid kõnes peaksid olema teatud määral

keeleülesed (Favaro et al. 2023; Vázquez-Correa et al. 2021). Favaro jt (Favaro et al. 2023) ulatuslikus uuringus kuues keeles leiti, et kuigi ka klassikalised tunnused näitasid teatavat keeleülest robustsust, olid isejuhendatud õppe mudelitest (nagu *TRILLsson*, *Wav2Vec 2.0*, *HuBERT*) saadud representatsioonid üldiselt täpsemad just ristkeelsetes (*cross-lingual*) stsenaariumides, viidates PT düsartria ühistele muustritele. Samas jääb täielik üldistamine täiesti uuele, treeningus nägemata keelele väljakutseks (Favaro et al. 2023; Vázquez-Correa et al. 2021). Vázquez-Correa jt (Vázquez-Correa et al. 2021) uurisid siirdeõpet (transfer learning) CNN-mudelitega hispaania, saksa ja tšehhi keele vahel ning leidsid, et teadmiste ülekanne toimib edukalt, parandades õigsust võrreldes nullist treenimisega, eriti kui lähtekeele mudel on piisavalt täpne. Need tulemused rõhutavad nii ristkeelse testimise (nagu U1) kui ka siirdõppe (nagu U3) potentsiaali ja väljakutseid PT tuvastamisel erinevates keeltes.

## 2.4 Hindamismetoodikad

Mudelite usaldusväärseks hindamiseks ja võrdlemiseks on oluline kasutada korrektsid valideerimismetoodikaid ja mõõdikuid. Levinud hindamismeetodiks Parkinsoni tõve hääleanalüüsi uuringutes on K-korda ristvalideerimine, kus K on tüüpiliselt 5 või 10 (Akila et al. 2024; La Quatra et al. 2024; Mallela et al. 2020; Orozco-Aroyave et al. 2014; Valarmathi et al. 2025; Vázquez-Correa et al. 2021). Selle meetodi puhul on kriitilise tähtsusega tagada **kõnelejast sõltumatu jaotus** (*subject-independent split*), kus sama isiku andmed ei satu kunagi samaaegselt treening- ja testandmestikku. Selline lähenemine aitab vältida olukorda, kus mudel õpib ära konkreetse kõneleja hääle eripärad, mis ei ole seotud haigusega, ja näitab seetõttu kunstlikult kõrgeid tulemusi (nn identiteedi segiajamine, *identity confounding*) (Favaro et al. 2023; La Quatra et al. 2024; Tougui et al. 2024; Vázquez-Correa et al. 2021).

Kuigi K-korda ristvalideerimine on laialt levinud ja kasulik meetod mudeli üldistusvõime hindamiseks, tuleb selle rakendamisel arvestada andmestiku suurusega. Väikeste andmemahdade korral, nagu ka käesolevas töös kasutatud Eesti andmestik ( $n=28$ ), võib K-korda ristvalideerimise tulemuste stabiilsus olla mõnevõrra piiratud. Kui testalamhulk igas korduses on väike, võivad üksikud andmepunktid tulemusi rohkem mõjutada, mis võib kaasa tuua suurema varieeruvuse mudeli jõudlushinnangutes erinevate korduste vahel (Kohavi 1995; Vabalas et al. 2019). Seetõttu on väikeste valimite puhul eriti oluline raporteerida lisaks keskmistele tulemustele ka nende hajuvust ning tõlgendada tulemusi ettevaatlikult. Mõnedes töödes kasutatakse väiksemate andmemahdade korral või tulemuste robustsuse suurendamiseks ka alternatiivseid või täiendavaid meetodeid, nagu "jätta-üks-välja" ristvalideerimine (LOOCV), pesastatud ristvalideerimine (*nested cross-validation*) hüperparameetrite häälestamiseks ja hindamiseks (Favaro et al. 2023; Vázquez-Correa et al. 2021) või korduvaid juhuslikke jaotusi (Jeancolas et al. 2021).

## 2.5 Süntees ja tehnilised uurimislüngad

Kirjanduse ülevaates analüüsitud uuringud Parkinsoni tõve häälepõhise tuvastamise kohta näitavad valdkonna kiiret arengut ja mitmeid olulisi tehnilisi suundumusi, kuid osutavad ka püsivatele puudustele.

Esiteks on täheldatav selge metoodiline trend, kus klassikalistelt, käsitsi konstrueeritud akustilistel ja prosoodilistel tunnustel ning klassikalistel masinõppemudelitel (nagu SVM või *Random Forest*) põhinevatelt süsteemidelt (Suppa et al. 2022; Orozco-Arroyave et al. 2014; Narendra et al. 2021; Carrón et al. 2021) liigutakse üha enam süvaõppe poole. Süvaõppe raames rakendatakse närvivõrke, sealhulgas konvolutsioonilisi närvivõrke, Transformereid ja SSL-mudeleid, mis võimaldavad õppida informatiivseid representatsioone otse heliandmetest. Sellist paradigmuuutust tõendab lai hulk uuemaid uurimusi (Akila et al. 2024; Boualoulou et al. 2023; Mallela et al. 2020; Nagasubramanian et al. 2021).

Teise domineeriva suundumusena on esile kerkinud siirdõppe laialdane kasutamine. See tähendab, et suurtes üldistes andmekogudes eeltreenitud mudeleid, eriti suuri alusmudeleid (*Foundational models*) nagu *WavLM*, *HuBERT* ja *AST*, peenhäälestatakse PT-spetsiifilistel andmetel. Mitmed uuringud, sealhulgas (Favaro et al. 2023), (Jeancolas et al. 2021), (La Quatra et al. 2024), (Tougui et al. 2024) ning (Vásquez-Correa et al. 2021), on kinnitanud, et siirdõppe annab sageli paremaid tulemusi kui mudelite treenimine nullist.

Kolmandaks on teadlased hakanud üha enam tähelepanu pöörama mudelite robustsusele reaalses tingimustes ja nende keelteülesele rakendatavusele. Selle ajendiks on tõdemus, et laboris saavutatud head tulemused ei pruugi alati üle kanduda mürrarikkamasse keskkonda või teise keelde. Robustsuse ja keelteülese üldistavuse problemaatikat käsitlevad mitmed autorid, näiteks (Carrón et al. 2021), (Favaro et al. 2023), (Iyer et al. 2023), (La Quatra et al. 2024), (Tougui et al. 2024) ja (Vásquez-Correa et al. 2021).

Masinõppemeetodite robustsust, keelteülest rakendatavust ning nende spetsiifilist uurimist Parkinsoni tõve tuvastamisel eestikeelsete hääleandmete põhjal – ongi peamiseks lähtekohaks ja motivatsiooniks bakalaureusetöö uurimisküsimustele (U1, U2, U3). Järgnevates peatükkides keskendutaksegi nendele küsimustele süstemaatiliste vastuste otsimisele läbi viidud eksperimentide kaudu, analüüsides erinevate masinõppelähenedemiste sobivust ja piiranguid Eesti keele ja andmete kontekstis.

## 3 Andmed

Selles peatükis kirjeldatakse uurimistöös kasutatud andmestikke, nende kogumise protsessi, osalejate demograafilisi ja kliinilisi andmeid ning andmete esmast analüüsi ja eeltöötlust masinõppe mudelite jaoks ettevalmistamisel.

### 3.1 Kasutatud andmestikud

Uurimistöös kasutati kahte peamist andmestikku: spetsiaalselt selle töö jaoks kogutud eestikeelset andmestikku ning Parkinsoni tuvastamise tüüpandmestikku - hispaaniakeelset PC-GITA .

#### 3.1.1 Eestikeelne andmestik

Eestikeelne andmestik koguti eesmärgiga analüüsida Parkinsoni tõve mõju eestikeelsele kõnele ning luua ressursid masinõppemudelite testimiseks ja treenimiseks eesti keele kontekstis.

**Osalejad** Uuringusse kaasati Tartu Ülikooli Kliinikumi Närvikliiniku abiga algselt 46 osalejat, kellest 18 olid diagnoositud Parkinsoni tõvega ning 28 olid kontrollisikud. Pärast esmast sobivuse hindamist moodustati valim, kuhu kuulus 18 PT diagnoosiga osalejat ja 18 demograafiliselt sarnast kontrollisikut (KG grupp). Nendes valitud 36 osalejast oli 5 salvestust eritingimustest: kolm Eesti Rahva Muuseumis ja kaks iPhone 12 nutitelefoni. Andmete kvaliteedi edasisel hindamisel ilmnis, et nende eritingimustes tehtud salvestistega osalejate hääleomadused erinesid oluliselt ülejäänud, standardsetes tingimustes salvestatud valimi näitajatest. Selleks, et tagada analüüsitava andmestiku suurem homogeensus ja vähendada salvestustingimuste varieeruvusest tulenevaid võimalikke segavaid mõjusid, otsustati need viis osalejat ja neile kõige paremini demograafiliselt sobivad kontrollgrupi liikmed edasisest analüüsist eemaldada. Seega koosnes lõplik valim, mida kasutati masinõppemudelite treenimiseks ja testimiseks, 14 Parkinsoni tõvega patsiendist ja 14 kontrollisikust. Osalejate detailsemad demograafilised andmed ( $N=14+14$ ) on esitatud tabelis 2.

Parkinsoni tõve diagnoos oli kinnitatud meditsiinitöötaja poolt ning haiguse raskusaste hinnatud Hoehn & Yahi (H&Y) ja Schwab-Englandi skaalade alusel. PT grupi osalejate vanused jäid vahemikku 55-87 aastat. Kontrollgruppi kuulusid isikud, kellel ei esinenud teadaolevalt neuroloogilisi haigusi ega muid kõnet või häält mõjutavaid seisundeid. Kõik uuringus osalenud olid eesti keelt emakeelena kõnelejad. Enne uuringus osalemist andsid kõik osalejad teadliku kirjaliku nõusoleku. Uuringu läbiviimiseks saadi luba Tartu Ülikooli inimuuringu eetikakomiteelt (otsuse number 397/T-17). Osalejate vanuse, soo ja H&Y skaala andmed on esitatud tabelites 1 ja 2.

Tabel 1. Lõplikku valimisse kuuluvate Parkinsoni tõve (PT) patsientide ( $n=14$ ) ja kontrollgrupi (KG) osalejate ( $n=14$ ) demograafilised ning kliinilised andmed kõrvuti vertikaalse eraldusjoonega (mõlemad grupid järjestatud vanuse järgi).

Parkinsoni tõve patsiendid (PT)						Kontrollgrupp (KG)		
Nr.	Vanus	Sugu	Diagnoosi kestus (a)	H&Y skaala	Schwab-Eng. %	Nr.	Vanus	Sugu
1	55.12	M	2	3	90	1	63.33	N
2	64.68	M	6	2	90	2	64.22	M
3	65.37	N	3	1	90	3	66.81	N
4	67.82	N	3	1	80	4	68.91	N
5	67.89	N	5	3	80	5	69.96	N
6	69.33	M	1	1	90	6	71.69	N
7	70.42	M	12	3	80	7	72.19	N
8	70.88	N	1	2	90	8	77.97	N
9	78.19	N	7	2	80	9	78.41	N
10	79.06	N	4	2	80	10	78.57	N
11	79.88	N	9	2	90	11	79.90	N
12	80.51	N	23	3	50	12	80.54	M
13	82.02	M	10	2	90	13	82.02	N
14	87.26	M	4	3	70	14	87.20	N

Tabel 2. Parkinsoni tõve (PT) ja kontrollgrupi (KG) osalejate (mõlemas  $n=14$ ) demograafiliste tunnuste võrdlus.

Tunnus	PT grupp ( $n=14$ )	KG grupp ( $n=14$ )
Vanus (aastad)		
Keskmine $\pm$ std	72.74 $\pm$ 8.66	74.41 $\pm$ 7.24
Vahemik [min-max]	[55.12 – 87.26]	[63.33 – 87.20]
Sugu (Mehi / Naisi)		
Arv (n)	6 M / 8 N	2 M / 12 N
Protsent (%)	42.86% M / 57.14% N	14.29% M / 85.71% N

**Salvestuskeskkond ja -seadmed** Põhiosa häälesalvestistest koguti standardiseeritud tingimustes Tartu Ülikooli Kliinikumi kabinetis, et minimeerida taustamüra ja tagada salvestiste kõrge kvaliteet. Salvestamiseks kasutati *RØDE NT-USB Mini* lauamikrofoni, mis oli asetatud lauale osalejast konstantsele 40 cm kaugusele. Helifailid salvestati WAV-formaadis, diskreetimissagedusega 44.1 kHz ja 16-bitise resolutsiooniga. Kuigi valdav

enamik häälesalvestisi koguti standardiseeritud tingimustes ja ettenähtud seadmestikuga, esines väheste salvestuste puhul andmekogumisel tehnilisi eksimusi või ettenägematuid kõrvalekaldeid tavapärasest protokollist. Täpsemalt, kolm salvestust tehti planeeritust erinevalt Eesti Rahva Muuseumis ning kahe Tartu Ülikooli Kliinikumis tehtud salvestuse puhul kasutati välise salvestusmikrofoni asemel iPhone 12 nutitelefonis sisseehitatud mikrofoni. Nende kogutud salvestiste andmekvaliteeti analüüsisiti põhiaandmestikuga ning otsustati erisalvestused jätta lõplikust valimist välja, et tagada uuritava valimi võimalikult suur homogeensus.

**Salvestusprotokoll ja Kõneülesanded** Andmete kogumisel järgiti protokoll, mis loodi eesmärgiga võimalikult täpselt jäljendada **PC-GITA andmestikku** (Orozco-Arroyave et al. 2014) ning hinnata kõne erinevaid aspekte: fonatsiooni, artikulatsiooni ja prosoodiat. Osalejatel paluti sooritada järgmised ülesanded:

- **Foneetika hindamine (vokaalide hääldamine):**

- Järjest viie sekundi vältel hääldada iga eestikeelset vokaali /a/, /e/, /i/, /o/, /u/, /õ/, /ä/, /ö/, /ü/ püsival, monotoonsel toonil.
- Korrata eelmist ülesannet, kuid hääldada vokaale alt-üles libistades (*glissando*).

- **Artikulatsiooni hindamine:**

- Diadokokineesia (DDK) hindamiseks korrata kolm korda järgmisi silbijadasid: /pa-ta-ka/, /pa-ka-ta/, /pe-ta-ka/, /pa/, /ta/, /ka/.
- Hääldada järgmised 39 eestikeelset sõna: Maasikas, auto, maja, kaev, küsimus, piim, klaver, karikakar, paber, saal, traktor, maantee, kvartal, õunapuu, küünal, röövel, vaarikas, kahvel, süda, mask, jalgratas, trellid, rajoon, vangla, erakond, kringel, saade, televiisor, õudne, lugema, sõitma, jooksuma, kõrge, igav, laisk, eputama, ujuma, jutustama, ovaalne.

- **Prosoodia hindamine:**

- Korrata järgnevat lauset kolm korda:
  1. Minu autol on neli ratast.
  2. Maarja, kes on minu naaber, kasvatab maasikaid.
  3. Tanel jookseb rongile.
  4. Uued raamatud ei mahu meie raamaturiulile.
  5. Noored kunstnikud kaasavad raha oksjonilt.
  6. Tädi ostab meile voodiraami, mida me oleme pikalt soovinud.

- Lugeda järgnevad laused etteantud rõhuasetusega (üks kord):
  1. Kas sina nägid uudiseid? Minu arust **VÕITJA** ei väärinud kuldmedalit.
  2. Lillede korjamisest **MIDAGI** paremat polegi.
  3. Meie majas on kaks magamistuba ning **KUUR** ning köök.
  4. Sõitsime kiiruga sõbrale külla ning **TERVITASIME** teda.
- Esitada ligikaudu ühe minuti pikkune spontaanne monoloog vastuseks küsimusele oma tavapärase päeva kohta.

Kogu salvestusprotsessi viis läbi koolitatud personal, tagades ülesannete korrektse selgitamise ja protokollidest kinnipidamise.

### 3.1.2 PC-GITA andmestik

Lisaks spetsiaalselt kogutud eestikeelsele andmestikule kasutati antud töös võrdlusena ning mudelite treenimiseks ja testimiseks rahvusvaheliselt laialdaselt kasutatavat PC-GITA andmestikku (Orozco-Arroyave et al. 2014). See on hispaaniakeelne andmestik, mis sisaldab 50 Parkinsoni tõvega patsiendi ja 50 terve kontrollisiku häälesalvestisi. Andmed on kogutud kontrollitud laboritingimustes, kus on kasutatud kvaliteetset helitehnikat ning salvestusprotokoll sarnaneb eestikeelsete kogutud andmetega. Salvestusprotokoll sisaldab vokaalide hääldusi, DDK ülesandeid, sõnade ja lausete lugemist ning monoloogi. PC-GITA andmestikku kasutati siin töös baasmudeli treenimiseks (U1), kombineeritud treeninguks (U2) ning SOTA mudelite testimiseks (U3), võimaldades hinnata loodud mudeleid standardse etaloni suhtes ja uurida ristkeelseid efekte. Andmestikule jagati ligipääs autoritega kontakteerumise teel.

## 3.2 Andmete esmane analüüs

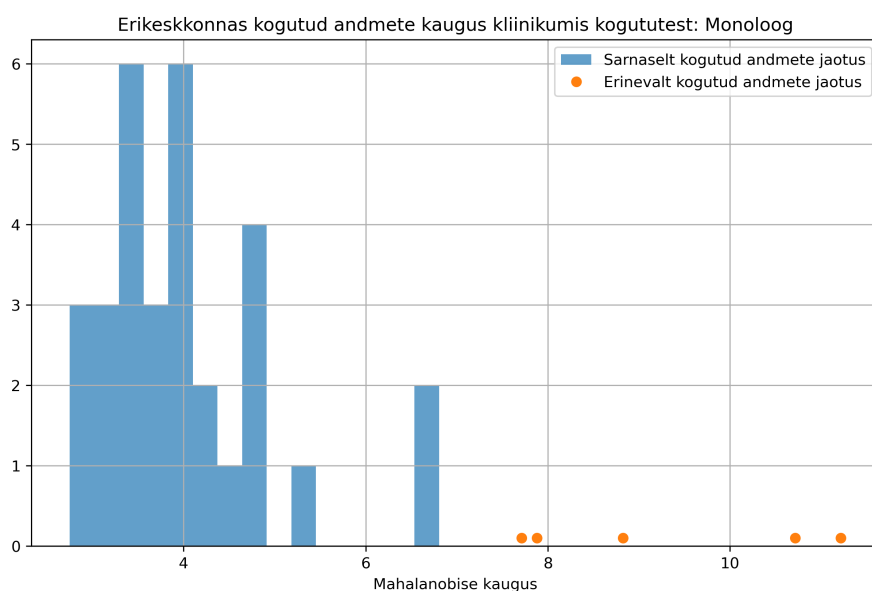
Enne masinõppe mudelite rakendamist teostati kogutud ja sobitatud eestikeelse andmestiku esmane analüüs, et iseloomustada andmestikku ja kirjeldada gruppidevahelisi erinevusi.

Esiteks analüüsiti demograafilisi andmeid. Pärast sobitamist olid gruppide keskmised vanused sarnased (PT grupi keskmine vanus  $72.74 \pm 8.66$  aastat; KG grupi keskmine vanus  $74.41 \pm 7.24$ ). Sooline jaotus erines gruppide vahel mõnevõrra: PT grupis oli 6 meest ja 8 naist, samas kui kontrollgrupis oli 2 meest ja 12 naist (vt Tabel 2). Kuigi soo mõju ei uuritud eraldi, on oluline märkida, et selline erinevus gruppide vahel võib teatud määral mõjutada hääle akustilisi karakteristikuid. Arvestades valimi suurust ja töö fookust, ei olnud võimalik soolist mõju süvitsi analüüsida, kuid see aspekt väärib tähelepanu tulevastes uuringutes.

Teiseks viidi läbi akustiliste põhitunnuste kirjeldav analüüs. Selleks arvutati, kasutades Pythoni teeki *parselmouth* (mis põhineb *Praat* tarkvaral (Jadoul et al. 2018)), erinevad hääle kvaliteedi mõõdikud.

Lisaks Parkinsoni tõve ja kontrollgrupi vaheliste tunnuste esmasele analüüsile uuriti eraldi ka neid hääleklippe, mis koguti tehniliste eksimuste või ettenägematute kõrvalekallete tõttu mittestandardsetes tingimustes (kolm Eesti Rahva Muuseumis (ERM) ja kaks iPhone 12 nutitelefoni). Kuigi nende salvestiste algne tekkimine ei olnud tingitud eesmärgist süstemaatiliselt uurida salvestustingimuste mõju, võimaldas nende hilisem akustiline analüüs siiski saada teatud esialgse pildi sellest, kuidas erinevad salvestuskeskkonnad ja -seadmed võivad hääleomadusi potentsiaalselt mõjutada. Selliste mõjude teadvustamine on oluline aspekt masinõppemudelite rakendatavuses.

Selleks võrreldi eritingimustes salvestatud klippide akustilisi tunnuseid standardsetes tingimustes salvestatud põhiandmestikuga. Võrdluse alusena Mahalanobise kauguse spetsiifiliselt monoloogi ülesandest eraldatud ja normaliseeritud akustilistele tunnustele, ilmnes eritingimustes salvestatud andmete puhul selgem eristumine (vt Joonis 1).



Joonis 1. Erikeskkondades (ERM, iPhone) ja kliinikumis kogutud monoloogide Mahalanobise kauguste jaotus, põhinedes normaliseeritud akustilistel tunnustel. Oranžid punktid tähistavad erikeskkondades ja sinine histogramm kliinikumis kogutud andmeid.

Jooniselt 1 on näha, et Eesti Rahva Muuseumis (ERM) ja iPhone nutitelefoni salvestatud monoloogid (kujutatud oranžide punktidenä) omasid oluliselt suuremaid Mahalanobise kaugusi võrreldes kliinikumis standardsetel tingimustel kogutud andmetega (kujutatud sinise histogrammina). See viitab, et monoloogiülesande tunnuste mitmemõõtmeline analüüs Mahalanobise kauguste kaudu tõi esile nende erikeskkondade salvestiste eristuva olemuse. Varieeruvus, mis salvestuskeskkonnast ja -tehnikast ei võimalda kasutada kogutud andmeid põhiandmestikuga.

### 3.3 Andmete eeltöötlus masinõppe jaoks

Mõlemad eksperimentides kasutatud andmestikud läbisid enne masinõppemudelite treenimist ja testimist mitu eeltötluse etappi, et tagada andmete ühtlus ja sobivus mudelitele. Esiteks viidi läbi kõneülesannetepõhine segmenteerimine, mille käigus jaotati pikemad helisalvestised lühemateks harjutuspõhisteks segmentideks, konkreetsele ülesandele vastavateks helifailideks; nende segmentide täpsem kasutus on kirjeldatud metodoloogia peatükis 4. Teiseks normaliseeriti helifailide helitugevus, kasutades tippväärtuse normaliseerimist, et vähendada salvestushelitugevuse erinevuste mõju (Favaro et al. 2023). Kolmandaks eraldati igast segmenteeritud helifailist 37 akustilist tunnuset, kasutades Pythoni teeki *parselmouth* (Jadoul et al. 2018); eraldatud tunnuste hulka kuulusid põhi- tooni (F0) statistikud, sagedushälve ja intensiivsushälve mõõdikud, harmoonilisuse ja müra suhe (HNR), formantide (F1-F4) keskmised sagedused, Mel-sageduse kepstraalkoefitsiendid (MFCC 1-12, keskmine ja standardhälve) ning intensiivsuse statistikud. Viimase sammuna jaotati andmestikud treening-, ja testalamhulkadeks vastavalt 4 peatükis kirjeldatud eksperimentidele, tagades alati rangelt kõnelejast sõltumatu jaotuse (La Quatra et al. 2024; Tougui et al. 2024; Vásquez-Correa et al. 2021).

## 4 Metodoloogia

### 4.1 Kasutatud andmestikud ja eeltöötlus

Eksperimentides kasutati peatükis 3 kirjeldatud eestikeelset ja hispaaniakeelset PC-GITA andmestikku. Mõlemad andmestikud läbisid ühtse eeltöötuse, mis hõlmas segmenteerimist kõneülesannete kaupa, tippväärtuse (peak) normaliseerimist ning akustiliste tunnuste eraldamist *parselmouth* teegi abil. Täpsem eeltöötuse kirjeldus on esitatud peatükis 3.

### 4.2 Masinõppemudelid

#### 4.2.1 Lihtne masinõppemudel (U1 ja U2 jaoks)

Uurimisküsimuste U1 ja U2 vastamiseks kasutati klassikalisi masinõppe klassifikaatorid, mis sobitati eelnevalt eraldatud ja töödeldud akustiliste tunnuste põhjal (kirjeldatud peatükis 3). Konkreetselt implementeeriti kaks laialt levinud mudelit, kasutades Pythoni *scikit-learn* teeki (Pedregosa, Gaël Varoquaux, et al. 2018). Esimene mudel oli tugivektormasin (SVM), mis sobitati tõenäosuste arvutamiseks. Teine mudel oli otsustusmets. Mõlema mudeli puhul kasutati tulemuste reprodutseeritavuse tagamiseks juhuslikkuse seemet 42. Need mudelid treeniti alati standardiseeritud tunnustel, nagu kirjeldatud alajaotuses 4.3.

#### 4.2.2 Eeltreenitud mudel (U3 jaoks)

Kolmanda uurimisküsimuste testimiseks kasutati tiptasemel (*State-of-the-Art*, SOTA) isejuhendatud õppe (*Self-Supervised Learning*, SSL) mudelit, mis pärineb La Quatra jt. (La Quatra et al. 2024) tööst ja on kättesaadav nende GitHubi repositooriumist *SSL4PR* (<https://github.com/K-STMLab/SSL4PR/>). Valituks osutus autorite poolt eelnevalt PC-GITA andmestikul peenhäälestatud *WavLM Base mudel*. Antud töö raames ei treenitud ega peenhäälestatud seda mudelit uuesti, vaid rakendati otse autorite loodud, 10-korda ristvalideerimise struktuuriga hindamisskripti. Eestikeelsed eeltöödeldud andmed sisestati treenitud mudelisse. Skript töötles andmeid läbi kõigi kümne eelnevalt treenitud korduse, arvutas igas korras klassifitseerimismõõdikud AUC, õigsus, F1, saagis (*recall*), tundlikkus, spetsiifilisus ning väljastas lõpptulemusena nende mõõdikute keskmised väärtused ja standardhälbed üle kõigi korduste. Selline *out-of-the-box* testimine võimaldas hinnata, kui hästi mujal peenhäälestatud SOTA SSL-mudel üldistub meie spetsiifilisele eestikeelsele andmestikule, ning võrrelda selle jõudlust lihtsamate, akustilistel tunnustel põhinevate mudelitega (U1, U2).

### 4.3 Eksperimentide läbiviimine ja tunnuste skaleerimine

Eksperimentides U1 ja U2, mis põhinesid eraldatud akustilistel tunnustel, oli oluliseks sammuks tunnuste standardiseerimine. Selleks kasutati 'StandardScaler' funktsiooni *scikit-learn* teegist (Pedregosa, Varoquaux, et al. 2011). Standardiseerimine teisendab tunnused nii, et nende keskmine on ligikaudu 0 ja standardhälve 1. See on vajalik samm, kuna tagab, et kõik tunnused on võrreldavas skaalas, mis on kriitilise tähtsusega paljude masinõppe algoritmide, eriti kauguspõhiste või gradientlaskumist kasutavate meetodite optimaalseks toimimiseks (Kotsiantis et al. 2006). Andmelekkete vältimiseks testandmestikku kohandati skaleerija alati ainult vastava ristvalideerimise korduse treeningandmetel. Seejärel rakendati sama kohandatud skaleerijat nii treening- kui ka testandmetele.

#### 4.3.1 U1: treenimine PC-GITA-l, testimine mõlemal

Esimese uurimisküsimuse (U1) testimiseks laeti esmalt PC-GITA ja Eesti andmestikud, kasutades eeldefineeritud akustiliste tunnuste nimekirja (vt Lisa 6, kus tunnused on loetletud).

Seejärel jaotati PC-GITA andmestik kuni 10-ks kõnelejast sõltumatuks korduseks, kasutades GroupKFold meetodit, ning viidi läbi ristvalideerimine. Igas korduses kohandati esmalt StandardScaler skaleerija, kasutades *ainult* jooksva korduse PC-GITA treeningandmeid. Pärast seda leiti ühised akustilised tunnused PC-GITA ja Eesti andmetike vahel. Saadud skaleerijaga transformeeriti (skaleeriti) nii PC-GITA treening- ja testandmed kui ka *kogu* eelnevalt töödeldud (sh filtreeritud ja ühiste tunnustega piiratud) eestikeelne andmestik. Järgnevalt trenniti valitud masinõppemudelit SVM või RF skaleeritud PC-GITA treeningandmetel. Lõpuks hinnati trennitud mudeli jõudlust kahel erineval andmestikul: esiteks skaleeritud PC-GITA testandmetel (keelesisene valideerimine) ja teiseks skaleeritud eestikeelsel andmestikul (ristkeelne testimine), eeldusel et Eesti andmestikus oli piisavalt andmeid ja klasse testimiseks.

#### 4.3.2 U2: Treenimine ja testimine liidetud andmestikul

Teise uurimisküsimuse (U2) testimiseks valmistati esmalt ette PC-GITA ja Eesti andmestikud eraldi, kasutades sama eeldefineeritud akustiliste tunnuste nimekirja (vt Lisa 6, kus tunnused on loetletud) ning sarnaseid eeltöötlusvõtteid, nagu kirjeldatud U1 eksperimendi puhul (alajaotus 4.3.1).

Nende kahe grupi vahel leiti ühised akustilised tunnused. Seejärel kombineeriti need kaks töödeldud andmestikku üheks suureks liitandmestikuks. Et tagada kõnelejate ID-de unikaalsus liitandmestikus GroupKFold ristvalideerimise korrektseks toimimiseks, lisati igale kõneleja ID-le vastava päritoluandmestiku tunnus. Seejärel jaotati see prefikseeritud ID-dega liitandmestik kuni 10-ks kõnelejast sõltumatuks korduseks.

Ristvalideerimise igas korduses kohandati esmalt *StandardScaler* skaleerija, kasutades ainult jooksva korduse liitandmestiku treeningosa. Selle skaleerijaga transformeeriti

nii liitandmestiku treening- kui ka testosa andmed. Pärast skaleerimist treeniti valitud masinõppemudelit, kas SVM või RF skaleeritud liitandmestiku treeningosal. Lõpuks hinnati treenitud mudelit sama korduse skaleeritud liitandmestiku testosal. Tulemustena koguti klassifitseerimismõõdikuid ning need esitati keskväärtus ja standardhälvetena üle kõigi valideerunud korduste.

#### 4.4 Hindamismõõdikud ja statistiline analüüs

Mudelite jõudluse hindamiseks kõigis eksperimentides (U1, U2 ja U3) kasutati standardseid binaarse klassifitseerimise mõõdikuid. Nende hulka kuulusid AUC (*Area Under the Receiver Operating Characteristic Curve, AUC*), mis mõõdab mudeli võimet eristada Parkinsoni tõve ja kontrollgrupi vahel sõltumata klassifitseerimislävendist, kusjuures väärtus 1.0 tähistab täiuslikku eristamist ja 0.5 juhuslikku arvamist. Samuti kasutati õigsust (*Accuracy*), mis näitab korrektselt klassifitseeritud juhtude osakaalu kõigist juhtudest. Lisaks hinnati täpsust (*precision*), mis näitab, kui suur osa mudeli poolt positiivseks klassifitseeritud juhtudest olid tegelikult positiivsed ( $TP/(TP + FP)$ ), ning tundlikkust (*Sensitivity*), tuntud ka kui saagis (*recall*), mis näitab positiivse klassi (antud juhul Parkinsoni tõbi) korrektse tuvastamise määra ( $TP/(TP + FN)$ ). Nende kahe, täpsuse ja saagise, kaalutud keskmisena arvutati F1-skoor ( $2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$ ), mis oli oluline eriti tasakaalustamata klassijaotusega andmestike puhul. Veel hinnati spetsiifilisust (*Specificity*), mis näitab negatiivse klassi (antud juhul kontrollgrupp) korrektse tuvastamise määra ( $TN/(TN + FP)$ ).

## 5 Eksperimendid

Käesolevas peatükis esitatakse ja analüüsitakse uurimisküsimustele U1, U2 ja U3 vastamiseks läbi viidud eksperimentide tulemusi. Peatükk keskendub mudelite hindamisele erinevates stsenaariumides, kasutades peatükis 4.4 kirjeldatud hindamismõõdikuid.

### 5.1 U1: Treenimine PC-GITA andmestikul, testimine mõlemal

Esimese uurimisküsimuse U1 eesmärk oli hinnata standardse, akustilistel tunnustel põhineva mudeli üldistusvõimet. Selleks treeniti mudeleid ainult hispaaniakeelsel PC-GITA andmestikul ning testiti nii PC-GITA (keelesisene valideerimine) kui ka eestikeelsel andmestikul (ristkeelne valideerimine). Kasutati kahte klassikalist masinõppemudelit: otsustusmetsa ja tugivektormasinat, mille implementatsioon ja tunnuste eeltöötlus on kirjeldatud peatükkides 4.2.1 ja 4.3

#### 5.1.1 Stsenaarium 1: Kõikidel harjutustel põhinev klassifitseerimine

Selles stsenaariumis kasutati kõiki 37 ühist akustilist tunnust, mis eraldati mõlemast andmestikust, mudelite treenimiseks ja testimiseks. Alljärgnev Tabel 3 esitab valitud RF ja SVM mudelite parameetrite ja nende tulemused PC-GITA ning Eesti andmestikel. Esitatud on keskmised ja standardhälbed (sulgudes) 10-kordse ristvalideerimise tulemusena.

Tabel 3. U1 koondtulemused: RF ja SVM mudelite jõudlus, kasutades 37 ühist tunnust. Treenitud PC-GITA andmestikul ja testitud PC-GITA ning Eesti andmestikel.

Klassif.	Parameetrid	Andmestik	AUC (std)	ACC (std)	F1 (std)	SENS (std)	SPEC (std)
RF	n_est=200	PC-GITA	0.672 (0.029)	0.625 (0.028)	0.617 (0.039)	0.603 (0.066)	0.649 (0.054)
RF	n_est=200	Eesti	0.448 (0.061)	0.464 (0.04)	0.548 (0.051)	0.686 (0.088)	0.261 (0.029)
SVM	C=1, rbf, scale	PC-GITA	0.682 (0.073)	0.637 (0.053)	0.616 (0.068)	0.586 (0.099)	0.688 (0.075)
SVM	C=1, rbf, scale	Eesti	0.552 (0.028)	0.490 (0.007)	0.61 (0.009)	0.835 (0.02)	0.177 (0.011)
SVM	C=10, poly, scale, d3	PC-GITA	0.614 (0.073)	0.592 (0.047)	0.595 (0.057)	0.6 (0.083)	0.585 (0.055)
SVM	C=10, poly, scale, d3	Eesti	0.613 (0.032)	0.494 (0.012)	0.609 (0.014)	0.828 (0.033)	0.189 (0.021)

Tabelist 3 on näha, et PC-GITA andmestikul treenitud ja testitud mudelid saavutasid mõõdukaid tulemusi. SVM klassifikaator RBF tuumaga (C=1.0, gamma=scale) näitas PC-GITA andmestikul parimat AUC väärtust (0.682). Mudelite rakendamisel eestikeelsele andmestikule ei olnud tulemused nii head. Eesti andmestiku puhul oli tundlikkus (SENS) SVM mudelite puhul suhteliselt kõrge, kuid spetsiifilisus (SPEC) väga madal. See viitab, et PC-GITA peal õpitud tunnused ei üldistunud hästi tervete Eesti kontrollisikute häälemustritele.

### 5.1.2 Stsenaarium 2: harjutuste kaupa klassifitseerimine (vokaalid ja monoloog)

Eelmises alampeatükis 5.1.1 analüüsi tulemusi, kus kõik 37 akustilist tunnust olid koon-  
datud ühtseks komplektiks. Selles alampeatükis uuritakse U1 stsenaariumi detailsemalt,  
keskendudes klassifikaatorite hindamisele spetsiifiliste kõneülesannete (harjutuste) kaupa,  
milleks on püsivalt hääldatud vokaalid (/a/, /e/, /i/, /o/, /u/) ning monoloog. Eesmärk on  
välja selgitada, kas need konkreetsed kõneülesanded on Parkinsoni tõve tuvastamisel  
ja selle hääleliste biomarkerite ristkeelsel ülekandmisel informatiivsemad. Nende üle-  
sannete täpsem kirjeldus on toodud peatükkides 3.1.1 ja 3.1.2. Iga kõneülesande jaoks  
eraldati vastavad tunnused ning treeniti ja testiti RF ning SVM klassifikaatoreid.

Alljärgnev Tabel 4 esitab vokaalide ja monoloogi klassifitseerimise tulemused, kus  
iga ülesande ja klassifikaatori kohta on valitud parim tulemus.

Tabel 4. U1 harjutuspõhised tulemused: Vokaalide ja monoloogi klassifitseerimine. Tree-  
nitud PC-GITA, testitud PC-GITA ja Eesti andmestikel.

Harjutus	Klassif.	Parameetrid	Andmestik	AUC (std)	ACC (std)	F1 (std)	SENS (std)	SPEC (std)
Vokaal /a/	RF	n_est=100	PC-GITA	0.530 (0.121)	0.547 (0.089)	0.501 (0.133)	0.472 (0.191)	0.625 (0.154)
Vokaal /a/	RF	n_est=100	Eesti	0.521 (0.134)	0.533 (0.097)	0.495 (0.159)	0.481 (0.248)	0.583 (0.201)
Vokaal /a/	SVM	C=1, rbf, scale	PC-GITA	0.542 (0.118)	0.558 (0.088)	0.524 (0.125)	0.500 (0.183)	0.619 (0.147)
Vokaal /a/	SVM	C=1, rbf, scale	Eesti	0.530 (0.129)	0.531 (0.095)	0.512 (0.147)	0.519 (0.231)	0.542 (0.187)
Vokaal /e/	RF	n_est=100	PC-GITA	0.518 (0.109)	0.528 (0.091)	0.489 (0.127)	0.461 (0.183)	0.597 (0.146)
Vokaal /e/	RF	n_est=100	Eesti	0.505 (0.117)	0.514 (0.092)	0.493 (0.148)	0.481 (0.226)	0.547 (0.183)
Vokaal /e/	SVM	C=1, rbf, scale	PC-GITA	0.529 (0.103)	0.539 (0.087)	0.507 (0.119)	0.483 (0.169)	0.597 (0.135)
Vokaal /e/	SVM	C=1, rbf, scale	Eesti	0.513 (0.112)	0.519 (0.089)	0.506 (0.139)	0.500 (0.211)	0.538 (0.172)
Vokaal /i/	RF	n_est=100	PC-GITA	0.559 (0.115)	0.568 (0.097)	0.535 (0.136)	0.518 (0.199)	0.621 (0.151)
Vokaal /i/	RF	n_est=100	Eesti	0.541 (0.126)	0.542 (0.103)	0.508 (0.164)	0.491 (0.253)	0.592 (0.207)
Vokaal /i/	SVM	C=1, rbf, scale	PC-GITA	0.570 (0.109)	0.577 (0.092)	0.551 (0.128)	0.533 (0.185)	0.623 (0.143)
Vokaal /i/	SVM	C=1, rbf, scale	Eesti	0.553 (0.121)	0.550 (0.099)	0.527 (0.153)	0.519 (0.236)	0.581 (0.193)
Vokaal /o/	RF	n_est=100	PC-GITA	0.537 (0.118)	0.549 (0.093)	0.510 (0.131)	0.483 (0.189)	0.618 (0.149)
Vokaal /o/	RF	n_est=100	Eesti	0.526 (0.130)	0.532 (0.099)	0.499 (0.157)	0.486 (0.241)	0.578 (0.196)
Vokaal /o/	SVM	C=1, rbf, scale	PC-GITA	0.551 (0.112)	0.560 (0.089)	0.529 (0.124)	0.506 (0.178)	0.617 (0.140)
Vokaal /o/	SVM	C=1, rbf, scale	Eesti	0.538 (0.125)	0.540 (0.096)	0.518 (0.149)	0.509 (0.228)	0.571 (0.185)
Vokaal /u/	RF	n_est=100	PC-GITA	0.542 (0.112)	0.553 (0.090)	0.518 (0.129)	0.494 (0.185)	0.614 (0.144)
Vokaal /u/	RF	n_est=100	Eesti	0.530 (0.122)	0.536 (0.095)	0.504 (0.153)	0.491 (0.237)	0.581 (0.191)
Vokaal /u/	SVM	C=1, rbf, scale	PC-GITA	0.555 (0.107)	0.563 (0.087)	0.534 (0.121)	0.512 (0.174)	0.615 (0.137)
Vokaal /u/	SVM	C=1, rbf, scale	Eesti	0.541 (0.118)	0.543 (0.092)	0.523 (0.146)	0.514 (0.223)	0.572 (0.181)
Monoloog	RF	n_est=100	PC-GITA	0.603 (0.085)	0.597 (0.071)	0.582 (0.093)	0.572 (0.135)	0.623 (0.108)
Monoloog	RF	n_est=100	Eesti	0.551 (0.099)	0.540 (0.078)	0.537 (0.115)	0.565 (0.173)	0.515 (0.142)
Monoloog	SVM	C=1, rbf, scale	PC-GITA	0.615 (0.080)	0.608 (0.068)	0.599 (0.087)	0.589 (0.127)	0.628 (0.101)
Monoloog	SVM	C=1, rbf, scale	Eesti	0.563 (0.093)	0.552 (0.073)	0.558 (0.108)	0.583 (0.161)	0.521 (0.133)

Tabelist 4 ilmneb harjutuspõhise analüüsi tulemusena, et üksikute püsivalt hääldatud  
vokaalide (/a/, /e/, /i/, /o/, /u/) eristusvõime jäi nii PC-GITA kui ka Eesti andmestikel  
tagasihoidlikuks. AUC väärtused PC-GITA andmestikul jäid enamasti vahemikku 0.52  
kuni 0.57 ning Eesti andmestikul olid need veelgi madalamad, sageli lähedal 0.50-le,  
mis viitab peaaegu juhuslikule klassifitseerimisele. Näiteks vokaali /i/ puhul saavutas  
SVM PC-GITA andmestikul AUC 0.570 ja Eesti andmestikul 0.553. Teiste vokaalide  
tulemused olid sarnased või veidi madalamad. Vokaalide puhul ei ilmnunud ühegi  
konkreetses vokaali selget paremust teiste ees ning standardhälbed olid sageli suured, mis  
viitab tulemuste ebastabiilsusele.

Monoloogi ülesanne näitas mõnevõrra paremaid tulemusi kui üksikud vokaalid. PC-GITA andmestikul saavutas SVM monoloogi puhul AUC väärtuse 0.615 ja RF 0.603. Ka Eesti andmestikul oli monoloogi AUC vastavalt 0.563 (SVM) ja 0.551 (RF), mis on küll madal, kuid siiski parem kui enamike vokaalide puhul. See võib tuleneda sellest, et monoloog sisaldab rikkalikumat ja mitmekesisemat kõneinformatsiooni (prosoodia, dünaamika, pikemad helilõigud) võrreldes lühikeste staatiliste vokaalidega.

Klassifikaatorite võrdluses ei ilmnenu selget eelistust RF või SVM vahel harjutuspõhises analüüsis; nende jõudlus oli sageli sarnane antud harjutuse piires. Sarnaselt koondtunnuste analüüsile (Alampeatükk 5.1.1), jäi ka harjutuspõhises analüüsis Eesti andmetel tundlikkus sageli kõrgemaks kui spetsiifilisus, eriti SVM klassifikaatorite puhul, mis viitab jätkuvalt paljudele valepositiivsetele ennustustele Eesti kontrollgrupi osas. Ükski testitud üksikharjutus ei näidanud piisavalt kõrget ja robustset jõudlust.

Uurimisküsimuse U1 teise stsenaariumi detailsem analüüs, mis keskendus üksikutele vokaalidele ja monoloogile, näitab, et kuigi monoloogi-tüüpi ülesanded võivad sisaldada rohkem informatsiooni Parkinsoni tõve tuvastamiseks kui üksikud lühikesed häälused nagu vokaalid, on klassikaliste masinõppemudelite ja akustiliste tunnuste ristkeelne ülekantavus ka harjutuspõhiselt endiselt suur väljakutse. Ükski testitud üksikharjutus ei näidanud piisavalt kõrget ja robustset jõudlust, et olla iseseisvalt usaldusväärne biomarker ristkeelses stsenaariumis.

### 5.1.3 Stsenaarium 3: Erinevate harjutuste andmepunktide kombineerimine

Kolmandas stsenaariumis uuriti, kas teatud kõneülesannete kombineerimine võib parandada Parkinsoni tõve tuvastamise õigsust võrreldes üksikute ülesannete või kõikide andmepunktide koondamisega. Testiti kahte peamist harjutuste kombinatsiooni:

- **Kombinatsioon 1 (K1):** Vokaalid, moduleeritud vokaalid, monoloog.
- **Kombinatsioon 2 (K2):** Vokaalid, moduleeritud vokaalid.

Nende kombinatsioonide jaoks eraldati vastavad tunnused ning treeniti ja testiti RF ning SVM klassifikaatoreid. Alljärgnev Tabel 5 esitab valitud klassifikaatorite tulemused nende kombinatsioonide kohta ning esiletoodud tulemused on kõige paremad.

Tabelist 5 on näha, et harjutuste kombineerimine andis PC-GITA andmestikul tulemusi, mis on võrreldavad või mõnevõrra paremad kui üksikute harjutuste (Tabel 4) puhul, kuid jäävad siiski alla kõikide 37 ühise tunnuse kasutamisele (Tabel 3). Kombinatsioon 1 (K1: Vokaalid, moduleeritud vokaalid, monoloog) saavutas SVM klassifikaatoriga PC-GITA andmestikul AUC väärtuse **0.708**, mis on parim tulemus selles stsenaariumis. Kombinatsioon 2 (K2: Vokaalid, moduleeritud vokaalid), kus monoloog oli eemaldatud, andis SVM-iga veidi madalama AUC **0.605**. See viitab, et monoloogi kaasamine kombinatsiooni annab täiendavat informatsiooni. RF klassifikaatorid andsid mõlema kombinatsiooni puhul veidi madalamaid tulemusi kui SVM.

Tabel 5. U1 kombineeritud harjutuste tulemused. Treenitud PC-GITA, testitud PC-GITA ja Eesti andmestikel.

Kombinatsioon	Klassif.	Parameetrid	Andmestik	AUC (std)	ACC (std)	F1 (std)	SENS (std)	SPEC (std)
K1: Vok, ModVok, Monol	SVM	C=10, rbf, scale	PC-GITA	0.695 (0.114)	0.634 (0.082)	0.616 (0.101)	0.604 (0.111)	0.656 (0.097)
K1: Vok, ModVok, Monol	SVM	C=10, rbf, scale	Eesti	0.632 (0.011)	0.589 (0.014)	0.649 (0.016)	0.75 (0.033)	0.423 (0.029)
K2: Vok, ModVok	SVM	C=10, rbf, scale	PC-GITA	0.605 (0.079)	0.597 (0.053)	0.576 (0.08)	0.564 (0.085)	0.617 (0.07)
K2: Vok, ModVok	SVM	C=10, rbf, scale	Eesti	0.61 (0.022)	0.557 (0.024)	0.603 (0.026)	0.665 (0.045)	0.446 (0.043)
K1: Vok, ModVok, Monol	SVM	C=10, poly, deg=3, scale	PC-GITA	0.708 (0.129)	0.654 (0.098)	0.64 (0.116)	0.638 (0.11)	0.67 (0.093)
K1: Vok, ModVok, Monol	SVM	C=10, poly, deg=3, scale	Eesti	0.602 (0.024)	0.578 (0.0117)	0.645 (0.014)	0.758 (0.035)	0.393 (0.032)
K1: Vok, ModVok, Monol	RF	n_est=300	PC-GITA	0.651 (0.069)	0.613 (0.065)	0.596 (0.083)	0.581 (0.096)	0.639 (0.072)
K1: Vok, ModVok, Monol	RF	n_est=300	Eesti	0.492 (0.027)	0.509 (0.037)	0.55 (0.043)	0.594 (0.061)	0.421 (0.044)
K2: Vok, ModVok	RF	n_est=200	PC-GITA	0.641 (0.069)	0.604 (0.063)	0.587 (0.079)	0.573 (0.094)	0.628 (0.07)
K2: Vok, ModVok	RF	n_est=200	Eesti	0.494 (0.024)	0.503 (0.035)	0.542 (0.04)	0.583 (0.057)	0.42 (0.042)
K1: Vok, ModVok, Monol	RF	n_est=50	PC-GITA	0.636 (0.068)	0.608 (0.063)	0.591 (0.08)	0.575 (0.092)	0.632 (0.07)
K1: Vok, ModVok, Monol	RF	n_est=50	Eesti	0.493 (0.025)	0.506 (0.026)	0.539 (0.032)	0.577 (0.052)	0.44 (0.049)

Ristkeelsel testimisel Eesti andmestikul jäid tulemused sarnaselt eelmistele stsenaariumidele tagasihoidlikuks. Kombinatsioon 1 (SVM) saavutas Eesti andmestikul AUC **0.632** ja kombinatsioon 2 (SVM) AUC **0.61**. Kuigi K1 oli ka siin veidi parem, on mõlemad väärtused madalad ning näitavad piiratud üldistusvõimet. Jätakuvalt ilmnes probleem madala spetsiifilisusega Eesti andmetel.

Võrreldes kolme U1 stsenaariumi, võib järeldada, et kuigi teatud harjutuste nagu monoloog või nende kombinatsioonide kasutamine võib anda marginaalselt paremaid tulemusi kui kõige lihtsamad üksikud harjutused (nagu staatilised vokaalid), ei ületa nende jõudlus oluliselt kõikide olemasolevate ühiste tunnuste korraga kasutamist (stsenaarium 1), eriti PC-GITA andmestikul. Ristkeelses kontekstis (Eesti andmed) jäi aga kõigi U1 stsenaariumide puhul klassikaliste masinõppemudelite ja akustiliste tunnuste põhine lähenemine piiratuks, rõhutades vajadust täpsemate meetodite või keelespetsiifilise kohandamise järele.

## 5.2 U2: Treenimine ja testimine liidetud andmestikul

Teise uurimisküsimuse (U2) eesmärk oli selgitada, kas hispaania- (PC-GITA) ja eestikeelsete andmete kombineerimine treeningfaasis parandab akustilistel tunnustel põhineva mudeli jõudlust võrreldes ainult ühel keelel (PC-GITA) treenimisega (nagu uuritud U1 raames, Alampeatükk 5.1). Selleks liideti PC-GITA ja eestikeelne andmestik üheks suureks andmestikuks, millel seejärel trenniti ja testiti RF ning SVM klassifikaatoreid, kasutades 10-kordset kõnelejast sõltumatut ristvalideerimist.

### 5.2.1 Stsenaarium 1: Kõik ühised tunnused liidetud andmestikul

Selles stsenaariumis kasutati samu 37 ühist akustilist tunnust nagu U1 eksperimentides, kuid mudelite treenimine ja testimine viidi läbi kombineeritud PC-GITA ja Eesti andmestikul. Alljärgnev Tabel 6 esitab valitud RF ja SVM mudelite parimad tulemused sellel liitandmestikul. Esitatud on keskmised ja standardhälbed (sulgudes).

Tabel 6. U2 koondtulemused: RF ja SVM mudelite tulemused liidetud (PC-GITA + Eesti) andmestikul (37 ühist tunnust).

Klassif.	Parameetrid	AUC (std)	ACC (std)	F1 (std)	SENS (std)	SPEC (std)
RF	n_est=100, max_depth=None	0.661 (0.058)	0.62 (0.04)	0.597 (0.066)	0.578 (0.073)	0.659 (0.074)
RF	n_est=200, max_depth=None	0.666 (0.06)	0.623 (0.038)	0.602 (0.066)	0.584 (0.076)	0.66 (0.077)
RF	n_est=300, max_depth=None	0.666 (0.06)	<b>0.625 (0.039)</b>	0.605 (0.066)	0.588 (0.078)	0.659 (0.078)
RF	n_est=50, max_depth=None	0.651 (0.056)	0.61 (0.041)	0.58 (0.066)	0.552 (0.072)	0.666 (0.069)
SVM	RBF, C=10, gamma=scale	<b>0.666 (0.037)</b>	0.621 (0.026)	<b>0.605 (0.039)</b>	<b>0.590 (0.042)</b>	0.645 (0.069)
SVM	Poly, C=10, degree=3, gamma=scale	0.621 (0.032)	0.594 (0.025)	0.572 (0.052)	0.555 (0.054)	0.625 (0.045)

Tabelist 6 nähtuvad tulemused, mis on saadud PC-GITA ja Eesti andmestike kombineerimisel ning mudelite treenimisel ja testimisel sellel liitandmestikul. Parimaid tulemusi näitasid RF klassifikaator (parameetritega n\_estimators=300, max\_depth=None), saavutades AUC 0.666 ja ACC 0.625, ning SVM klassifikaator RBF tuumaga (parameetritega C=10, gamma=scale), saavutades AUC 0.666 ja ACC 0.621. Mõlemad klassifikaatorid näitasid suhteliselt head tasakaalu tundlikkuse ja spetsiifilisuse vahel, mis jäid enamasti vahemikku 0.55 kuni 0.66.

Kõige olulisem on võrrelda neid tulemusi U1 eksperimendi tulemustega (Tabel 3), kus mudeleid treeniti ainult PC-GITA andmestikul ja testiti ristkeelselt Eesti andmestikul. U1 raames oli SVM (RBF tuumaga) parim AUC Eesti andmetel 0.55 ja RF puhul 0.448. U2 tulemused on nendest väärtustest märkimisväärselt kõrgemad. See viitab selgelt, et eesti-keelsete andmete lisamine treeninghulka parandas oluliselt mudeli võimekust eristada Parkinsoni tõvega patsiente tervetest kontrollidest. Kombineeritud treening võimaldas mudelitel õppida robustsemaid ja üldistatavamaid tunnuseid, mis on relevantes mõlema keele ja andmestiku jaoks. Lisaks paranes oluliselt tundlikkuse ja spetsiifilisuse tasakaal võrreldes U1 Eesti andmete tulemustega, kus spetsiifilisus oli väga madal.

Kokkuvõttes näitavad U2 esimese stsenaariumi tulemused, et hispaania- ja eesti-keelsete andmete kombineerimine treeningfaasis omab märkimisväärset positiivset mõju klassikaliste masinõppemudelite jõudlusele Parkinsoni tõve tuvastamisel, parandades oluliselt nii üldist õigsust kui ka tundlikkuse ja spetsiifilisuse tasakaalu võrreldes mudelitega, mis on treenitud vaid ühel keelel ja rakendatud teisele.

### 5.2.2 Stsenaarium 2: Harjutuste kaupa klassifitseerimine liidetud andmestikul (vokaalid ja monoloog)

Analoogselt U1 harjutuspõhisele analüüsile (Alampeatükk 5.1.2), uuriti ka U2 raames klassifikaatorite jõudlust spetsiifiliste kõneülesannete kaupa, kuid seekord treeniti ja testiti mudeleid liidetud (PC-GITA + Eesti) andmestikul. Analüüs keskendus püsivalt häälstatud vokaalidele (/a/, /e/, /i/, /o/, /u/) ning monoloogile. Eesmärk oli hinnata, kas andmestike liitmine parandab üksikute harjutuspõhist tuvastusõigsust võrreldes U1 ristkeelse testimisega ning kas mõni neist ülesannetest osutub liitandmestikul eriti informatiivseks.

Alljärgnev Tabel 7 esitab vokaalide ja monoloogi klassifitseerimise tulemused liidetud andmestikul, kus iga ülesande ja klassifikaatori kohta.

Tabel 7. U2 harjutuspõhised tulemused: Vokaalide ja monoloogi klassifitseerimise tulemused liidetud (PC-GITA + Eesti) andmestikul.

Harjutus	Klassif.	Parameetrid	AUC (std)	ACC (std)	F1 (std)	SENS (std)	SPEC (std)
Vokaal /a/	RF	n_est=200	<b>0.712 (0.097)</b>	<b>0.683 (0.094)</b>	<b>0.674 (0.097)</b>	<b>0.671 (0.129)</b>	<b>0.693 (0.138)</b>
Vokaal /a/	SVM	C=1, rbf, scale	0.713 (0.127)	0.671 (0.106)	0.658 (0.109)	0.646 (0.138)	0.69 (0.147)
Vokaal /a/	RF	n_est=300	0.716 (0.093)	0.668 (0.091)	0.658 (0.093)	0.65 (0.122)	0.686 (0.123)
Vokaal /e/	SVM	C=1, rbf, scale	<b>0.666 (0.149)</b>	<b>0.626 (0.116)</b>	<b>0.616 (0.13)</b>	<b>0.622 (0.157)</b>	<b>0.617 (0.148)</b>
Vokaal /e/	RF	n_est=300	0.668 (0.119)	0.614 (0.084)	0.588 (0.115)	0.569 (0.133)	0.658 (0.124)
Vokaal /e/	RF	n_est=200	0.664 (0.111)	0.602 (0.099)	0.583 (0.109)	0.564 (0.146)	0.642 (0.127)
Vokaal /i/	SVM	C=1, rbf, scale	<b>0.701 (0.081)</b>	<b>0.666 (0.063)</b>	<b>0.64 (0.101)</b>	<b>0.618 (0.125)</b>	<b>0.719 (0.101)</b>
Vokaal /i/	RF	n_est=300	0.682 (0.04)	0.636 (0.037)	0.63 (0.041)	0.62 (0.055)	0.651 (0.048)
Vokaal /i/	SVM	C=0.1, rbf, scale	0.683 (0.088)	0.63 (0.065)	0.617 (0.085)	0.596 (0.103)	0.675 (0.084)
Vokaal /o/	RF	n_est=300	<b>0.681 (0.084)</b>	<b>0.658 (0.073)</b>	<b>0.642 (0.083)</b>	<b>0.624 (0.099)</b>	<b>0.689 (0.095)</b>
Vokaal /o/	SVM	C=1, rbf, scale	0.676 (0.101)	0.638 (0.081)	0.627 (0.086)	0.607 (0.104)	0.669 (0.098)
Vokaal /o/	RF	n_est=200	0.657 (0.046)	0.616 (0.041)	0.608 (0.047)	0.597 (0.065)	0.636 (0.055)
Vokaal /u/	SVM	C=1, rbf, scale	<b>0.683 (0.0411)</b>	<b>0.636 (0.0365)</b>	<b>0.629 (0.041)</b>	<b>0.618 (0.054)</b>	<b>0.654 (0.047)</b>
Vokaal /u/	RF	n_est=300	0.671 (0.043)	0.629 (0.038)	0.621 (0.043)	0.611 (0.059)	0.647 (0.05)
Vokaal /u/	SVM	C=0.1, rbf, scale	0.679 (0.072)	0.62 (0.067)	0.61 (0.071)	0.589 (0.091)	0.661 (0.075)
Monoloog	SVM	C=1, rbf, scale	<b>0.749 (0.032)</b>	<b>0.686 (0.032)</b>	<b>0.681 (0.036)</b>	<b>0.672 (0.046)</b>	<b>0.7 (0.04)</b>
Monoloog	RF	n_est=300	0.736 (0.034)	0.675 (0.033)	0.67 (0.038)	0.662 (0.05)	0.688 (0.042)
Monoloog	SVM	C=0.1, rbf, scale	0.741 (0.038)	0.669 (0.035)	0.664 (0.037)	0.655 (0.051)	0.684 (0.041)

Tabelist 7 on näha, et liidetud (PC-GITA + Eesti) andmestikul treenides paranes üksikute harjutuspõhine klassifitseerimisjõudlus märgatavalt võrreldes U1 ristkeelse testimisega (kus mudeleid treeniti ainult PC-GITA andmestikul ja testiti Eesti andmestikul, vt Tabel 4).

Üksikute vokaalide puhul jäid AUC väärtused liitandmestikul vahemikku ligikaudu 0.65 kuni 0.69 nii RF kui ka SVM klassifikaatoritega. Parimaid tulemusi vokaalidest näitas vokaal /i/, mille puhul SVM saavutas AUC 0.701 ja RF AUC 0.682. See on oluline paranemine võrreldes U1 Eesti andmete tulemustega, kus vokaali /i/ SVM AUC oli 0.55. Ka teiste vokaalide (/a/, /e/, /o/, /u/) puhul oli jõudluse kasv märgatav, ületades sageli AUC 0.65 piiri.

Monoloogi ülesanne osutus ka liitandmestikul kõige informatiivsemaks üksikuks harjutusteks. SVM klassifikaator saavutas monoloogi puhul AUC väärtuse 0.748 ja RF AUC 0.736. Need tulemused on selgelt paremad kui üksikutel vokaalidel ning näitavad, et pikem ja mitmekesisem kõneülesanne võimaldab mudelitel paremini õppida haigusega seotud mustreid, kui treeningandmestik on piisavalt suur ja mitmekesine.

Võrreldes U1 harjutuspõhiste tulemustega, eriti U1 Eesti andmestiku osas, on peamine erinevus U2 harjutuspõhises analüüsis oluliselt paranenud AUC ja ACC väärtused ning ka parem tasakaal tundlikkuse (SENS) ja spetsiifilisuse (SPEC) vahel. Kui U1 ristkeelses testimises oli spetsiifilisus sageli väga madal, siis liitandmestikul treenides on nii tundlikkus kui ka spetsiifilisus enamasti **0.60-0.70** vahemikus, mis on praktilisest vaatepunktist palju paljulubavam. Näiteks monoloogi puhul SVM-iga oli SENS **0.672** ja

## SPEC 0.7.

Klassifikaatorite osas näitas SVM sageli veidi paremaid tulemusi kui RF, kuigi erinevused ei olnud drastilised.

Uurimisküsimuse U2 teise stsenaariumi analüüs kinnitab, et andmestike liitmine avaldab tugevat positiivset mõju ka üksikute kõneülesannete tuvastusõigsusele. Monoloog jääb kõige informatiivsemaks üksikuks ülesandeks, kuid ka üksikute vokaalide (eriti /i/ ja /u/) eristusvõime paraneb märgatavalt, kui treeningandmestik on mitmekesisem. Tulemused on oluliselt paremad kui U1 ristkeelse testimise puhul, mis rõhutab piisava treeningandmestiku olulisust.

### 5.2.3 Stsenaarium 3: Kombineeritud harjutused liidetud andmestikul

Uurimisküsimuse U2 raames analüüsiti lisaks üksikutele harjutustele ka teatud kõneülesannete kombinatsioonide mõju Parkinsoni tõve tuvastamisele liidetud andmestikul. Eesmärk oli hinnata, kas teadlikult valitud harjutuste kombinatsioonid võivad anda paremaid tulemusi kui kõikide tunnuste lihtne koondamine (nagu stsenaariumis 5.2.1) või üksikud harjutused (nagu stsenaariumis 5.2.2). Testiti kahte peamist harjutuste kombinatsiooni, mille "Stsenaariumi Kirjeldus" oli vastavalt:

- **Kombinatsioon 1 (K1):** Vokaalid, moduleeritud vokaalid, monoloog.
- **Kombinatsioon 2 (K2):** Vokaalid, moduleeritud vokaalid.

Nende kombinatsioonide jaoks eraldati vastavad tunnused ning treeniti ja testiti RF ning SVM klassifikaatoreid. Alljärgnev Tabel 8 esitab valitud klassifikaatorite parimad tulemused nende kombinatsioonide kohta liidetud andmestikul.

Tabel 8. U2 kombineeritud harjutuste tulemused liidetud andmestikul.

Kombinatsioon	Klassif.	Parameetrid	AUC (std)	ACC (std)	FI (std)	SENS (std)	SPEC (std)
K1: Vok, ModVok, Monol	RF	n_est=300	0.776 (0.029)	0.704 (0.03)	0.7 (0.034)	0.692 (0.044)	0.717 (0.041)
K1: Vok, ModVok, Monol	SVM	C=1, rbf, scale	0.789 (0.027)	0.718 (0.028)	0.712 (0.032)	0.703 (0.041)	0.732 (0.037)
K2: Vok, ModVok	RF	n_est=200	0.75 (0.032)	0.683 (0.033)	0.677 (0.037)	0.669 (0.049)	0.698 (0.044)
K2: Vok, ModVok	SVM	C=0.1, rbf, scale	0.764 (0.031)	0.696 (0.031)	0.69 (0.035)	0.68 (0.045)	0.711 (0.041)
K1: Vok, ModVok, Monol	SVM	C=10, poly, deg=3, scale	0.61 (0.073)	0.583 (0.05)	0.553 (0.086)	0.529 (0.096)	0.63 (0.062)
K2: Vok, ModVok	SVM	C=10, poly, deg=3, scale	0.61 (0.073)	0.583 (0.05)	0.553 (0.086)	0.529 (0.096)	0.63 (0.062)
K1: Vok, ModVok, Monol	SVM	C=10, rbf, scale	<b>0.673 (0.065)</b>	<b>0.621 (0.049)</b>	<b>0.598 (0.077)</b>	<b>0.576 (0.079)</b>	<b>0.659 (0.058)</b>
K2: Vok, ModVok	SVM	C=10, rbf, scale	<b>0.673 (0.065)</b>	<b>0.621 (0.049)</b>	<b>0.598 (0.077)</b>	<b>0.576 (0.079)</b>	<b>0.659 (0.058)</b>
K1: Vok, ModVok, Monol	SVM	C=10, rbf, scale (RunID: VMM)	<b>0.702 (0.071)</b>	<b>0.65 (0.058)</b>	<b>0.636 (0.08)</b>	<b>0.629 (0.093)</b>	<b>0.661 (0.09)</b>

Tabelist 8 on näha, et harjutuste kombineerimine liidetud andmestikul andis tugevaid tulemusi, mis mõnel juhul lähenevad või on võrreldavad kõikide 37 ühise tunnuse kasutamisega saadud tulemustega (Tabel 6).

Kombinatsioon 1 (K1: Vokaalid, moduleeritud vokaalid, monoloog) saavutas SVM klassifikaatoriga (parameetritega C=1, kernel=rbf, gamma=scale) parima AUC väärtuse 0.789 (std 0.027) ja õigsuse 0.718 (std 0.028). *Random Forest* (parameetritega n\_estimators=300) andis samuti hea tulemus AUC 0.776 (std 0.029). Võrreldes U2

esimese stsenaariumi tulemustega, kus kõikide ühiste tunnuste kasutamisel saavutas SVM klassifikaator parimaks AUC väärtuseks 0.666, on siinse K1 kombinatsiooni tulemused SVM mudelil AUC 0.789 oluliselt paremad. See näitab, et spetsiifiliste, informatiivsete harjutuste kombinatsioon liitandmestikul andis tugevama eristusvõime kui kõikide tunnuste lihtne koondamine.

Kombinatsioon 2 (K2: Vokaalid, moduleeritud vokaalid), kus monoloog oli eemaldatud, näitas oodatult mõnevõrra madalamaid tulemusi. SVM (parameetritega  $C=0.1$ ,  $\text{kernel}=\text{rbf}$ ,  $\text{gamma}=\text{scale}$ ) saavutas K2 puhul AUC 0.764 (std 0.031) ja RF ( $n\_estimators=200$ ) AUC 0.75 (std 0.032). Erinevus K1 ja K2 tulemuste vahel (nt SVM AUC K1 0.789 vs K2 0.764) kinnitab monoloogi kui informatiivse kõneülesande olulisust, mis lisab eristusvõimet ka siis, kui kombineerida seda vokaaliharjutustega.

Võrreldes U2 raames eelnevalt analüüsitud lähenemistega, eriti kõikide ühiste akustiliste tunnuste korruga kasutamisega, kui kõikide tunnuste koondamisel jäi parim SVM AUC 0.666 tasemele, siis vokaale, moduleeritud vokaale ja monoloogi ühendav kombinatsioon K1 saavutas märgatavalt kõrgema SVM AUC (0.789), mis rõhutab sellise sihitud tunnuste valiku potentsiaalset eelist.

Kõigi U2 stsenaariumide tulemused näitavad selget paranemist võrreldes U1 ristkeelse testimisega. Andmestike liitmine on oluliselt parandanud mudelite võimet tuvastada Parkinsoni tõbe ja saavutada parem tasakaal tundlikkuse ning spetsiifilisuse vahel.

Uurimisküsimuse U2 kolmanda stsenaariumi põhjal võib järeldada, et monoloogi sisaldav harjutuste kombinatsioon (K1) oli kõige efektiivsem U2 raames testitud akustilistel tunnustel põhinevatest meetoditest, ületades nii ainult vokaale ja moduleeritud vokaale sisaldavat kombinatsiooni (K2) kui ka kõikide 37 ühise tunnuse kasutamist liitandmestikul (U2.1)

U2 eksperimendid, kus PC-GITA ja Eesti andmestikud liideti treeninghulgaks, näitasid märkimisväärset positiivset mõju mudelite jõudlusele võrreldes U1 ristkeelse testimisega. Kui kõikide 37 ühise tunnuse kasutamisel liitandmestikul saavutasid SVM ja RF klassifikaatorid AUC väärtused vastavalt 0.666 ja 0.666, siis veelgi paremaid tulemusi andsid spetsiifilisemad lähenemised liitandmestikul. harjutuspõhine analüüs näitas, et monoloog osutus kõige informatiivsemaks üksikuks ülesandeks (SVM AUC 0.749). Kõige tugevamad tulemused saavutati aga harjutuste kombinatsiooniga, mis sisaldas vokaale, moduleeritud vokaale ja monoloogi (K1), kus SVM saavutas AUC 0.789. Need leiud viitavad selgelt, et treeningandmestiku mitmekesistamine erinevate keelte andmetega parandab oluliselt mudelite võimekust, kusjuures parimaid tulemusi võib anda just spetsiifiliste, informatiivsete kõneülesannete või nende kombinatsioonide kasutamine.

### 5.3 U3: Mujal peenhäälestatud SSL-mudeli otserakendatavus Eesti andmetel

Kolmanda uurimisküsimuse (U3) eesmärk oli hinnata tiptasemel, isejuhendatud õppel põhineva Parkinsoni tõve tuvastamiseks peenhäälestatud mudeli jõudlust otse eestikeelsele hääleandmetel. Selleks kasutati La Quatra jt (2024) (La Quatra et al. 2024) poolt välja töötatud SSL4PR raamistiku mudeleid. SSL4PR projekt kasutab siirdeõpet, et rakendada suurtes kõnekorpustes eeltreenitud alusmudeleid (ingl *foundational models*) Parkinsoni tõve häälepõhisele tuvastamisele.

Testiti kahte SSL4PR projekti WavLM Base põhjal loodud mudelit, mis on avalikult kättesaadavad Hugging Face platvormil:

- **SSL4PR WavLM Base** (moreno1q/SSL4PR-wavlm-base): Mudelite kogum, mis on treenitud ja valideeritud PC-GITA andmestikul 10-kordse ristvalideerimisega. Iga korduse (ingl *fold*) jaoks on eraldi mudel.
- **SSL4PR WavLM Base Full** (moreno1q/SSL4PR-wavlm-base-full): Üksik mudel, mis on treenitud kogu PC-GITA standardse treeningandmestiku (s-PC-GITA) peal.

Mõlemad mudelid on algselt eeltreenitud *WavLM Base* arhitektuuriga ning seejärel peenhäälestatud La Quatra jt poolt spetsiifiliselt Parkinsoni tõve tuvastamiseks hispaaniakeelsel PC-GITA andmestikul.

Mudel rakendati otse eestikeelsetele andmetele ilma täiendava peenhäälestamiseta või kohandamiseta Eesti andmestikul. Testimiseks kasutati eestikeelse andmestiku eeltöödelatud (vt alapeatükk 3.3) vokaalide hääldamise ja monoloogi ülesannetest pärinevaid helilõike. SSL4PR WavLM Base (10-kordne) mudelite puhul kasutati autorite poolt pakutavat hindamiskripti, mis töötles eestikeelseid andmeid läbi kõigi kümne eelnevalt treenitud mudelikorduse ning arvutas keskmised tulemusmõõdikud. SSL4PR WavLM Base Full mudelit testiti sarnaselt, kasutades ühtset treenitud mudelit.

Ekspirimendi tulemused näitasid, et SSL4PR WavLM Base ja WavLM Base Full mudelite otsene rakendamine eestikeelsetele lühikestele helilõikudele ei andnud edukaid tulemusi Parkinsoni tõve ja kontrollgrupi eristamisel. Mõlema testitud mudeli puhul jäi AUC ligikaudu 0.5 lähedale, mis viitab juhuslikule klassifitseerimisele. Detailsemal vaatlusel ilmsel, et mudelid kaldusid ainult ennustama ühte klassi ehk parkinsoni tõbe põdevat, mis tähendas madalat spetsiifilisust ja F1-skoori.

Saadud negatiivne tulemus on siiski oluline leid. See rõhutab, et isegi tiptasemel SSL-põhiste alusmudelite peal peenhäälestatud spetsiifilised mudelid ei pruugi uues keele- ja andmekontekstis ilma täiendava kohandamiseta oodatud tulemusi saavutada. Võimalikud madala jõudluse põhjused võivad olla mitmesugused, sealhulgas:

- Olulised erinevused treening- (hispaania keel, PC-GITA) ja sihtandmestiku (eesti keel, käesoleva töö andmestik) keelelistes ning akustilistes omadustes.

- Kasutatud eestikeelsete helilõikude potentsiaalsed erinevused (nt kestus, ülesande spetsiifika) võrreldes PC-GITA andmestiku segmentidega, millele mudelid algelt optimeeriti.

Need tulemused on kooskõlas varasemate uuringutega (nt (Favaro et al. 2023)), mis on näidanud, et kuigi SSL-mudelid omavad head potentsiaali ristkeelseks siirdeõppeks, on parimate tulemuste saavutamiseks sageli vajalik vähemalt mingil määral mudeli kohandamine sihtkeele või -andmestiku spetsiifikale. Seega, vaatamata SSL-mudelite võimekusele õppida üldisi helirepresentatsioone, jääb nende otserakendatavus *zero-shot transfer* keerukates meditsiinilistes rakendustes, nagu Parkinsoni tõve tuvastamine, väljakutseks.

## 6 Järeldused

Antud bakalaureusetöö keskendus Parkinsoni tõve tuvastamisele eestikeelsete häälek-lippide analüüsi abil, kasutades erinevaid masinõppe meetodeid ja stsenaariume. Töö peamiseks eesmärgiks oli uurida, kuidas rahvusvahelistes uuringutes levinud lähenemised kohanduvad ja toimivad spetsiifiliselt eesti keele kontekstis ning milline on erinevate strateegiate (ristkeelne testimine, andmestike kombineerimine, tipptasemel SSL-mudelite rakendamine) potentsiaal. Selleks püstitati kolm uurimisküsimust, millele otsiti vastuseid läbi süstemaatiliste eksperimentide.

Vastused uurimisküsimustele ja peamised tulemused on järgmised:

**Uurimisküsimus 1 (U1): Kui hästi üldistuvad standardsetel akustilistel tunnustel põhinevad mudelid, mis on treenitud hispaaniakeelsel PC-GITA andmestikul, otse eestikeelsele andmestikule?**

U1 raames läbi viidud eksperimendid näitasid, et klassikalised masinõppemudelid (*Random Forest* ja *SVM*), mis treeniti ainult hispaaniakeelsel PC-GITA andmestikul, omasid piiratud üldistusvõimet eestikeelsele andmestikule. Kuigi PC-GITA sisesel testimisel saavutati mõõdukaid tulemusi (parim SVM AUC 0.682 koondtunnustega), langes jõudlus Eesti andmetel märgatavalt (parim SVM AUC 0.613 polünoomse tuumaga koondtunnustel, kuid sageli AUC väärtused 0.55 lähedal või alla selle). Eriti problemaatiline oli madal spetsiifilisus Eesti andmetel, mis viitas suurele hulgale valepositiivsetele ennustustele. harjutuspõhine analüüs (üksikud vokaalid, monoloog) ja spetsiifiliste harjutuste kombinatsioonide testimine U1 raames ei toonud esile meetodit, mis oleks oluliselt parandanud ristkeelset ülekantavust võrreldes kõikide ühiste tunnuste kasutamisega; monoloogi ja seda sisaldavad kombinatsioonid näitasid siiski veidi paremaid tulemusi kui üksikud vokaalid. Need leiud rõhutavad traditsiooniliste akustiliste tunnuste põhiste mudelite otsese ülekandmise väljakutseid ühest keelest teise. See on probleem, millele on tähelepanu juhitud ka varasemas kirjanduses (nt Favaro et al. (2023); Vásquez-Correa et al. (2021)), kus on leitud, et mudelite täielik üldistamine uutele, treeningandmestikus

esindamata keelele on keeruline, eriti kui sihtkeele andmestik on väike ja erineb oluliselt treeningkeele omast.

**Uurimisküsimus 2 (U2): Kas hispaania- ja eestikeelsete andmete kombineerimine treeningfaasis parandab akustilistel tunnustel põhineva mudeli jõudlust võrreldes ainult ühel keelel treenimisega?**

U2 eksperimendid, kus PC-GITA ja Eesti andmestikud liideti treeninghulgaks, näitasid märkimisväärset positiivset mõju mudelite jõudlusele. Kõikide ühiste tunnuste kasutamisel liitandmestikul saavutas SVM klassifikaator AUC väärtuse 0.66 ja RF AUC 0.666. Need tulemused on oluliselt paremad kui U1 ristkeelse testimise tulemused Eesti andmetel ning ületavad ka U1 PC-GITA siseseid tulemusi. Oluliselt paranes ka tundlikkuse ja spetsiifilisuse tasakaal. Harjutuspõhine analüüs (vokaalid, monoloog) ja harjutuste kombinatsioonide testimine liitandmestikul kinnitasid seda trendi: monoloog osutus kõige informatiivsemaks üksikuks harjutusteks (SVM AUC 0.749) ning monoloogi sisaldav harjutuste kombinatsioon (vokaalid, moduleeritud vokaalid, monoloog) andis samuti tugevaid tulemusi (SVM AUC 0.789). Need leiud viitavad selgelt, et treeningandmestiku mitmekesistamine erinevate keelte andmetega võib oluliselt parandada mudelite võimekust õppida robustsemaid ja üldistatavamaid haiguspetsiifilisi mustreid.

**Uurimisküsimus 3 (U3): Milline on tipptasemel, isejuhendatud õppel põhineva ja mujal peenhäälestatud WavLM Base mudeli jõudlus otse eestikeelsetel andmestikul?**

U3 raames testiti kahte SSL4PR raamistikust pärinevat, PC-GITA andmestikul peenhäälestatud WavLM Base mudelit<sup>1,2</sup> otse eestikeelsetel vokaali- ja monoloogilõikudel. Mõlema mudeli puhul ilmnis, et need ei suutnud eesti keeles klasse eristada, kuigi andmed olid töödeldud mudelile sobivaks sisendiks. Mudel kaldus ennustama enamusklassi (AUC ligikaudu 0.5). See negatiivne tulemus on siiski oluline, rõhutades, et isegi tipptasemel SSL-mudelite "karbist välja" rakendamine uues keele- või andmekontekstis (eriti lühikestel helilõikudel) ei pruugi anda soovitud tulemusi ilma sihtandmestikul täiendava peenhäälestamiseta või mudeli spetsiifilisema kohandamiseta.

**Peamised järeldused ja panus**

Bakalaureusetöö peamised järeldused on järgnevalt:

- Traditsioonilistel akustilistel tunnustel ja klassikalistel masinõppemudelitel põhinevate süsteemide otsene ülekantavus hispaaniakeelselt PC-GITA treeningandmestikult eestikeelsele sihtandmestikule Parkinsoni tõve tuvastamiseks on piiratud. See leid on kooskõlas varasemas kirjanduses täheldatud üldisemate väljakutsetega, mis puudutavad mudelite ristkeelset rakendatavust (Favaro et al. 2023; Vásquez-Correa et al. 2021). Nende artiklites on samuti osutatud, et täielik üldistamine uuele, treeningus nägemata keelele on keeruline, eriti kui sihtkeele andmestik on väike või erineb akustiliste või keeleliste omaduste poolest oluliselt treeningandmestikust, nagu see oli ka Eesti andmestiku näitel.

<sup>1</sup><https://huggingface.co/moreno1q/SSL4PR-wavlm-base-full>

<sup>2</sup><https://huggingface.co/moreno1q/SSL4PR-wavlm-base>

- Erinevate keelte andmestike kombineerimine treeningfaasis parandab oluliselt klassikaliste masinõppemudelite jõudlust ja üldistusvõimet, võimaldades saavutada märksa paremaid tulemusi kui ainult ühel keelel treenides. See on praktilisest vaatepunktist oluline leid, mis soosib mitmekeelsete andmekogude loomist ja kasutamist.
- Üksikute kõneülesannete (harjutuste) informatiivsus varieerub. Monoloogi tüüpi ülesanded, mis sisaldavad rikkalikumat kõneinformatsiooni, osutusid üldiselt informatiivsemaks kui lühikesed staatilised vokaalihäldused,
- Tiptasemel, mujal peenhäälestatud SSL-mudelite otsene rakendamine eesti keelele ja spetsiifilistel lühikestel helilõikudel ei pruugi anda edukaid tulemusi ilma täiendava kohandamise või sihtandmestikul peenhäälestamiseta.

Töö panus seisneb esiteks Parkinsoni tõve häälepõhise tuvastamise uurimises spetsiifiliselt eesti keele kontekstis, kasutades nii traditsioonilisi kui ka kaasaegsemaid masinõppeläheneid. Teiseks pakub töö empiirilisi tulemusi ristkeelse ülekantavuse ja andmestike kombineerimise efektiivsuse kohta, mis on olulised edasiste uuringute ja potentsiaalsete rakenduste väljatöötamisel. Kolmandaks toob töö esile väljakutsed, mis on seotud tiptasemel süvaõppemudelite rakendamisega piiratud ressurssidega keeltes ja spetsiifilistes andmekogudes.

#### **Edasised uurimisvõimalused**

Saadud tulemuste põhjal võib välja pakkuda mitmeid edasisi uurimissuundi:

- SSL-mudelite (nagu WavLM) peenhäälestamine spetsiifiliselt eestikeelse Parkinsoni tõve andmestikul, et hinnata nende tegelikku potentsiaali eesti keeles. Seejuures tuleks uurida erinevaid peenhäälestamise strateegiaid ja andmemahatude mõju.
- Täiendavate ja robustsemate akustiliste tunnuste uurimine ning nende kombineerimine masinõppemudelitega, pöörates erilist tähelepanu tunnustele, mis on vähem keelespetsiifilised.
- Helisignaali eeltöötluse täiustamine, sealhulgas erinevate müraeemaldusmeetodite katsetamine ja nende mõju hindamine Parkinsoni tõve tuvastamise täpsusele.
- Suurema ja mitmekesisema eestikeelse Parkinsoni tõve hääleandmestiku kogumine, mis võimaldaks robustsemate mudelite treenimist ja valideerimist.
- Erinevate kõneülesannete (harjutuste) ja nende kombinatsioonide süstemaatilisem uurimine, et leida optimaalseimad ja kliiniliselt kõige relevantsemad testid Parkinsoni tõve häälepõhiseks hindamiseks Eestis.

- Andmete augmenteerimise tehnikate rakendamine, et kunstlikult suurendada eesti-keelse treeningandmestiku mahtu ja varieeruvust.
- Mudelite interpreteeritavuse uurimine, et paremini mõista, millistele hääleomadustele mudelid oma otsuste tegemisel tuginevad.

Kokkuvõttes annab käesolev töö esialgse, kuid olulise panuse Parkinsoni tõve häälepõhise tuvastamise uurimisse eesti keeles, luues baasi edasisteks põhjalikumateks uuringuteks selles valdkonnas.

## Viited

- Akila, B ja J. J. V. Nayahi (juuni 2024). „Parkinson classification neural network with mass algorithm for processing speech signals“. *Neural Computing and Applications* 36.17, lk. 10165–10181. DOI: 10.1007/s00521-024-09596-z.
- Alexei Baevski, Henry Zhou ja Michael Auli Abdelrahman Mohamed (2020). „wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations“. *CoRR* abs/2006.11477. arXiv: 2006.11477. URL: <https://arxiv.org/abs/2006.11477>.
- Boualoulou, Nabil, Taoufiq Belhoussine Drissi ja Benayad Nsiri (juuni 2023). „CNN AND LSTM FOR THE CLASSIFICATION OF PARKINSON’S DISEASE BASED ON THE GTCC AND MFCC“. *Applied Computer Science* 19.2, lk. 1–24. DOI: 10.35784/acs-2023-11.
- Carrón, J. *et al.* (detsember 2021). „A mobile-assisted voice condition analysis system for Parkinson’s disease: assessment of usability conditions“. *Biomedical Engineering OnLine* 20.1, lk. 114. DOI: 10.1186/s12938-021-00951-y.
- Cortes, Corinna ja Vladimir Vapnik (jaanuar 2009). „Support-vector networks“. *Chem. Biol. Drug Des.* 297, lk. 273–297. DOI: 10.1007/%2FBF00994018.
- Cutler, Adele, David Cutler ja John Stevens (jaanuar 2011). „Random Forests“. Teoses: köide 45. Springer Science+Business Media, LLC, lk. 157–176. ISBN: 978-1-4419-9325-0. DOI: 10.1007/978-1-4419-9326-7\_5.
- Davis, S. ja P. Mermelstein (august 1980). „Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences“. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28.4, lk. 357–366. ISSN: 0096-3518. DOI: 10.1109/TASSP.1980.1163420.
- Favaro, Anna *et al.* (november 2023). „Interpretable speech features vs. DNN embeddings: What to use in the automatic assessment of Parkinson’s disease in multi-lingual scenarios“. *Computers in Biology and Medicine* 166, lk. 107559. DOI: 10.1016/j.combiomed.2023.107559.
- Feigin, Valery *et al.* (mai 2019). „Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016“. *The Lancet Neurology* 18, lk. 459–480. DOI: 10.1016/S1474-4422(18)30499-X.
- Iyer, Anu *et al.* (november 2023). „A machine learning method to process voice samples for identification of Parkinson’s disease“. *Scientific Reports* 13.1, lk. 20615. DOI: 10.1038/s41598-023-47568-w.
- Jadoul, Yannick, Bill Thompson ja Bart de Boer (2018). „Introducing Parselmouth: A Python interface to Praat“. *Journal of Phonetics* 71, lk. 1–15. ISSN: 0095-4470. DOI: <https://doi.org/10.1016/j.wocn.2018.07.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0095447017301389>.

- Jeancolas, Louise *et al.* (veebruuar 2021). „X-Vectors: New Quantitative Biomarkers for Early Parkinson’s Disease Detection From Speech“. *Frontiers in Neuroinformatics* 15, lk. 578369. DOI: 10.3389/fninf.2021.578369.
- Kadastik-Eerme, L. *et al.* (september 2018). „The increasing prevalence of Parkinson’s disease in Estonia“. *Acta Neurologica Scandinavica* 138.3, lk. 251–258. DOI: 10.1111/ane.12948.
- Kohavi, Ron (jaanuar 1995). „A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection“. *International Joint Conference on Artificial Intelligence* 14.
- Kotsiantis, Sotiris, I. Zaharakis ja P. Pintelas (november 2006). „Machine learning: A review of classification and combining techniques“. *Artificial Intelligence Review* 26, lk. 159–190. DOI: 10.1007/s10462-007-9052-3.
- La Quatra, Moreno *et al.* (september 2024). „Exploiting Foundation Models and Speech Enhancement for Parkinson’s Disease Detection from Speech in Real-World Operative Conditions“. Teoses: *Proceedings of Interspeech 2024*. ISCA, lk. 1405–1409. DOI: 10.21437/Interspeech.2024-522.
- LeCun, Yann, Y. Bengio ja Geoffrey Hinton (mai 2015). „Deep Learning“. *Nature* 521, lk. 436–44. DOI: 10.1038/nature14539.
- Logemann, Jeri *et al.* (1978). „Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients.“ *The Journal of speech and hearing disorders* 43 1, lk. 47–57. URL: <https://api.semanticscholar.org/CorpusID:25865019>.
- Mallela, Jakkula *et al.* (oktoober 2020). „Raw Speech Waveform Based Classification of Patients with ALS, Parkinson’s Disease and Healthy Controls Using CNN-BLSTM“. Teoses: *Proceedings of Interspeech 2020*. ISCA, lk. 4586–4590. DOI: 10.21437/Interspeech.2020-2221.
- Nagasubramanian, Gnaneswaran ja Muthukrishnan Sankayya (mai 2021). „Multi-Variate vocal data analysis for Detection of Parkinson disease using Deep Learning“. *Neural Computing and Applications* 33.10, lk. 4849–4864. DOI: 10.1007/s00521-020-05233-7.
- Narendra, N. P., Björn Schuller ja Paavo Alku (2021). „The Detection of Parkinson’s Disease From Speech Using Voice Source Information“. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29, lk. 1925–1936. DOI: 10.1109/TASLP.2021.3078364.
- Orozco-Arroyave, Juan Rafael ja Elmar Nöth (2014). *New Spanish speech corpus database for the analysis of people suffering from Parkinson’s disease*. Appeared as reference [20] in Vásquez-Correa *et al.* 2021 (doi: 10.1016/j.patrec.2021.04.011) and reference [23] in La Quatra *et al.* 2024 (doi: 10.21437/Interspeech.2024-522), likely LREC 2014 paper. Original publication details like conference/journal name,

- pages etc. not provided in user's list. Assumed to be the LREC 2014 paper based on citations in other provided articles.
- Pedregosa, Fabian, Alexandre Gramfort Gaël Varoquaux *et al.* (2018). *Scikit-learn: Machine Learning in Python*. arXiv: 1201.0490 [cs.LG]. URL: <https://arxiv.org/abs/1201.0490>.
- Pedregosa, Fabian, Gaël Varoquaux *et al.* (2011). „Scikit-learn: Machine Learning in Python“. *Journal of Machine Learning Research* 12, Oct, lk. 2825–2830.
- Shen, Matthew, Pouria Mortezaagha ja Arya Rahgozar (aprill 2025). „Explainable artificial intelligence to diagnose early Parkinson's disease via voice analysis“. *Scientific Reports* 15. DOI: 10.1038/s41598-025-96575-6.
- Suppa, Antonio *et al.* (veebruuar 2022). „Voice in Parkinson's Disease: A Machine Learning Study“. *Frontiers in Neurology* 13, lk. 831428. DOI: 10.3389/fneur.2022.831428.
- Teixeira, João, Carla Oliveira ja Carla Lopes (detsember 2013). „Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters“. *Procedia Technology* 9, lk. 1112–1122. DOI: 10.1016/j.protcy.2013.12.124.
- Titze, Ingo (juuni 2008). „Nonlinear source-filter coupling in phonation: Theory“. *The Journal of the Acoustical Society of America* 123, lk. 2733–49. DOI: 10.1121/1.2832337.
- Tougui, Ilias *et al.* (detsember 2024). „Transformer-based transfer learning on self-reported voice recordings for Parkinson's disease diagnosis“. *Scientific Reports* 14.1, lk. 30131. DOI: 10.1038/s41598-024-81824-x.
- Vabalas, Andrius *et al.* (november 2019). „Machine learning algorithm validation with a limited sample size“. *PLOS ONE* 14, e0224365. DOI: 10.1371/journal.pone.0224365.
- Valarmathi, P. *et al.* (märts 2025). „Enhancing parkinson disease detection through feature based deep learning with autoencoders and neural networks“. *Scientific Reports* 15.1. Year and month might indicate future publication or pre-print status, check original source if needed., lk. 8624. DOI: 10.1038/s41598-025-88293-w.
- Vásquez-Correa, Juan Camilo *et al.* (oktoober 2021). „Transfer learning helps to improve the accuracy to classify patients with different speech disorders in different languages“. *Pattern Recognition Letters* 150, lk. 272–279. DOI: 10.1016/j.patrec.2021.04.011.
- Wang, Chengyi *et al.* (oktoober 2022). „WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing“. *IEEE Journal of Selected Topics in Signal Processing* 16, lk. 1–14. DOI: 10.1109/JSTSP.2022.3188113.
- Wei-Ning Hsu, Benjamin Bolte, Kushal Lakhotia Yao-Hung Hubert Tsai ja Abdelrahman Mohamed Ruslan Salakhutdinov (2021). „HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units“. *CoRR* abs/2106.07447. arXiv: 2106.07447. URL: <https://arxiv.org/abs/2106.07447>.

Yumoto, Eiji, W Gould ja Thomas Baer (juuli 1982). „Harmonics-to-noise ratio as an index of the degree of hoarseness“. *The Journal of the Acoustical Society of America* 71, lk. 1544–9.

# Lisad

## I. Sõnastik

**AUC (*Area Under the Curve*)** Klassifitseerimismudeli jõudluse mõõdik, mis arvutatakse ROC-kõvera aluse pindalana. ROC-kõver (*Receiver Operating Characteristic curve*) ise on graafik, mis illustreerib klassifikaatori diagnostilist võimekust erinevatel otsustuslavedel. See konstrueeritakse, kandes y-teljele tundlikkuse (tõepositiivsete osakaal, TPR) ja x-teljele valepositiivsete osakaalu (FPR ehk 1 - spetsiifilisus). AUC, olles selle kõvera alune pindala (väärtus vahemikus 0 kuni 1), peegeldab mudeli üldist võimet eristada korrektseid positiivseid ja negatiivseid juhtumeid üle kõigi võimalike otsustuslavede. Väärtus 1 tähistab täiuslikku eristamist, 0.5 juhuslikku arvamist (ning väärtused alla 0.5 viitavad, et mudel toimib halvemini kui juhuslikkus).

**CNN (*Convolutional Neural Network*)** Konvolutsiooniline närvivõrk. Süvaõppe mudeli tüüp, mis on eriti efektiivne ruumiliste hierarhiate tuvastamisel andmetes, nt piltides või spektrogrammides.

**DL (*Deep Learning*)** Süvaõpe. Masinõppe haru, mis kasutab mitmekihilisi (sügavaid) tehisnärvivõrke, et õppida andmetest keerukaid mustreid ja hierarhilisi tunnuseid.

**DDK (*Diadochokinesis*)** Diadokokineesia. Kõne mootorika hindamise ülesanne, mis mõõdab võimet kiiresti ja täpselt korrata vahelduvaid silpe (nt /pa-ta-ka/), hinnates artikulatsiooni kiirust ja regulaarsust.

**Embedding** Representatsioon ehk vektorkujutis. Süvaõppe mudelite poolt õpitud madalamõõtmeline arvuline esitus sisendandmetest (nt heliklipist), mis püüab tabada selle olulisi tunnuseid.

**F0 (*Fundamental Frequency*)** Põhisagedus. Häälekurdude võnkumise sagedus hertsides (Hz) arvestuses; tajutakse kui hääle helikõrgust.

**Fine-tuning** Peenhäälestamine. Protsess, kus suurtes andmehulkades eeltreenitud masinõppemudeli parameetreid kohandatakse täiendavalt väiksema, spetsiifilise ülesande (nt PT tuvastamise) andmestiku põhjal.

**Formant** Häälekulgla akustiline resonants, mis võimendab teatud sagedusribasid häälespektris. Formantide sagedused on olulised vokaalide ja teiste kõnehäälikute identifitseerimisel.

**Foundational Model** Alusmudel. Suur masinõppemudel (sageli Transformer-arhitektuuriga), mis on eeltreenitud väga laiaulatuslikel andmetel ja mida saab kohandada (peenhäälestada) paljude erinevate spetsiifiliste ülesannete jaoks.

**HNR (*Harmonics-to-Noise Ratio*)** harmoonia-müra suhe hääles. Akustiline mõõdik, mis väljendab häälesignaali perioodilise (harmoonilise) energia ja aperioidilise (müra) energia suhet, tavaliselt detsibellides (dB). Kõrgem HNR viitab vähem kahedale või mürarikkale häälele.

**Jitter** Hääle põhisageduse ( $F_0$ ) lühiajaline varieeruvus ehk ebastabiilsus järjestikuste võnkeperioodide kestuste vahel. Mõõdetakse tavaliselt protsentuaalselt või absoluutväärtusena (nt mikrosekundites). Suurem väärtus viitab suuremale ebastabiilsusele.

**otsustusmets** Vt RF.

**Kepstraalkoefitsient** Vt MFCC.

**KG (Kontrollgrupp)** Uuringus osalejate rühm, kellel puudub uuritav seisund (antud juhul Parkinsoni tõbi) ja kelle tulemusi kasutatakse võrdluseks patsiendigrupi tulemustega.

**Klassifitseerimine** Masinõppe ülesanne, mille eesmärk on määrata andmepunktile (nt hääleklipile) eelnevalt defineeritud kategooria ehk klass (nt 'PT patsient' või 'Terve kontrollisik').

**MFCC (Mel-Frequency Cepstral Coefficients)** Mel-sageduse kepstraalkoefitsiendid. Heli spektraalse kuju tunnused, mis on arvatud Mel-sagedusskaalal (jäljendab inimese kuulmistaju) ja mida kasutatakse laialdaselt kõnetöötlikes.

**ML (Machine Learning)** Masinõppe. Tehisintellekti valdkond, mis keskendub algoritmidele, mis võimaldavad arvutisüsteemidel andmetest õppida ja oma jõudlust parandada ilma spetsiifilise programmeerimiseta.

**PC-GITA** Laialdaselt kasutatav hispaaniakeelne avalik kõneandmestik Parkinsoni tõvega patsientidelt ja tervetelt kontrollisikutelt (Orozco-Arroyave et al. 2014).

**PT (Parkinsoni tõbi)** Parkinsoni tõbi. Progresseeruv neurodegeneratiivne haigus, mida iseloomustavad peamiselt motoorsed sümptomid, sh kõne- ja häälemuutused (hüpokineetiline düsartria).

**PCA (Principal Component Analysis)** Põhikomponentide analüüs. Statistiline meetod andmestiku dimensioonide vähendamiseks, leides uued tunnused (põhikomponendid), mis on algsete tunnuste lineaarsed kombinatsioonid ja kirjeldavad maksimaalset andmete varieeruvust.

**RF (*Random Forest*)** otsustusmets. Masinõppe ansamblimeetod, mis koosneb paljudest otsustuspuudest. Ennustuse tegemiseks kasutatakse enamasti puude häälte enamust (klassifitseerimisel) või keskmist (regressioonil) (Cutler et al. 2011).

**Ristvalideerimine (*cross-validation*)** Mudeli hindamise tehnika, kus andmestik jaotatakse korduvalt treening- ja testalamhulkadeks (kordusteks), et hinnata mudeli üldistusvõimet ja vähendada juhuslikkusest tingitud tulemuste kõikumist.

**Shimmer** Häälesignaali amplituudi lühiajaline varieeruvus ehk ebastabiilsus järjestikuste võnkeperioodide vahel. Mõõdetakse tavaliselt protsentuaalselt või detsibellides (dB). Suurem väärtus viitab suuremale ebastabiilsusele.

**siirdeõpe (Transfer Learning)** Masinõppe meetod, kus ühes ülesandes (lähteülesanne) õpitud teadmisi (nt eeltreenitud mudeli parameetreid) rakendatakse teise, sarnase ülesande (sihtülesanne) lahendamiseks.

**Spektrogramm** Helisignaali visuaalne esitus, mis näitab signaali sageduskomponentide intensiivsust ajas. Tavaliselt on x-teljel aeg, y-teljel sagedus ja värvi intensiivsus tähistab amplituudi.

**SSL (Self-Supervised Learning)** Isejuhendatud õpe. Juhendamata õppe alaliik, kus mudel õpib representatsioone, luues sisendandmetest ise õpiülesandeid (nt ennustades puuduvaid osi), ilma et oleks vaja väliseid silte.

**Subject-independent split** Kõnelejast sõltumatu jaotus. Andmete jaotamine treening- ja testalamhulkadeks nii, et sama kõneleja salvestised ei satu kunagi samaaegselt mõlemasse alamhulka. See on oluline, et vältida mudeli üleõppimist kõneleja identiteedile.

**SVM (Support Vector Machine)** Tugivektormasin. Juhendatud masinõppe algoritm, mida kasutatakse peamiselt klassifitseerimiseks. See leiab optimaalse hüpertasandi, mis maksimeerib erinevate klasside andmepunktide vahelist eralduspiiri (Cortes et al. 2009).

**Süvaõpe** Vt DL.

**Tugivektormasin** Vt SVM.

**Tunnus (Feature)** Mõõdetav omadus, karakteristik või atribuut, mis on eraldatud või tuletatud andmetest (nt helisignaalist) ja mida kasutatakse sisendina masinõppe-mudelile.

**FN (False Negative)** Valenegatiiv. Juhtum, kus mudel ennustas negatiivset klassi (nt 'Terve'), kuid tegelik klass oli positiivne (nt 'Parkinsoni tõbi'). Teisisõnu, mudel jättis positiivse juhu tuvastamata.

**FP (False Positive)** Valepositiiv. Juhtum, kus mudel ennustas positiivset klassi (nt 'Parkinsoni tõbi'), kuid tegelik klass oli negatiivne (nt 'Terve'). Teisisõnu, mudel

tuvastas positiivse juhu seal, kus seda tegelikult polnud (nimetatakse ka I tüübi veaks).

**TN (True Negative)** Tõenegatiiv. Juhtum, kus mudel ennustas korrektselt negatiivset klassi (nt 'Terve').

**TP (True Positive)** Tõeposiitiv. Juhtum, kus mudel ennustas korrektselt positiivset klassi (nt 'Parkinsoni tõbi').

## II. Kasutatud akustilised tunnused

Antud töös kasutati *parselmouth* teegi abil eraldatud järgmist 37 akustilist tunnust:

**Põhitooni (F0) statistikud:** • Keskmise F0 (Hz)

- F0 standardhälve (Hz)
- F0 miinimum (Hz)
- F0 maksimum (Hz)

**sagedushälve mõõdikud:** • sagedushälve (local,

**intensiivsushälve mõõdikud:** • intensiivsushälve (local, dB)

**Harmonilisuse ja müra suhe:** • Keskmise HNR (dB)

**Formantide keskmised sagedused:** • F1 keskmise sagedus (Hz)

- F2 keskmise sagedus (Hz)
- F3 keskmise sagedus (Hz)
- F4 keskmise sagedus (Hz)

**Intensiivsuse statistikud:** • Keskmise intensiivsus (dB)

- Intensiivsuse standardhälve (dB)

**Mel-sageduse kepstraalkoefitsiendid (MFCC):** • MFCC 1 (keskmise)

- MFCC 1 (standardhälve)
- MFCC 2 (keskmise)
- MFCC 2 (standardhälve)
- MFCC 3 (keskmise)
- MFCC 3 (standardhälve)

- MFCC 4 (keskmine)
- MFCC 4 (standardhälve)
- MFCC 5 (keskmine)
- MFCC 5 (standardhälve)
- MFCC 6 (keskmine)
- MFCC 6 (standardhälve)
- MFCC 7 (keskmine)
- MFCC 7 (standardhälve)
- MFCC 8 (keskmine)
- MFCC 8 (standardhälve)
- MFCC 9 (keskmine)
- MFCC 9 (standardhälve)
- MFCC 10 (keskmine)
- MFCC 10 (standardhälve)
- MFCC 11 (keskmine)
- MFCC 11 (standardhälve)
- MFCC 12 (keskmine)
- MFCC 12 (standardhälve)

## II. Litsents

### **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, **Aleksis Krass**,  
(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose  
**Parkinsoni tõve tuvastamine eestikeelsete hääleklippide analüüsi abil kasutades masinõppe meetodeid**,  
(lõputöö pealkiri)  
mille juhendaja(d) on Joonas Järve ja Pille Taba,  
(juhendaja nimi)  
reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Aleksis Krass  
**15.05.2025**