

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Ida Tuule Jõgi 223123IAIB

**PARKINSONI DIAGNOOSIMISEKS KASUTATAVATE
ARCHIMEDESE SPIRAALIDE HULGA SUURENDAMISE
ANALÜÜS**

Bakalaureusetöö

Juhendaja: Sven Nõmm
PhD

Tallinn 2025

Autorideklaratsioon

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

Author: Ida Tuule Jõgi

01.06.2025

Annotatsioon

Parkinsoni diagnoosimiseks kasutatavate Archimedese spiraalide hulga suurendamise analüüs

Lõputöö eesmärk on uurida Parkinsoni tõve tuvastamist Archimedese spiraalide põhjal ning hinnata, kuidas erinevad andmehulkade suurendamise meetodid mõjutavad spiraalide sarnasust ja masinõppe mudelite klassifitseerimisvõimet. Töö aluseks on DraWritePD andmestik, mis sisaldab nii Parkinsoni patsientide kui ka kontrollgrupi poolt joonistatud spiraale. Andmehulkade suurendamiseks rakendati afiinseid teisendusi: roteerimine ja skaleerimine, ning andmehulkade tasakaalustamiseks kasutati Synthetic Minority Oversampling Technique algoritmi, sarnasust mõõdeti Dynamic Time Warping põhiste meetoditega. Klassifitseerimiseks kasutati kuut erinevat masinõppemudelit ning 19 tunnuse põhjal koostatud seitset erinevat tunnuste valikut.

Tulemused näitasid, et suurendamine sarnaseks muutvate teisenduste abil parandas klassifitseerimist, samas kui ebasarnaste teisenduste kasutamine või liigne moonutus võis täpsust vähendada. Parimaid tulemusi saavutati, kui suurendatud andmestikku täiendati ka esialgsete treeningandmetega, mis toetas mudelite üldistusvõimet ja täpsust.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 36 leheküljel, 6 peatükki, 5 joonist, 3 tabelit.

Abstract

Analysis of Increasing the Number of Archimedean Spirals Used to Diagnose Parkinson's

This thesis explores the detection of Parkinson's disease through a drawing task involving Archimedean spirals. The goal was to assess how data augmentation techniques affect spiral similarity and the classification performance of machine learning models. The dataset, sourced from the DraWritePD database, included spirals drawn by individuals with Parkinson's disease and healthy controls. Spiral similarity was measured using two Dynamic Time Warping based methods. Data augmentation was performed using affine transformations: rotation and scaling, and the Synthetic Minority Over-sampling Technique method was applied for class balancing. A total of 19 features were extracted and grouped into seven feature subsets for use in six machine learning models.

Similarity analysis showed that control group spirals were the most consistent, while Parkinson's spirals were more variable. Affine transformations did not increase within-class similarity and sometimes made Parkinson's spirals resemble control spirals more than each other. Classification results were best when augmentation used similarity-preserving transformations, especially with scaling. Dissimilar transformations reduced performance, particularly in sensitive models like support vector machine and logistic regression.

Synthetic Minority Over-sampling Technique improved results on the original dataset, but had mixed effects when combined with augmented data. The best outcomes were achieved when original training data was included alongside augmented samples. Overall, the findings suggest that data augmentation can improve Parkinson's detection from spiral drawings, provided transformations are selected carefully and the original data structure is preserved.

The thesis is written in Estonian and is 36 pages long, including 6 chapters, 5 figures and 3 tables.

Lühendite ja mõistete sõnastik

AB	AdaBoost klassifikaator
ACGAN	<i>Auxiliary Classifier Generative Adversarial Networks</i> , generatiivse võistlev võrgustiku mudel
ADASYN	<i>Adaptive Synthetic Sampling</i> , andmehulga tasakaalustamise meetod
BRIEF	<i>Binary Robust Independent Elementary Features</i> , bitipõhine tunnuste ekstraktor
Borderline-SMOTE	<i>Borderline Synthetic Minority Over-sampling Technique</i> , andmehulga tasakaalustamise meetod
CNN-BLSTM	<i>Convolutional Neural Networks Bidirectional Long Short Term Memory Network</i> , konvolutsiooniline närvivõrk, millel on kahepoolne mälu mehhanism
CW-SSIM	<i>Complex Wavelet Structural Similarity</i> , lainepõhine piltide sarnasuse mõõdik
DBA	<i>Dynamic Time Warping Barycentric Averaging</i> , aeGRIDade keskmistamise meetod aeGRIDade hulga suurendamiseks
DCGAN	<i>Deep Convolutional Generative Adversarial Networks</i> , sügav konvolutsiooniline generatiivne võistlev närvivõrk
DFM	<i>Deviation From Mean</i> , aeGRIDade augmenteerimismeetod, mille korral rakendatakse keskmisest tulemusest kõrvalekallet
DT	<i>Decision Tree</i> , otsustuspuu klassifikaator
DTW	<i>Dynamic Time Warping</i> , aeGRIDade augmenteerimise ja sarnasuse leidmise meetod, põhineb dünaamilisel ajas sobitamisel
EMDA	<i>Equalized Mixture Data Augmentation</i> , aeGRIDadest koosneva andmestiku suurendamine mitmestatud allikatest
EMD	<i>Empirical Mode Decomposition</i> , aeGRIDade empiiriliste režiimide dekomponeerimine
F1-skoor	<i>F1-score</i> , harmooniline keskmine masinõppe mudeli täpsusest ja saagisest, hindab mudeli üldist täpsust [1]
FAST	<i>Features from Accelerated Segment Test</i> , kiire nurgatuvastusmeetod

GAN	<i>Generative Adversarial Networks</i> , generatiivne võistlev närvivõrk
GRATIS	<i>GeneRAtIng TIme Series with Diverse and Controllable Characteristics</i> , tööriist mitmekesiste aegridade loomiseks
HOG	<i>Histogram of Oriented Gradients</i> , tunnustepõhine piltide sarnasusemõõdik, põhineb gradientide suundade histogrammil
ICA	<i>Independent Component Analysis</i> , dekompositsioonipõhine aegridade augmenteerimismeetod, tugineb sõltumatute komponentide analüüsil
KNN	<i>K-Nearest Neighbor</i> , k-lähima naabri klassifikaator
KT	kontrollgrupi klassi kuuluv Archimedese spiraal
<i>K-fold</i>	ristvalideerimise meetod, kus andmestik jagatakse <i>k</i> -alamhulka mudeli hindamiseks
LCSS	<i>Longest Common SubSequence</i> , elastne aegridade sarnasusemõõdik
LGT	<i>Local and Global Trend</i> , aegridade statistiline genereeriv augmenteerimismeetod
LR	<i>Logistic Regression</i> , logistilise regressiooni klassifikaator
LSTM	<i>Long Short-Term Memory</i> , pika- ja lühimäluvõrk, rakendatakse aegridade augmenteerimisel
MAR-mudel	<i>Mixture Autoregressive</i> mudel, autoregressiivsete protsesside segul põhinev mudel
NCD	<i>Normalized Compression Distance</i> , normaliseeritud pakkimispõhine kaugusmõõdik
ORB	<i>Oriented FAST and Rotated BRIEF</i> , kiire tunnuste ja kirjeldajate kombinatsioon piltide sarnasuse leidmiseks
PD	Parkinsoni klassi kuuluv Archimedese spiraal
PRDC	<i>Pattern Representation based on Data Compression</i> , tiheduspõhine piltide sarnasuse leidmise meetod
RF	<i>Random Forest</i> , otsustusmetsa klassifikaator
Saagis	<i>Recall, sensitivity</i> , näitab, kui suurt osa tegelikest positiivsetest juhtumitest suudab mudel õigesti tuvastada [1]
SFM	<i>Stochastic Feature Mapping</i> , stokastiline tunnuseruumi teisendus, mida kasutatakse aegridade augmenteerimisel
SIFT	<i>Scale-Invariant Feature Transform</i> , skaalast sõltumatu tunnuste teisendus, tunnustepõhine piltide sarnasuse leidmise meetod
SMOTE	<i>Synthetic Minority Over-sampling Technique</i> , andmehulga tasakaalustamise meetod

SPAWNER	<i>Suboptimal Warped Time Series Generator</i> , moonutatud aegridade genereerija
Spetsiifilisus	<i>Specificity</i> , näitab, kui suur osa tegelikult negatiivsetest juhtumitest klassifitseeriti mudeli poolt õigesti negatiivseteks [1]
SSIM	<i>Structural Similarity Index Measure</i> , struktuuripõhine sarnasuse mõõdik
S-SSIM	<i>Sliding Structural Similarity Index Measure</i> , libiseva aknaga struktuuripõhine sarnasuse mõõdik
STL	<i>Seasonal-Trend Decomposition Procedure Based on Loess</i> , aegridade augmenteerimismeetod, põhineb hooajalise trendi eraldamisel
SURF	<i>Speeded Up Robust Features</i> , kiire ja robustne tunnusekstraktor, kasutatakse piltide sarnasuse leidmiseks
SVM	<i>Support Vector Machine</i> , tugivektor-masin klassifikaator
Täpsus	<i>Precision</i> , näitab, kui suur osa mudeli poolt positiivseks ennustatud juhtumitest olid tegelikult positiivsed [1]
VAE	<i>Variational Autoencoder</i> , varieeruv automaatkodeerija
VGG-16	<i>Visual Geometry Group 16</i> , 16-kihiline konvolutsiooniline närvivõrk
VGG-19	<i>Visual Geometry Group 19</i> , 19-kihiline konvolutsiooniline närvivõrk
Õigsus	<i>Accuracy</i> , näitab, kui suur osa kõikidest ennustustest olid mudeli poolt õiged ehk kui paljud juhtumid ennustati õigesse klassi [1]

Sisukord

Jooniste loetelu	9
Tabelite loetelu	10
1 Sissejuhatus	11
1.1 Probleemipüstitus	12
2 Taust	13
2.1 Parkinsoni diagnoosimiseks kasutatavad joonistustestid	13
2.2 Varasemate uuringute tulemused	14
2.3 Andmehulkade suurendamise meetodite mõju	16
3 Kirjanduse ülevaade	17
3.1 Pildiliste andmete augmenteerimismeetodid	17
3.2 Aegridade augmenteerimismeetodid	18
3.3 Synthetic Minority Over-sampling Technique	19
3.4 Piltide sarnasus	20
3.5 Aegridade sarnasus	21
3.6 Masinõppe mudelid	22
4 Metoodika	23
4.1 Töövoog	23
4.2 DraWritePD andmehulk	24
4.3 Andmehulkade sarnasusskoori leidmismeetod	25
4.4 Kasutatud andmehulkade suurendamise meetodid	26
4.5 Spiraalidepõhine klassifitseerimine ja mudelite hindamine	28
5 Tulemused	30
5.1 Esialgsete spiraalide jooniste sarnasus	30
5.2 Afiinsete teisendustega suurendatud andmehulkade sarnasus	31
5.2.1 Roteeritud spiraalide jooniste sarnasus	31
5.2.2 Skaleeritud spiraalide jooniste sarnasus	32
5.2.3 Roteeritud ja skaleeritud spiraalide jooniste sarnasus	33
5.3 Andmehulkade tasakaalustamise mõju sarnasusele	34
5.4 Klassifikaatorite tulemused	34
5.4.1 Esialgsete spiraalide jooniste põhjal klassifitseerimine	35

5.4.2	Roteeritud spiraalide jooniste põhjal klassifitseerimine	35
5.4.3	Skaleeritud spiraalide jooniste põhjal klassifitseerimine	37
5.4.4	Roteeritud ja skaleeritud spiraalide jooniste põhjal klassifitseerimine	39
5.4.5	Tasakaalustatud andmehulkade põhjal klassifitseerimine	40
5.5	Tulemuste järelused	43
6	Kokkuvõte	46
	Kasutatud kirjandus	47
	Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaa-	
	davaks tegemiseks	56
	Lisa 2 – Spiraalide sarnasust hindavate meetodite soojuskaardid	57
	Lisa 3 – Klassifitseerimiseks kasutatud spiraalide tunnuste valemid	62
	Lisa 4 – Spiraalide klassidesisesed ja -vahelised keskmised sarnasusskoorid	
	arvutatud Dynamic Time Warpingu'ga	66
	Lisa 5 – Suurendatud andmehulkade suurused	95
	Lisa 6 – Suurendatud andmehulkade ja tunnuste valikute põhjal saadud klassi-	
	fitseerimistulemused	100

Jooniste loetelu

Joonis 1. Parkinsoni diagnoosimiseks kasutatavad joonistustestid (vasakult): Archimedese spiraal [17], Luria vahelduva seeria testid [12] ja Poppel- reuteri test [14].	14
Joonis 2. Eksperimentaalne töövoog.	24
Joonis 3. Archimedese spiraalid: Parkinsoni spiraal ja kontrollgrupi spiraal. . .	24
Joonis 4. Roteeritud Parkinsoni klassi kuuluv spiraal.	27
Joonis 5. Skaleeritud Parkinsoni klassi kuuluv spiraal.	27

Tabelite loetelu

Tabel 1.	Andmehulga jaotus klassifitseerimiseks.	28
Tabel 2.	Spiraalide klassifitseerimiseks kasutatud tunnused.	29
Tabel 3.	Spiraalide jooniseid sarnaseks ja ebasarnaseks muutvad skaleerimis- faktorid.	32

1 Sissejuhatus

Parkinsoni tõbi on Alzheimeri järel teine kõige levinum neurodegeneratiivne häire, põhjustades selle põdejatel lihaste jäikust, jäsemete värisemist ja tasakaaluhäireid [2]. 2016. aastal mõjutas Parkinson kogu maailmas 6,1 miljonit inimest. Haiguste globaalse koormuse uuringu kohaselt on Parkinsoni esinemissagedus ja levimus kiiresti kasvanud ning prognoosi järgi võib 2030. aastaks haigestunute arv kasvada umbes 10 miljonini [3]. Kuna Parkinsoni tõbi süveneb järk-järgult, on haiguse varajane avastamine oluline, sest aitab määrata patsiendile õige ravi ja parandada seeläbi tema elukvaliteeti [2].

Hetkel puudub Parkinsoni kinnitav diagnostiline test, mistõttu tugineb diagnoosimine tavaliselt ühe või mitme tüüpilise sümptomi kombinatsioonil: rütmiline värin, liigutuste bradükineesia, lihasjäikus ja kehahoiaku ebastabiilsus [4], [5]. Siiski on ka nende sümptomite põhjal diagnoosimine keeruline, sest need kliinilised sümptomid võivad kattuda teiste neurodegeneratiivsete seisunditega [6]. Diagnoosi saab lõplikult kinnitada ainult patoloogilise analüüsi abil lahkamisel, mis omakorda näitab diagnoosimise keerukust [7]. Seega on vaja leida lahendusi, mis aitaksid arstidel Parkinsoni varakult diagnoosida.

Parkinsoni tõve uuringute põhjal on jõutud arusaamisele, et käekirja saab kasutada haiguse diagnoosimisel biomarkerina [8], [9]. Selle oletuse põhjuseks on asjaolu, et Parkinsoni põdejate jaoks ei ole käekiri enam automaatne protsess ning nende käekiri sõltub visuaalsest tagasisidest [10]. Parkinsoni patsientide käekirja analüüsimiseks ja haiguse põhjustatud motoorse häire paremaks mõistmiseks on välja pakutud erinevaid käekirjajharjutusi, kuid tõenäoliselt on kõige populaarsemaks värina hindamise käekirjajharjutuseks praegu Archimedese spiraal [7]. Käesolevas töös keskendutaksegi ainult Archimedese spiraalide analüüsile, mitte teistele joonistustestidele.

Masinõppemudelite kasutamine spiraaliandmetel põhinevate tunnuste abil on üks võimalus Parkinsoni tuvastamiseks. Mudelite kasutust piirab aga treeningandmete vähesus – usaldusväärseks ennustamiseks vajavad mudelid piisavalt palju näiteid. Selle probleemi lahendamiseks kasutatakse andmehulkade suurendamist. Levinumateks meetoditeks on afiinsed teisendused, nagu pööramine ja skaleerimine, mis kuuluvad traditsiooniliste geomeetriliste andmeaugmenteerimise võtete hulka [11].

Samas võib suurendatud andmestikus esineda oht, et klassidevahelised andmed muutuvad liiga sarnaseks, mis raskendab mudelil nende eristamist. Kui kontrollgrupi ja Parkinsoni klassi spiraalid muutuvad liiga sarnaseks, võib mudel anda ekslikke ennustusi. Seetõttu on oluline hinnata, kuidas suurendamine mõjutab spiraalide omavahelist sarnasust.

Lõputöö käigus loodetakse pakkuda lahendusi masinõppe mudelite treeningandmete vähesuse probleemile, uurides afiinsete teisenduste abil joonistustestide hulga suurendamise võimalusi. Uuritakse, kuidas need teisendused mõjutavad Archimedese spiraalide jooniste sarnasust ning seeläbi ka masinõppe mudelite ennustamisoskust.

1.1 Probleemipüstitus

Bakalaureusetöö eesmärk on leida afiinsete teisenduste parameetrid, mille rakendamisel õnnestub suurendada Parkinsoni tõve diagnoosimiseks kasutatavate Archimedese spiraalide joonistustestide hulka selliselt, et kattuvus erinevate klasside vahel oleks minimaalne. Samuti leida ka parameetrid, mida rakendades muutub kattuvus erinevate klasside vahel suuremaks ehk spiraalid muutuvad sarnasemaks. Nii loodetakse treenida Parkinsoni ennustavaid masinõppe mudeleid ja vaadelda, kuidas mõjutavad erinevad andmehulga mudelite klassifitseerimisoskust. Eesmärgi saavutamiseks on autor püstitanud järgmised uurimisküsimused:

1. Kuidas mõjutab Archimedese spiraalide hulga suurendamine, kasutades afiinseid teisendusi, spiraalide jooniste omavahelist sarnasust?
2. Milliste afiinsete teisenduste parameetrid muudavad Archimedese spiraalide joonised sarnaseks ning millised muudavad joonised ebasarnaseks?
3. Milline on afiinsete teisenduste abil suurendatud andmehulkade mõju masinõppe mudelite ennustamisoskusele?

Uurimisküsimustele vastuste leidmiseks valitakse välja sarnasusskoori arvutamismeetod, mida kasutades saab leida Archimedese spiraalide joonistustestide sarnasust. Seejärel rakendatakse joonistustestidele erinevate väärtustega väljavalitud afiinseid teisendusi, nagu roteerimine ning skaleerimine. Suurendatud andmehulkade peal viiakse läbi sarnasuse valideerimine, et leida parameetrid, mille puhul suurendatud andmehulga kattuvus Parkinsoni põdejate ja kontrollgrupi vahel muutus minimaalseks ja milliste puhul maksimaalseks. Kasutades nii minimaalse kui ka maksimaalse kattuvusega andmehulki, demonstreeritakse, et nende abil treenitud samad masinõppe mudelid ennustavad erinevalt.

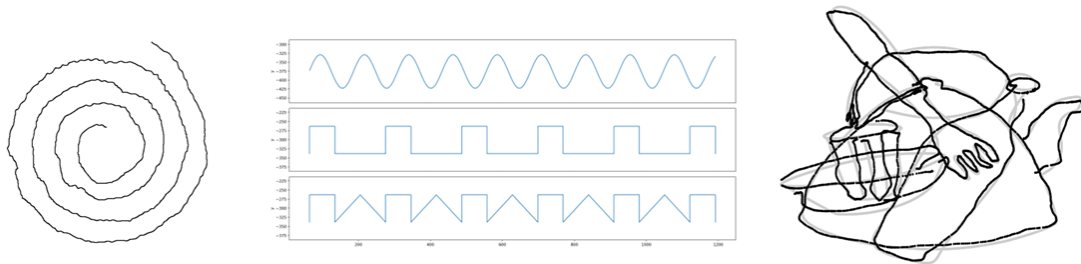
2 Taust

Töö olemusest parema ülevaate saamiseks kirjeldab autor Parkinsoni diagnoosimiseks kasutavaid joonistusteste ja nende kogumise viisi. Samuti tuuakse välja varasemalt tehtud meditsiiniliste uuringute tulemused nii Parkinsoni kui ka teiste haiguste ennustamisel. Viimases alapealkirjas on kirjeldatud, millised võivad olla andmehulkade suurendamise mõjud andmete klassifitseerimisele.

2.1 Parkinsoni diagnoosimiseks kasutatavad joonistustestid

Peenmotoorika testid, nagu joonistamis- ja kirjutamistestid, on juba pikka aega olnud kasutusel Parkinsoni tõve ning teiste neurodegeneratiivsete haiguste diagnoosimisel. Tavaliselt viidi need testid läbi, kasutades paberit ja pliiatsit, ning testide hindamine toimus spetsialistide poolt. Niiviisi toimides olid testi ainsad mõõdetavad parameetrid aeg ja vaatlajate poolt täheldatud vigade arv. Samas on pisikeste vigade märkamine inimhindajatele keeruline ning puudub ka ametlik määratlus, mida täpselt loetakse testi puhul veaks. Digitaalsete tahvlite ja hiljem tahvelarvutite kasutuselevõtt on võimaldanud registreerida ning analüüsida testide kinemaatilisi parameetreid ja pliiatsi avaldatud survet ekraanile, jättes testimisel välja inimefaktori [12].

Archimedese spiraali digitaalsed joonistustestid, mida kasutatakse ka käesolevas töös, viiakse läbi ergonoomilise puutetundliku pliiatsi ja tahvelarvuti abil. Testis osalejad peavad joonistama spiraali oma tavapärase kirjutamiskäega, püüdes võimalikult täpselt jäljendada ekraanil kuvatud eeljoonistatud kujundit, ilma kätt või käsivart toetamata [13]. Spiraali joonistamise alguspunkt, kas alustatakse seestpoolt või väljastpoolt, sõltub suuresti testi läbiviija juhistest. Sarnaselt Archimedese testile põhineb ka Poppelreuteri test etteantud kujundite jälgimisel [14]. Luria vahelduva seeria test seevastu sisaldab lisaks eeljoonistatud mustrite jäljendamisele ka vabakäelist joonistamist [15]. Kõigi kolme nimetatud testi näited on toodud joonisel 1. Lisaks nendele kasutatakse ka muid teste, nagu eraldiseisvalt kirjutatud tähed ja kirjutatud lauseid. Näiteks andmehulk HandPDMultiMC [16] koosneb seitsmest käekirjalisest testist: kursiivse l-tähe järjest kirjutamine, kolmnurkse laine joonistamine, ristkülikulise laine joonistamine, sõna "esmaspäev" ja "teisipäev" vaheldumisi kirjutamine (loetakse neljandaks ja viiendaks testiks), eesnime kirjutamine ning perekonnanime kirjutamine.



Joonis 1. Parkinsoni diagnoosimiseks kasutatavad joonistustestid (vasakult): Archimedese spiraal [17], Luria vahelduva seeria testid [12] ja Poppelreuteri test [14].

2.2 Varasemate uuringute tulemused

Drotar *et al.* [7] uurisid tuginedes enda loodud andmehulgale PaHaW käekirja kinemaatiliste ja survetunnuste põhjal Parkinsoni ennustamist. Uuringus kasutati Parkinsoni tõve prognoosimiseks kolme mudelit: k-lähim naaber (ingl. k *K-Nearest Neighbor*), AdaBoost ja tugivektor-masin (ingl. k *Support Vector Machine*). Mudelitest osutus parimaks tugivektor-masin, mille Archimedese spiraalide klassifitseerimise õigsus (ingl. k *accuracy*) oli 62.8% ja kogu andmehulga klassifitseerimise õigsus 81,3%. Eraldi hinnates osutused survetunnused Parkinsoni tõve diagnoosimisel olulisemaks kui kinemaatilised tunnused, andes õiguseks 82,5%, kinemaatiliste tunnuste kasutamisel saadi õiguseks aga 75,4%.

Valla *et al.* [17] pakkusid välja käekirjaliste andmetele tuginedes kõrgemat järku tuletistel põhinevad nurgatüüpi ja integraalilaadsed tunnused. Pakutud tunnused kirjeldavad mikro-muutusi käekirja trajektooris, mida on visuaalse vaatlusega raske või võimatu tuvastada. Siiski sisaldavad valitud tunnused väärtuslikku teavet värinataoliste sümptomite analüüsimisel. Uuringus kasutati kahte andmehulka kuuluvaid Archimedese spiraale: DraWritePD, autorite endal loodud, ja PaHaW [7]. DraWritePD andmehulga ennustamise õigsuseks saadi 84.33%, tundlikuseks (ingl. k *sensitivity*) 70.00% ja spetsiifilisuseks (ingl. k *specificity*) 93.20%. Andmehulgale PaHaW tuginedes saadi õiguseks 73.71%, tundlikuseks 75.00% ja spetsiifilisuseks 71.43%.

Lisaks statistilisele masinõppele, kus kasutatakse Parkinsoni ennustamiseks käekirja kinemaatilisi ja survepõhiseid tunnuseid, on uuringutes tuginetud ka joonistustestide pildilistele andmetele. Archimedese spiraalide pildiliste andmete põhjal ennustamiseks kasutati töös [18] Parkinson's Drawings andmehulka [19], mida suurendati spiraale kolm korda 90 kraadi võrra roteerides. Pärast piltide andmehulga suurendamist kasutati neid eeltreenitud konvolutsioonilise närvivõrgu mudeli VGG-19 (*Visual Geometry Group 19*) peal. Mudeli ennustamistulemused olid järgmised: õigsus 88.5%, tundlikkus 86.5% ja spetsiifilisus 92.2%.

Sarnaseid meetodeid Parkinsoni ennustamiseks on kasutatud ka teistes töödes [20], [21], [22]: esmalt suurendatakse joonistustestide hulka, kasutades geomeetrilisi teisendusi, värvipõhiseid moonutusi või mõlemat korraga, ning hiljem kasutatakse suurendatud andmehulka eeltreenitud tehisnärvivõrgu mudelitel, nagu Alexnet, Resnet ja Inception.

Generatiivsete võistlevate võrgustike kasutuselevõttuga 2014. aastal [23] on ka seda andmehulkade suurendamise meetodit rakendatud meditsiiniliste piltide abil haiguste ennustamiseks. Maksakahjustuste tüüpide ennustamiseks rakendasid Frid-Adar *et al.* [24] nii klassikalisi andmehulkade suurendamise meetodeid kui ka GAN-e (*Generative Adversarial Networks*): DCGAN (*Deep Convolutional Generative Adversarial Networks*) ja ACGAN (*Auxiliary Classifier Generative Adversarial Networks*). Võrreldes traditsiooniliste andmehulkade suurendamise meetoditega ennustasid GAN-id maksakahjustusi paremini, saades DCGAN-i abil tundlikuseks 85.7% ja spetsiifilisuseks 92.4%. Dzotsenidze *et al.* [25] uurisid, kuidas GAN-ide abil suurendatud Archimedese spiraalide andmehulk mõjutab spiraalide klassifitseerimist. Uuringus kasutati nelja GAN-i arhitektuuri ning võrreldi nende põhjal tehtud ennustusi traditsiooniliste andmehulkade suurendamise meetoditega. Uuringu tulemused näitasid, et GAN-iga genereeritud joonistustestide lisamine klassifitseeritavasse andmehulka parandas Parkinsoni tõve ennustamist. Mõne mudeli puhul ületas see lähenemine tavapäraste andmete suurendamise meetodite tulemusi. Siiski tuleb märkida, et ka GAN-ide abil andmehulkade suurendamiseks kasutati mõlemas töös juba traditsiooniliste meetodite abil suurendatud andmehulka.

Haiguste ennustamiseks ei ole varasemalt kasutatud ainult pildiliste andmete augmenteerimist, vaid ka aeGRIDade andmehulkade suurendamise meetodeid. Alzheimer ennustamiseks rakendasid Dao *et al.* [26] andmehulgale nii traditsioonilisi aeGRIDade suurendamise meetodeid kui ka aeGRIDade genereerimist DoppelGANger GAN-iga. AeGRIDade hulga suurendamiseks traditsiooniliselt kasutati töös Gaussi müra lisamist, skaleerimist, *time warping*'ut ja *window warping*'ut ning SPAWNER-it (*Suboptimal Warped Time Series Generator*). Parimad ennustamistulemused saadi treenides 1D konvolutsioonilist närvivõrku DoppelGANger-i abil genereeritud andmetega ning sama tulemus saadi ka genereeritud andmete ja SPAWNER-i abil augmenteeritud andmete koos rakendamisel. Parkinsoni ennustamist aeGRIDade järgi on käsitlenud Taleb *et al.* [27]. Nad kasutasid oma uuringus käekirjalisi andmeid, mis kuuluvad PDMultiMC [16] andmehulga alamhulka HandPDMultiMC. Käekirjaliste andmete esialgset aeGRIDade hulka suurendati kasutades Gaussi müra lisamist, skaleerimist ja *time warping*'ut. Traditsioonilistele aeGRIDade juurdetekitamise meetoditele lisaks rakendati uute aeGRIDade genereerimiseks ka DTW-d (*Dynamic Time Warping*). Parim õigsus, 97.62%, saavutati, kui kasutati mudeli CNN-BLSTM (*Convolutional Neural Networks Bidirectional Long Short Term Memory Network*) treenimiseks eraldi nii Gaussi müraga suurendatud andmehulki kui ka DTW-ga genereeritud andmehulki.

2.3 Andmehulkade suurendamise meetodite mõju

Masinõppe mudelid vajavad hea jõudluse tagamiseks tavaliselt suures koguses kvaliteetseid märgendatud andmeid. Andmete kogumine ja märgendamine toimub tavaliselt aga käsitsi ning nõuab palju aega ja ressursse. Kureeritud andmete kvaliteet ja esinduslikkus sõltuvad tavaliselt valdkonnas kättesaadavate puhaste andmete olemasolust ning andmete kogujate asjatundlikkuse tasemest. Kuna paljudes valdkondades ei ole sageli võimalik hankida piisavalt treeningandmeid, on praegu andmehulkade suurendamise meetodid kõige tõhusamad viisid selle probleemi leevendamiseks [11].

Andmehulkade suurendamise üheks eeliseks võib pidada väikeste andmestike klassifitseerimise õigsuse ja üldistusvõime paranemist, mis tuleneb andmete mitmekesisistamisest. Moreno-Barea *et al.* [28] uurisid VAE-de (*Variational Autoencoder*) ja GAN-ide abil väikeste andmehulkade suurendamist ning jõudsid järeldusele, et selliseid meetodeid kasutades võib masinõppe mudelite õigsus suureneda 1–3% võrra. Lisaks parandavad andmete juurdetektamise meetodid mudeli jõudlust tasakaalustamata andmestike korral. Andmehulkade suurendamise meetodite mõju uuriti töös [29]. Uuringus selgus, et suurendamise meetodite rakendamine tasakaalustamata andmetele põhjustab märkimisväärsed muutusi nii konvolutsiooniliste närvivõrkude mudelite kaaludes kui ka tugivektor-masinas kasutatavate tugivektorite ja ennustavate tunnuste valikus. Need muutused esinevad isegi siis, kui globaalsetes mõõdikutes, nagu tasakaalustatud õigsus või F1-skoor (ingl. k *F1-Score*), on näha vaid tagasihoidlikke erinevusi.

Andmehulkade suurendamisel klassifitseerimistulemuste parandamiseks võivad olla ka omad miinused. Uuringus, mis käsitles andmete suurendamise ja ansambelõppe meetodite kombinatsiooni tasakaalustamata klasside probleemide lahendamiseks, leiti, et kuigi generatiivsed mudelid, nagu GAN-id, võivad parandada klassifitseerimise jõudlust, on need sageli arvutuslikult kulukamad võrreldes traditsiooniliste meetoditega [30]. Kuigi andmete suurendamine võib küll parandada masinõppe mudeli jõudlust, võib see ka põhjustada üleõppimist, eriti kui suurendatud andmed ei esinda hästi tegelikku andmestikku [31].

3 Kirjanduse ülevaade

Peatükis antakse ülevaade kirjanduses käsitletud pildiliste andmete ja aeGRIDade augmenteerimismeetoditest ning andmehulkade suurendamise võtetest, mida kasutatakse tasakaalustamata andmehulkade tasakaalustamiseks. Lisaks käsitletakse meetodeid, mille abil on võimalik hinnata piltide või aeGRIDade sarnasust. Töös kasutatav Archimedese spiraalide andmehulk on esitatud numbriliste andmetena, sisaldades nii spiraalide jooniste koordinaate kui ka nende joonistamiseks kulunud aega. Seetõttu saab samu andmeid käsitleda nii piltide kui ka aeGRIDadena. Vastavalt sellele uuritakse augmenteerimismeetodeid mõlema andmetüübi jaoks. Kuna töös kasutatav andmehulk sisaldab rohkem kontrollgrupi andmeid kui Parkinsoni tõve patsientide andmeid, on vajalik käsitleda ka tasakaalustamise meetodeid, et vältida andmehulkade ebaühtlusest tulenevat mõju klassifikaatorite tulemustele [32].

3.1 Pildiliste andmete augmenteerimismeetodid

Üheks pildiliste andmete traditsiooniliseks augmenteerimismeetodiks on geomeetrilised teisendused. Geomeetrilised teisendused muudavad piltide geomeetrilist struktuuri, nihutades pikslid nende algsetelt asukohtadelt uutele positsioonidele, ilma et pikslite väärtusi muudetaks. Neid teisendusi kasutatakse treeningandmete mitmekesistamiseks, et paremini kirjeldada reaalse maailma välimuse muutusi, mis on põhjustatud näiteks vaatenurga erinevustest, mittejäikadest deformatsioonidest, perspektiivi ja skaalamuutustest. Geomeetrilised teisendused jagunevad omakorda afiinseteks ja mitteafiinseteks teisendusteks [11].

Afiinseks teisenduseks peetakse niisugust tasandi või ruumi teisendust, mis kujutab sirged sirgeteks ning tasandid tasanditeks. Afiinsed teisendused säilitavad paralleelsuse ja lõikumise ning pikkuste, pindalade ja ruumalade suhte [33]. Sellist tüüpi teisendused on kõige lihtsamad ja enimkasutatavad geomeetrilised teisendused pildiliste andmete juurdetekitamiseks. Afiinsete teisenduste alla kuuluvad piltide roteerimine, kallutamine, nihutamine, skaleerimine, suuruse muutmine (ilma suumimise ja kärpimiseta), peegeldamine ning ümberpööramine (ingl. k *flipping*). Suumimine ja kärpimine on levinud meetodid piltide skaleerimiseks, kuid need ei kuulu afiinsete teisenduste alla [11]. Kuna afiinsed teisendused on lihtsad ja tõhusad, kasutatakse neid sageli esmase andmetäiendamise meetodina enne keerukamate meetodite rakendamist [11].

Mitteafiinsed geomeetrilised teisendused kujutavad endast keerulisemaid geomeetrilisi teisendusi, mida kasutatakse rakenduskeskkonnale omaste geomeetriliste muudatuste jäljendamiseks. Levinud mitteafiinsed teisendused hõlmavad projektiivseid ehk perspektiivseid teisendusi, mille eesmärk on luua uusi andmeid, mis esindavad erinevaid vaatenurki. See tähendab tavaliselt algse pildi punktide teisendamist uude koordinaatsüsteemi vastavalt eelnevalt määratletud reeglitele. Projektiivsel teisendusel põhinevaid andmetäiendamise tehnikaid kasutatakse laialdaselt, et laiendada standardsetel pildikomplektidel treenitud mudeleid geomeetriliste moonutuste ja deformatsioonide käsitlemiseks. Teine oluline mitteafiinne teisendus perspektiivsete teisenduste kõrval on mittelineaarne deformatsioon, mida kasutatakse sageli andmetäienduseks. Seda saab simuleerida, lisades geomeetrilistele teisendustele rohkem vabadusastmeid, võimaldades teisenduse tugevusel varieeruda pildi erinevates piirkondades. Mittelineaarne deformatsioon on eriti kasulik välimuse muutuste kompenseerimiseks, mis on tingitud looduslikest teguritest, näiteks mittejäikade kehade deformatsioonid, või kujutise artefaktidest, nagu objektiivi moonutused [11].

Lisaks geomeetrilistele teisendustele kuuluvad ka fotomeetrilised teisendused piltide traditsiooniliste augmenteerimismeetodite alla [11]. Samuti loetakse tavaliste piltide manipuleerimise alla ka kerneli filtrite rakendamist, piltide kokkusegamist ja suvalist kustutamist. Nende augmenteerimismeetodite kõrval on olemas ka süvaõppel põhinevad andmehulkade suurendamise meetodid, nagu võistlev treenimine (ingl. k *adversarial training*), GAN-põhised meetodid ja närvistiili ülekanne (ingl. k *neural style transfer*) [34].

3.2 Aegridade augmenteerimismeetodid

Aegridade andmehulkade suurendamise meetodid jagunevad nelja põhikategooriasse: juhuslikud teisendused, mustrite segamine, generatiivsed mudelid ja dekompositsioonipõhised tehnikad. Iga kategooria jaguneb omakorda domeenipõhisteks alamtüüpideks: suurusjärgu, aja ja sageduse domeeni meetoditeks [35].

Juhuslikud teisendused on paljuski üle võetud pilditööstlusest. Teisenduste suurusjärgu domeeni alla kuuluvad näiteks müra lisamine, roteerimine, skaleerimine ja suurusjärgu moonutamine. Aja domeenis kasutatakse lõikamist, permutatsioone ja ajamoonutusi ning sageduse domeenis rakendatakse näiteks sagedusmoonutust, Fourier-teisendusel põhinevaid meetodeid ja spektrogrammi põhist täiustamist [35]. Olemas on ka hübriidlahendused, mis kombineerivad mitut domeeni korraga.

Mustrite segamisel võib samuti eristada suurusjärgu, aja ja sageduse domeeni. Suurusjärgu domeenis segatakse andmeid näiteks keskmistamise või interpoleerimise kaudu – siia kuuluvad SMOTE (Synthetic Minority Over-sampling Technique) [36], LSTM-il

(*Long Short-Term Memory*) põhinev interpoleerimine [37] ja DFM (*Deviation From Mean*) [38]. Aja domeenis kasutatakse meetodeid nagu Guided Warping [39], mis põhineb DTW-l [40]. Sageduse domeenis on näideteks EMDA (*Equalized Mixture Data Augmentation*), mis segab helisid juhuslikes ajahetkedes ja muudab sagedusribasid [41], ning SFM (*Stochastic Feature Mapping*), mis teisendab kõne ühelt esitajalt teisele [42]. Mitme domeeni kombinierimist esindavad näiteks SPAWNER [43] ja DBA (*Dynamic Time Warping Barycentric Averaging*), mis leiab mitme aegrea ühise tsentroidi iteratiivselt DTW abil [44].

Generatiivsetest mudelitest kasutatakse nii statistilisi kui ka närvivõrgu-põhiseid lähene-misi. Statistiliste hulka kuuluvad näiteks LGT (*Local and Global Trend*), mis ühendab mittelineaarsed globaalsed trendid ja lokaalsed lineaarsed suunad [45], ning GRATIS (*GeneRAting Time Series with Diverse and Controllable Characteristics*), mis simuleerib aegridasid MAR-mudelite (*Mixture Autoregressive* mudel) abil [46]. Närvivõrgupõhised meetodid hõlmavad kodeerija-dekoodri võrgustikke ja generatiivseid võistlevaid võrke ehk GAN-e [35].

Dekompositsioonipõhised meetodid keskenduvad aegrea jagamisele allkomponentideks, tuues välja eristatavad mustrid. Tuntuimad lähenemised on EMD (*Empirical Mode Decomposition*): mittelineaarsete ja mittestatsionaarsete signaalide lahutamiseks [47], ICA (*Independent Component Analysis*): sõltumatute komponentide eraldamiseks [48], ning STL (*Seasonal-Trend Decomposition Procedure Based on Loess*), mis jagab signaali hooajalis-teks, trendi- ja jääkkomponentideks [49].

3.3 Synthetic Minority Over-sampling Technique

Synthetic Minority Over-sampling Technique ehk SMOTE on meetod, mida kasutatakse klassifitseerimisel tasakaalust väljas olevatel andmehulkade tasakaalu viimiseks. Andme-hulk pole tasakaalus, kui ühte klassi kuuluvaid andmeid on rohkem kui teise klassi omasid. Andmete ebavõrdne jaotumine viib selleni, et masinõppe mudel ei suuda vähemusklassi täpselt ennustada, põhjustades seeläbi erinevaid klassifitseerimisvigu. Antud meetodi abil on võimalik andmehulka suurendada nii, et klasside kohta oleks andmeid võrdselt [32].

SMOTE pakkusid välja 2002. aastal Chawla *et al.* [36]. SMOTE erines juba olemasoleva-test andmehulga tasakaalustamise meetoditest selle poolest, et andmehulka suurendatakse, luues esialgsete andmete põhjal sünteetilisi näidiseid, selle asemel, et lihtsalt dubleerida olemasolevaid algandmeid. Meetod suurendab vähemuses olevat klassi, võttes arvesse iga sinna kuuluvat andmerida ja luues neid ühendavate tunnusjoonte põhjal sünteetilisi andmeid, mis ühendavad kõiki või osa selle k -lähimatest naabritest. Ülevalimise määra-st sõltuvalt valitakse sünteesimise lähteandmeteks juhuslikult mõned k -lähimad naabrid.

Sünteesilised näited luuakse interpoleerimise teel järgmiselt: võetakse erinevused vaadeldava tunnusvektori ja selle lähimate naabrite vahel. Erinevus korrutatakse juhusliku arvuga vahemikus 0 kuni 1 ning see liidetakse vaadeldavale tunnusvektorile. Selline teguviis põhjustab juhusliku punkti valimise kahe konkreetse tunnuse vahel ja muudab vähemusklassi otsustuspiirkonda üldisemaks, aidates vähendada andmete põhjal üleõppimist ning parandades klassifitseerimise täpsust [36]. Aastate jooksul on erialakirjanduses välja pakutud üle 85 SMOTE-i laienduse, et muuta meetodit klassifitseerimise jaoks veelgi täpsemaks. Nendest laiendustest ühed tuntuimad ja enimkasutatavad on Borderline-SMOTE (*Borderline-Synthetic Minority Over-sampling Technique*) [50] ning ADASYN (*Adaptive Synthetic Sampling*) [51], [52].

3.4 Piltide sarnasus

Piltide sarnasuse hindamise meetodid jaotuvad üldiselt kolmeks: normipõhised, tunnuspõhised ja tihenduspõhised lähenemised [53].

Normipõhised meetodid, nagu eukleidiline kaugus, keskmine ruutviga ja tippsignaali-müra suhe, hindavad piltide sarnasust punkt-punktilt [54], kuid ei suuda tabada väikseid muutusi ega vasta hästi inimese visuaalsele tajule [55], [56]. Selliste piirangute ületamiseks on välja töötatud SSIM (*Structural Similarity Index Measure*), mis võtab arvesse struktuurilist infot ja inimese nägemise omadusi [57]. SSIM-i variatsioonid, nagu S-SSIM (*Sliding Structural Similarity Index Measure*) [57] ja CW-SSIM (*Complex Wavelet Structural Similarity*) [58], arvestavad vastavalt ruumi- ja sagedusdomeeni mustreid, kuid SSIM põhineb siiski punktipõhisel korrelatsioonil, mistõttu ei sobi hästi tekstuuride võrdlemiseks [54]. Normipõhiste meetodite hulka kuulub ka OpenCV funktsioon `cv.matchShapes()`, mis hindab kujundite sarnasust Hu-momentide alusel [59].

Tunnuspõhised meetodid keskenduvad huvipunktide ja nende deskriptorite leidmisele. Näiteks SIFT (*Scale-Invariant Feature Transform*) [60], [61], SURF (*Speeded Up Robust Features*) [62], ORB (*Oriented FAST and Rotated BRIEF*) [63], HOG (*Histogram of Oriented Gradients*) [64] ja nende variandid võimaldavad piltidelt leida olulisi struktuure. ORB on SIFT-ist kiirem ja mürakindlam, kuna ühendab FAST-nurkdetektori (*Features from Accelerated Segment Test*) [65] ja BRIEF- kirjeldaja (*Binary Robust Independent Elementary Features*) [66], võimaldades tõhusat objekti tuvastust reaajas.

Tihenduspõhised meetodid põhinevad ideel, et kui ühe pildi info aitab teist efektiivselt tihendada, on pildid sarnased. Levinumad on NCD (*Normalized Compression Distance*) [67] ja PRDC (*Pattern Representation based on Data Compression*) [68], kuid nende puhul jääb sageli ebaselgeks, milline osa tihendusalgoritmist sarnasust kõige enam mõjutab [53].

Lisaks klassikalistele meetoditele kasutatakse üha enam piltide sarnasuse leidmiseks tehishärvivõrke. Sarnasuse leidmiseks on kasutatud nii konvulsioonilisi härvivõrke kui ka Siamese härvivõrke. Näiteks [53] autorid kasutasid konvulsioonilist härvivõrku piltide sarnasuse deskriptori loomiseks, deskriptorit kasutati kahest pildist noorema näo tuvastamiseks, pildi klassifitseerimisel, näo kontrollil ja pildi kvaliteedi hindamisel. Siamese härvivõrke on kasutatud ühekordse pildi klassifitseerimise parandamiseks [69]. Samuti on kasutatud ka ülekantavat õpet, mis põhineb õppija jõudluse parandamisel ühes domeenis, kandes üle teavet seotud domeenist, ja koosinussarnasust, et tehishärvivõrkude abil leida piltide sarnasust [70].

3.5 Aegridade sarnasus

Aegridade sarnasuse või ebasarnasuse mõõtmine on oluline paljudes andmekaevandamise, otsingu, klasterdamise ja klassifitseerimise ülesannetes. Masinõppes on üheks keskseks probleemiks kahe aegrea vahelise sarnasuse hindamine ehk kuivõrd üks aegrida teisele vastab. Sarnasuse mõõtmise meetodid jagunevad nelja kategooriasse: sünkroonsed, tunnustepõhised, mudelipõhised ja elastsed mõõdikud [71].

Sünkroonsed mõõdikud võrdlevad kahe aegrea vastavaid punkte (i -nda punkti i -ndaga), mistõttu sobivad need ainult sama pikkusega aegridadele. Fikseeritud vastavuse tõttu on need meetodid tundlikud nii müra kui ka ajalise nihke suhtes ning ei suuda tuvastada sarnaseid, kuid ajas nihkunud segmente. Sünkroonsete mõõdikute hulka kuuluvad näiteks eukleidiline ja Manhattani kaugus ning teised L_p normidel põhinevad mõõdikud [72].

Sünkroonsetest mõõdikutest sammu võrra keerukamaks mõõdikuteks aegridade sarnasuse leidmisel võib pidada tunnusepõhiseid sarnasuse meetodeid. Erinevalt sünkroonsetest mõõdikutest, mis arvutavad aegridade sarnasust toorandmete põhjal, kasutavad tunnuspõhised meetodid aegridade algandmetest eraldatavaid omadusi ja arvutavad nende põhjal sarnasust. Üheks tunnusepõhiseks mõõdikuks on eukleidilise kauguse arvutamine, kasutades Fourier'i kordajaid: esialgsetele aegridade arvutatakse vastavad Fourier'i kordajad ning seejärel käsitletakse neid kordajaid justkui võrreldatavaid aegridu, mille järgi leitakse eukleidiline kaugus [71].

Sarnaselt tunnuspõhistele mõõdikutele kasutavad mudelipõhised mõõdikud samuti aegridade algandmetest eraldatavaid omadusi, kuid ei arvuta koheselt nende põhjal aegridade sarnasust, vaid kasutavad kahe aegrea omadusi, et koostada nende põhjal mudel. Saadud mudeli parameetreid kasutatakse aegridade sarnasuse arvutamiseks [71].

Elastsed mõõdikud võimaldavad mõõta sarnasust aegridade vahel, mis on asünkroonsed. Mõõdikute seas on olemas meetodid, mis võimaldavad võrrelda punkte üks-mitmele, näiteks DTW [40], ning ka meetodid, mis suudavad võrrelda punkte üks-mitmele või üks-mitteühelegi, näiteks LCSS (*Longest Common SubSequence*) [72]. DTW on üks klassikalisemaid elastseid mõõdikuid [71]. Meetodi tööpõhimõte seisneb aegridade joondamises ajas optimaalselt nii, et selle joondamise kogukulu oleks minimaalne. Optimaalseks joondamiseks arvutatakse kahele aegreale tuginedes kaugusmaatriks, kaugusmaatriksi põhjal omakorda kulumaatriks ja kulumaatriksil viiakse läbi tagasijälgimine [73], [74]. DTW on peamine võrdlusmõõdik, millega uusi sarnasuse mõõdikuid võrreldakse [75], [72]. Väga vähesed meetodid on suutelised DTW-d süstemaatiliselt ületama mitmesuguste andmeallikate puhul. Need vähesed meetodid on tavaliselt keerukamad kui DTW ja nõuavad sageli ühe või enama parameetri põhjalikku häälestamist [71].

3.6 Masinõppe mudelid

Masinõppe on arvutiteaduse ja tehisintellekti haru, mis rakendab statistilisi meetodeid, et võimaldada arvutitel teadmisi omandada ja oma sooritust järk-järgult parandada ilma otsese programmeerimiseta. Keskmes on idee, et arvutid on võimelised õppima andmetest, tuvastama mustreid ning tegema järeldusi minimaalse inimsekkumisega [76], [77], [78]. Masinõppe mudeleid treenitakse sildistatud, sildistamata või hübriidsete andmeti-ke abil ning konkreetse ülesande jaoks sobiv andmetüüp määrab ära kasutatava mudeli tüübi. Andmetüüpide põhjal eristatakse nelja peamist masinõppe kategooriat: juhendatud õpe, juhendamata õpe, vähese juhendamisega õpe (ingl. *semi-supervised learning*) ning stiimulõpe [79]. Antud töös kasutatakse ainult juhendatud õpet.

Juhendatud õpe on masinõppe meetod, mis kasutab sildistatud andmeid esmalt mudeli treenimiseks ning seejärel tulemuste ennustamiseks uutel andmetel. See jälgendab inimõpet juhendaja toel, kus konkreetsete näidete põhjal tuletatakse üldised reeglid [79]. Juhendatud õpe jaguneb peamiselt kaheks: regressiooniks ja klassifitseerimiseks [78], [80], [81]. Kui mudel peab ennustama pidevaid arväärtusi, kasutatakse regressioonimeetodeid, kui väljundid kuuluvad piiratud klasside hulka, rakendatakse klassifitseerimistehnikaid [82].

Klassifitseerimise valdkonnas eristatakse nelja põhikategooriat: binaarne, mitmeklassiline, mitmesildiline ja tasakaalustamata klassifitseerimine. Sildistatud andmeid saab kasutada ka tehishärvivõrkudes: masinõppe algoritmide kogumis, mille ülesehitus ja toimimispõhimõtted on inspireeritud bioloogilistest härvivõrkudest ning mis kasutavad nii juhendatud kui ka juhendamata õppimist. Tuntuimad klassikalised klassifikaatorid on Bayesi mudelid, otsustuspuud (ingl. k *Decision Tree*), ansamblimeetodid, logistiline regressioon (ingl. k *Logistic Regression*), k-lähim naaber ning tugivektor-masin [79].

4 Metoodika

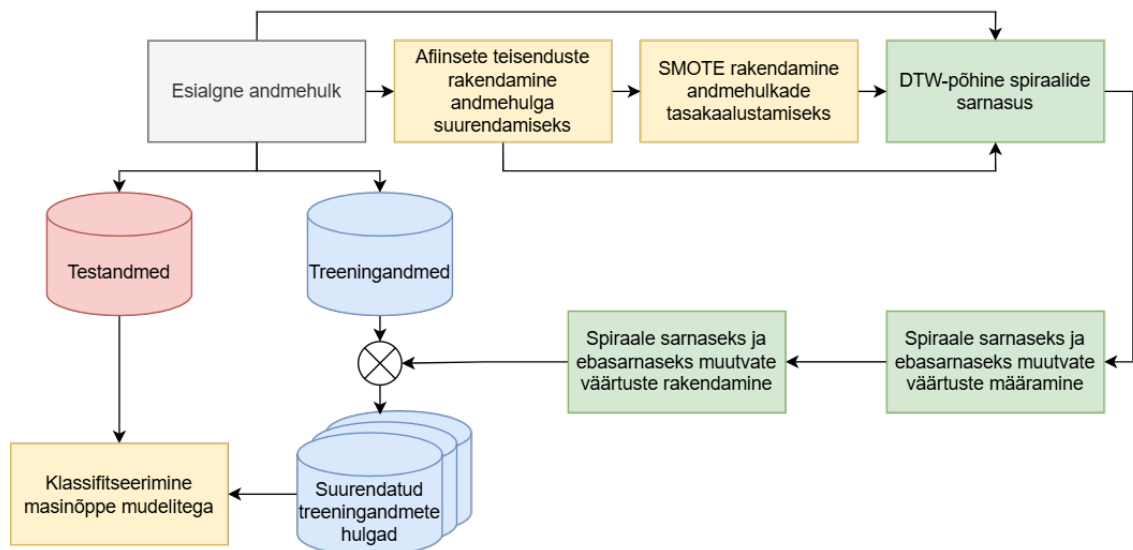
Peatükis kirjeldatakse lõputöös rakendatud metoodikat, mille eesmärk on uurida Archimede spiraalide andmehulga suurendamise mõju Parkinsoni tõve tuvastamisel. Alustuseks antakse ülevaade töö üldisest töövoost, mis hõlmab andmete töötlemise, analüüsi ja klassifitseerimise etappe. Seejärel tutvustatakse kasutatud andmestikku ja selle kogumise põhimõtteid. Edasi käsitletakse spiraalide jooniste sarnasuse hindamise meetodit ning andmehulkade suurendamiseks rakendatud afinseid teisendusi ja tasakaalustamistehnikat. Viimasena kirjeldatakse klassifitseerimiseks kasutatud masinõppemudeleid, treeningandmeid ning spiraalide jooniste põhjal arvutatud tunnuseid.

4.1 Töövoog

Lõputöö üldine töövoog oli järgmine: esmalt tutvuti kirjandusega, et saada teemast ülevaade. Seejärel katsetati erinevaid spiraalide joonistustestide sarnasuse hindamise meetodeid nii pildipõhiselt kui ka aegridade alusel, et leida sobivaim lähenemine. Valitud meetodit kasutati andmehulkade suurendamise mõju hindamiseks, rakendades afinseid teisendusi ja andmestiku tasakaalustamist. Arvutati klassisisese ja -vahelise sarnasuse muutused, tuvastades, millised teisendused muutsid spiraalide joonised sarnasemaks või ebasarnasemaks. Nende tulemuste põhjal koostati uued suurendatud andmehulgad, tuginedes kas ainult spiraalide jooniseid sarnaseks või ebasarnaseks muutvatele väärtustele. Lõpuks arvutati vastavad tunnused ning viidi läbi klassifitseerimine valitud masinõppe mudelitega. Saadud spiraalide jooniste sarnasus- ja klassifitseerimistulemuste põhjal tehti järeldused. Töövoo eksperimentaalset osa kujutab joonis 2.

Projekti halduseks kasutati Tallinna Tehnikaülikooli hallatavat GitLabi¹ platvormi. Kogu lõputöö protsess jaotati viieks tähtpunktiks (ingl. k *milestone*), millel olid omad eesmärgid ja vastavad piletid projekti halduskeskkonnas. Viis suuremat tähtpunkti olid järgmised: kirjandusega tutvumine, sobiva spiraalide jooniste sarnasusskoori leidmismeetodi otsimine, afinsete teisenduste rakendamine esialgsele andmehulgale ja suurendatud andmehulkade sarnasuse arvutamine, suurendatud andmehulkade klassifitseerimine ning lõputöö kirjaliku osa koostamine. GitLabi projektis on samuti dokumenteeritud kõik koosolekud juhendajaga. Lisaks sisaldab projekt ka kõiki lõputöö jooksul saadud tulemusi: nii spiraalide jooniste vahelisi sarnasusskoore kui ka masinõppe mudelite klassifitseerimistulemusi.

¹<https://gitlab.cs.taltech.ee/idjogi/iaib>

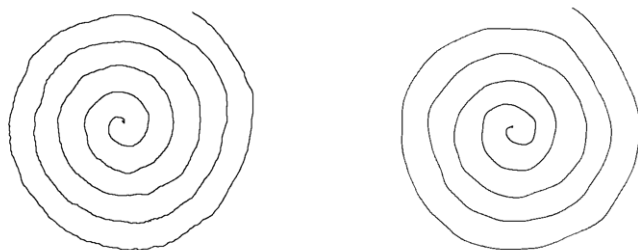


Joonis 2. Eksperimentaalne töövoog.

4.2 DraWritePD andmehulk

Töös kasutatud Archimedese spiraalide joonistustestid pärinevad DraWritePD andmehulgast, mis koguti Apple iPad Pro (2016) ja Apple Pencil'i abil. Tahvelarvuti (10,5-tolline, 240 Hz proovivõtusagedus) salvestas spetsiaalse iOS-i rakenduse kaudu spiraalide joonistamisel dünaamilisi tunnuseid: x - ja y -koordinaat (mm), ajatempel (s), surve ($[0, \dots, 6.0]$), kõrgus- ja asimuutnurk (rad). Andmestiku loomisel osales 24 Parkinsoni tõvega patsienti (keskmine vanus $74,1 \pm 6,7$) ja 34 vastendatud kontrollisikut ($74,1 \pm 9,1$). Kokku viidi läbi 12 erinevat joonistamis- ja kirjutamistesti [17].

Andmehulgast kasutatakse kokku lõputöös 48 spiraali, millest 19 kuulub Parkinsoni klassi ja 29 spiraali kontrollgrupi omasse. Joonis 3 kujutab kahte andmebaasi kuuluvat Archimedese spiraali.



Joonis 3. Archimedese spiraalid: Parkinsoni spiraal ja kontrollgrupi spiraal.

4.3 Andmehulkade sarnasusskoori leidmismeetod

Archimedese spiraalide jooniste sarnasuse leidmisel lähtuti algselt piltide põhjal sarnasuse arvutamisest. Kirjanduse põhjal eristati kolme peamist pildipõhist sarnasuse hindamise kategooriat: normipõhised, tunnuspõhised ja tihendus põhised meetodid [53]. Normipõhiseid meetodeid ei katsetatud, kuna need ei suuda tabada väikseid visuaalseid erinevusi ega vasta hästi inimese visuaalsele tajule [56]. Erandina kasutati siiski OpenCV funktsiooni `cv.matchShapes()`, mis arvutab kujundite sarnasust Hu-momentide alusel [59]. Tihendus põhised meetodid jäeti kõrvale, kuna nende toimimist on keeruline interpreteerida ning need pole visuaalsete andmete puhul niivõrd laialdaselt levinud [53].

Sarnasuse leidmiseks katsetati seetõttu tunnuspõhiseid meetodeid: SIFT [61], HOG [64] ja ORB [63], lisaks kasutati eelmainitud `cv.matchShapes()` meetodit. Tulemusi visualiseeriti normaliseeritud soojuskaartide abil, kus suurema sarnasusskooriga (ühele lähemal) spiraalide joonised kajastusid punasemas toonis. Tunnuspõhiste meetodite soojuskaardid on lisas 2. Parimaks osutus SIFT, mis tuvastas sarnasuse Parkinsoni ja kontrollgrupi spiraalide jooniste siseselt. HOG suutis määrata, et spiraali joonis on kõige sarnasem iseendale, kuid ei tuvastanud laiemat sarnasust. ORB leidis kõik spiraalide joonised omavahel üsna sarnas- teks, ent suutis samuti määrata spiraali iseendale kõige sarnasemaks. `cv.matchShapes()` andis eksitavaid tulemusi, liigitades kaks Parkinsoni spiraali joonist täiesti samasuguseks, kuigi need polnud identsed.

Tunnuspõhised meetodid, mis on mõeldud objektiderikastele piltidele, ei sobinud hästi musta spiraali analüüsimiseks valgel taustal. Seetõttu kasutati alternatiivina ImageNet-andmestikul eeltreenitud tehisnärvivõrke: DenseNet121, DenseNet169, DenseNet201, EfficientNet, Inception V3, NasNet, ResNet50, ResNet101, ResNet152, VGG-16 (*Visual Geometry Group 16*) ja Xception, mille puhul arvutati piltide *embedding*’ute koosinussarnasust [70], [83]. Selle põhjal hinnati piltide vahelist sarnasust ilma otsese objektituvastuseta. Kõikide proovitud eeltreenitud tehisnärvivõrkude spiraalide jooniste sarnasust illustreerivad soojuskaardid on leitavad lõputöö GitLabi projektist ².

Eeltreenitud mudelid ei andnud oodatud tulemusi: enamik neist ei suutnud spiraalide jooniseid eristada või andsid kõigile sarnase skoori. Seetõttu treeniti VGG-16, DenseNet121, ResNet50 ja ResNet101 mudelite viimaseid kihte edasi ka DraWritePD spiraalidega. Parima tulemuse andis VGG-16, mis tuvastas klassisisese sarnasuse, kuid spiraalide vähese arvu tõttu esines mudelitel üleõppimist: sarnasusskoorid muutusid liiga ühtlasteks. Mudeli peenhäälestamine: õppimiskiirus, epohhid, kihtide arv, parandas tulemusi veidi, ent ei võimaldanud spiraalide jooniseid piisavalt täpselt eristada.

²<https://gitlab.cs.taltech.ee/idjogi/iaib>

Kuna nii tunnuspõhised meetodid kui ka eeltreenitud närvivõrgud ei võimaldanud sarnasust usaldusväärselt hinnata, otsustati kasutada Archimedese spiraalide jooniste esialgseid koordinaate ja käsitleda neid aegridadena. Sellisel juhul sobib sarnasuse mõõtmiseks Dynamic Time Warping, mis sobib erineva pikkusega aegrealiste andmete võrdlemiseks [40]. DTW arvutati spiraalide jooniste x - ja y -koordinaatide põhjal, jättes välja surveandmed, kuna andmehulkade suurendamine rakendati vaid positsioonilistele andmetele.

Soojuskaardi, joonis 10, põhjal tuvastati, et kontrollgrupi spiraalide joonised olid omavahel sarnasemad kui Parkinsoni spiraalide joonised. Samuti oli KT–PD (Kontrollgrupi klass– Parkinsoni klass) klassidevaheline sarnasus madalam. Kuna DTW suutis tuvastada kontrollgrupi spiraalide jooniste omavahelise sarnasuse, Parkinsoni spiraalide jooniste omavahelise sarnasuse ning kontrollgrupi ja Parkinsoni spiraalide jooniste sarnasuse erinevuse esialgsete andmete peal, otsustati töös edaspidi kasutada DTW-d spiraalide jooniste sarnasuse arvutamiseks.

Sarnasus arvutati Pythonis funktsiooniga `fastdtw()` [84], mis töötab kiiremini kui klassikaline DTW. Kasutati kahte lähenemist: 1D DTW (x - ja y -koordinaadid ühendatud) koos absoluutkaugusega ning 2D DTW koos eukleidilise kaugusega. Spiraalide jooniste selgemaks sarnasuse arvutamiseks kasutati Parkinsoni spiraalide joonistest, mis olid omavahel ebasarnased, ning kontrollgrupi spiraalide joonistest, mis olid omavahel sarnased. Seega koosnes lõplik sarnasuse arvutamiseks kasutatav valim 15 Parkinsoni (PD1–PD15) ja 17 kontrollgrupi spiraali joonisest (KT13–KT29). Sarnasus kirjeldati kolme näitaja kaudu: keskmine, mediaan ja standardhälve.

4.4 Kasutatud andmehulkade suurendamise meetodid

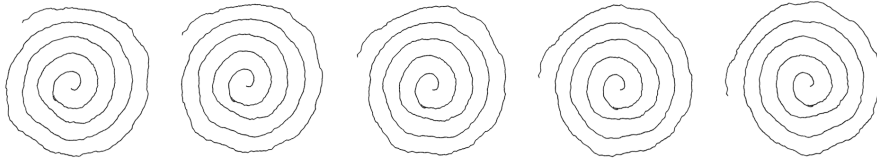
Töös kasutati Archimedese spiraalide andmehulga suurendamiseks afinseid teisendusi, et uurida, kuidas need mõjutavad masinõppe mudelite klassifitseerimisoskust. Afinsetest teisendustest rakendati roteerimist ja skaleerimist, kuna need on lihtsalt rakendatavad, laialdaselt kasutusel ning sobivad nii aegridade kui ka pildiliste andmete töötlemiseks. See võimaldab eeldada, et aegridadele rakendatud teisendused annavad sarnase mõju ka spiraalide jooniste visuaalsel tasandil. Joonis 4 kujutab ühe Parkinsoni klassi kuuluva spiraali roteerimist ja joonis 5 kujutab sama spiraali skaleerimist.

Kõik spiraalid roteeriti, kasutades nende x - ja y -koordinaate. Roteerimine viidi läbi 15-kraadise sammuga vahemikus 15 kuni 360 kraadi, et hinnata roteerimisnurga mõju sarnasusele võimalikult täpselt. Kuna koordinaate käsitleti maatriksina, kasutati roteerimiseks maatriksarvutust. Kasutatud roteerimismaatriks erineb klassikalisest 2D-mudelist, kuna arvestab pildi koordinaattelgede orientatsiooni.

Maatriksarvutus põhines järgmisel teisendusel:

$$\theta_{\text{rad}} = \frac{\pi}{180} \cdot \theta \quad (4.1)$$

$$\begin{bmatrix} x' & y' \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \cdot \begin{bmatrix} \cos(\theta_{\text{rad}}) & \sin(\theta_{\text{rad}}) \\ \sin(\theta_{\text{rad}}) & -\cos(\theta_{\text{rad}}) \end{bmatrix} \quad (4.2)$$

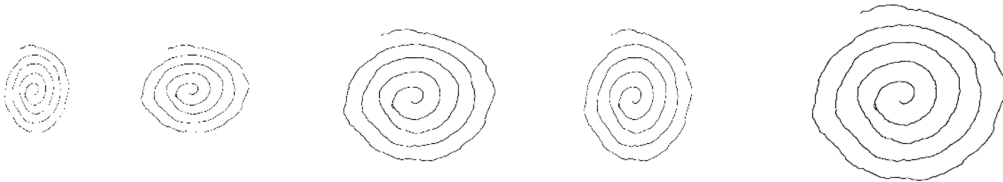


Joonis 4. Roteeritud Parkinsoni klassi kuuluv spiraal.

Spiraalide skaleerimine toimus sarnaselt: kasutati x - ja y -koordinaate ning maatriksarvutust. Skaleerimisfaktorite vahemik oli 0.05 kuni 3.00, sammuga 0.05. Rakendati kahte erinevat lähenemist: esimesel juhul alustati x -telje skaleerimist väiksema väärtusega, teisel juhul y -telje skaleerimist. Skaleerimine arvutati meetodiga, kus x on skaleerimisfaktor x -teljel ja y on skaleerimisfaktor y -teljel, järgmiselt:

$$\mathbf{X} = \begin{bmatrix} x & y \end{bmatrix} \quad (4.3)$$

$$\mathbf{S} = \mathbf{X} \cdot \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \quad (4.4)$$



Joonis 5. Skaleeritud Parkinsoni klassi kuuluv spiraal.

Lisaks roteerimise ja skaleerimise eraldi rakendamisele testiti ka nende koos rakendamist. Selleks roteeriti spiraalid esmalt valemi (4.2) abil ja seejärel skaleeriti roteeritud punktid vastavalt valemile (4.4), mille tulemusel saadi teisendatud spiraal.

Kuna esialgne andmestik oli tasakaalustamata: Parkinsoni klassis oli 19 ja kontrollgrupis 29 spiraali, kasutati andmehulkade tasakaalustamiseks SMOTE meetodit, mille puhul loodi uusi näiteid vähemuses olevasse klassi. SMOTE meetodi rakendamisel katsetati erinevaid k -naabrite väärtusi: 3, 5 ja 10, et mudeli treenimisel tagada võimalikult kõrged tulemused.

4.5 Spiraalidepõhine klassifitseerimine ja mudelite hindamine

Erinevalt spiraalide jooniste vahelise sarnasuse hindamisest, kus kasutati valitud alamhulka, kaasati klassifitseerimisesse kõik olemasolevad Archimedese spiraalid. Kuna enamik spiraale olid joonistatud seestpoolt väljapoole, alustades keskpunkti, pöörati kahe kontrollgrupi spiraali (KT6 ja KT7) x -, y -koordinaadid ning surveandmed ümber, et tagada ühtne alguspunkt kõigi spiraalide puhul. See välistas olukorra, kus joonistamise suund mõjutaks tunnuste arvutamist ja seeläbi klassifitseerimise täpsust.

Klassifitseerimiseks jagati andmestik juhuslikult treening- ja testandmeteks, järgides põhimõtet, et 70% andmetest kuulub treening- ja 30% testandmetesse. Kuna andmestik oli tasakaalustamata: Parkinsoni klassi kuulus vähem spiraale kui kontrollgruppi, valiti testandmetesse võrdses mahus mõlema klassi näiteid. Kontrollgrupist kaasati ülejäänud andmed treeningandmetesse. Lõplik jaotus on toodud tabelis 1.

Tabel 1. Andmehulga jaotus klassifitseerimiseks.

Andmed	Parkinsoni klass	Kontrollgrupp
Testandmed	6	6
Treeningandmed	13	23

Klassifitseerimiseks kasutati kuut masinõppemudelit: AdaBoost (AB), otsustuspuu (DT), k -lähima naabri meetod (KNN), logistiline regressioon (LR), otsustusmets (RF, ingl. *Random Forest*) ja tugivektor-masin (SVM). Mudelid valiti Valla *et al.* töö [17] põhjal, kus uuriti Parkinsoni tuvastamist spiraalide käekirjaliste tunnuste alusel. Sama allika põhjal arvutati ka klassifitseerimiseks vajalikud spiraalide tunnused, mida oli kokku 19. Kasutatud tunnuste valemid on leitud lisas 3. 19 tunnuse seas oli kaheksa surveandmetel põhinevat tunnust, mida afiinsete teisenduste rakendamine ei mõjuta. Need on: *Pressure_median*, *Shake_mass*, *Shake_max*, *Shake_mean*, *Shake_median*, *Pressure_diff_min*, *Pressure_diff_max*, *Snatch_mean*. Lisaks ka tunnus *Duration*, mis kirjeldab joonistamise kestust.

Tunnused jaotati seitsmesse alamhulka, mis koostati Valla *et al.* töös saadud parimate klassifitseerimistulemuste põhjal. Alamhulgad on esitatud tabelis 2.

Tabel 2. Spiraalide klassifitseerimiseks kasutatud tunnused.

Valiku nr	Kasutatud tunnused
1	ϕ_mass , <i>Duration</i> , <i>Pressure_median</i>
2	<i>Pressure_median</i> , <i>Velocity_median</i> , α_accel_max
3	<i>Shake_mass</i> , <i>Shake_max</i> , <i>Snap_mass</i> , <i>Crackle_mass</i> , <i>Pop_mass</i>
4	<i>Shake_mean</i> , <i>Accel_x_mass</i> , α_accel_max , <i>Shake_median</i> , α_accel_min
5	<i>Pressure_diff_min</i> , <i>Shake_mean</i> , <i>Shake_median</i>
6	<i>Accel_x_min</i> , α_accel_min , <i>Shake_mean</i> , <i>Pressure_diff_max</i> , <i>Shake_max</i>
7	$\alpha_velocity_max$, α_accel_min , <i>Snatch_mean</i>

Treeningandmete hulga suurendamiseks rakendati roteerimist, skaleerimist ning nende kombinatsiooni. Lisaks kasutati andmehulkade tasakaalustamiseks SMOTE meetodit. Roteerimisel rakendati kõiki pöördnurki ning lisaks kahte varianti: ainult sarnaseks muutvad nurgad ja ainult ebasarnaseks muutvad nurgad (lähtuvalt peatükis 4.3.1 toodud sarnasustulemustest). Sama põhimõtet järgiti ka skaleerimisel: suurendamine toimus kas kõigi skaleerimisfaktoritega, ainult sarnaseks muutvate või ainult ebasarnaseks muutvate väärtustega. Roteerimise ja skaleerimise kombinatsioonis kasutati samuti ainult sarnaseid või ebasarnaseid teisendusi. Lisaks testiti andmehulkade klassifitseerimist ka sellisel kujul, kus suurendatud andmestikele lisati juurde esialgne treeningandmete hulk. Eesmärk oli uurida, kas suurendatud andmestik koos algsete näidetega parandab mudeli üldistusvõimet võrreldes olukorraga, kus kasutatakse ainult teisendatud andmeid.

Suurendatud andmestike põhjal arvutati vastavad tunnused ning treeniti valitud klassifikaatorid. Mudelite ennustamisvõimet hinnati testandmetel, kasutades õigsust, täpsust (ing. *precision*), saagist (ing. *recall*), spetsiifilisust ja F1-skoori.

Klassifikaatorite ületreenimise vältimiseks kasutati töös pesastatud ristvalideerimist (ingl. *k nested cross-validation*). Ristvalideerimine on levinud meetod mudeli üldistusvõime hindamiseks ning hüperparameetrite optimeerimiseks, kuid tavalise ristvalideerimise puhul võivad samad andmed sattuda nii mudeli häälestamisse kui ka hindamisse, põhjustades kallutatust. Pesastatud ristvalideerimine lahendab selle, jagades protsessi kaheks: sisemine ristvalideerimine optimeerib hüperparameetrid, välimine hindab mudelit seni nägemata andmetel. See tagab objektiivsema pildi mudeli toimivusest [85]. Töös kasutati *k-fold* lähenemist väärtustega $k = 5$ ja $k = 10$, jagades andmestiku vastavalt viieks ja kümneks osaks.

5 Tulemused

Järgnevat esipeatükikes esitletakse töö eksperimentaalset osa, mille eesmärk on hinnata andmehulkade sarnasust ning analüüsida erinevatel viisidel suurendatud andmehulkade põhjal treenitud mudelite klassifitseerimistulemusi. Esmalt antakse ülevaade Archimedese spiraalide jooniste sarnasusest: käsitletakse nii esialgsete spiraalide, afiinsete teisendustega suurendatud kui ka SMOTE'i abil täiendatud andmestike omavahelist sarnasust. Sarnasuse hindamiseks kasutati metoodika peatükis kirjeldatud kahte DTW meetodit: aegridade absoluutset kaugust ühetasandiliselt ning punktidevahelist kaugust eukleidiliselt. Spiraalide jooniste keskmisi sarnasustulemusi illustreerivad tabelid on leitavad lisast 4, mediaansarnasust ja spiraalide jooniste sarnasuse standardhälvet kirjeldavad tabelid on kättesaadavad lõputöö GitLabi projektis ¹. Mida väiksem on sarnasusskoor tabelis, seda sarnasemad on spiraalide joonised. Viimaks esitatakse töös kasutatud klassifikaatorite tulemused, leitavad samuti GitLabi projektist, ning võrreldakse mudelite toimivust erinevate andmehulkade suurendusmeetodite korral.

5.1 Esialgsete spiraalide jooniste sarnasus

Esialgsete spiraalide jooniste sarnasuse illustreerimiseks tehti soojuskaart, joonis 10. Soojuskaardilt ilmnes, et Parkinsoni põdejate spiraalide joonised olid üldiselt omavahel ning kontrollgrupiga võrreldes vähem sarnased, samas kui kontrollgrupi spiraalide joonised olid üksteisele üsna sarnased. Erandina paistsid silma PD16–PD19 ja KT13–KT29, mis moodustasid omavahel sarnasemad alamrühmad. Selgemate võrdluste saamiseks arvutati sarnasusskoorid kitsamate spiraalide rühmade vahel: PD1–PD15 ja KT13–KT29.

Väljavalitud esialgsete Archimedese spiraalide jooniste sarnasuse saamiseks arvutati spiraalide klasside vahelised sarnasused: Parkinsoni spiraalide sarnasus omavahel, kontrollgrupi spiraalide sarnasus omavahel ning Parkinsoni spiraalide sarnasus kontrollgrupi spiraalidega.

Mõlema meetodi põhjal arvatud tulemustest selgus, et Archimedese spiraalide joonised on kõige sarnasemad kontrollgrupis. See ilmnes nii madalaima keskmise sarnasustaseme kui ka väikseima standardhälbe kaudu, viidates kontrollgrupi joonistuste stabiilsusele. Vastupidiselt kontrollgrupi spiraalidele oli PD ja KT spiraalide jooniste sarnasus mõlema meetodi puhul madalaim, peegeldades erinevust spiraalide joonistamise liikumismustrites.

¹<https://gitlab.cs.taltech.ee/idjogi/iaib>

Parkinsoni patsientide (PD vs PD) joonistused olid omavahel keskmiselt sarnasemad kui PD ja KT vahelised joonistused, kuid suure standardhälbe tõttu võib järeldada, et haiguse mõju selle põdejatele on erinev.

5.2 Afiinsete teisendustega suurendatud andmehulkade sarnasus

Peatükk kirjeldab Archimedese spiraalidele rakendatud afiinsete teisenduste mõju spiraalide jooniste sarnasusele. Esmalt kirjeldatakse roteerimise mõju, seejärel skaleerimise mõju ning viimaks mõlema afiinse teisenduse koos rakendamise mõju. Samuti tuuakse välja roteerimisnurgad ning skaleerimisfaktorid, mille korral muutusid spiraalide joonised sarnaseks ja ka väärtused, mille korral muutusid spiraalide joonised ebasarnaseks.

5.2.1 Roteeritud spiraalide jooniste sarnasus

Sarnaselt esialgsete Archimedese spiraalidega arvutati ka roteeritud spiraalide jooniste puhul klassidevahelised sarnasused: näiteks esialgsete ja roteeritud spiraalide vaheline sarnasus nii Parkinsoni kui ka kontrollgrupi sees ja vahel. Tulemused näitasid, et ka pärast roteerimist jäid kontrollgrupi spiraalid omavahel kõige sarnasemaks. Roteerimine ei suurendanud aga kontrollgrupi spiraalide sisemist sarnasust: ühelgi roteerimisnurga korral ei olnud kontrollgrupi esialgsete ja roteeritud spiraalide vaheline sarnasus kõrgem kui esialgsete omavaheline sarnasus.

Parkinsoni klassi spiraalide joonised olid endiselt kõige vähem sarnased – isegi vähem kui erinevatesse klassidesse kuuluvad spiraalid. Roteerimine ei suurendanud ka nende klassisisest sarnasust. Siiski, esimese sarnasuse arvutamise meetodi järgi paranes Parkinsoni spiraalide jooniste sarnasus 90-kraadise roteerimise puhul, ületades nii nende algse sarnasuse kui ka klassidevahelise (KT–PD) sarnasuse. Samal pöördenurgal kasvas ka klassidevaheline sarnasus võrreldes esialgsete spiraalidega.

Esimese DTW-põhise meetodi järgi olid spiraalide joonised keskmiselt sarnaseimad, kui neid roteeriti 60, 75, 90, 105 või 120 kraadi võrra. Teise meetodi puhul ilmnes suurim sarnasus pöördenurkadel 45, 60, 75, 90 ja 105 kraadi. Mõlema meetodi puhul muutusid spiraalide joonised kõige ebasarnasemaks 225–285 kraadi juures.

Kokkuvõttes roteerimine spiraalide jooniseid sarnasemaks ei muutnud, võrreldes esialgse seisuga. Märkimisväärne on aga see, et roteerimine mõjutas enim Parkinsoni klassi spiraale, muutes need klassisisiselt veelgi ebasarnasemaks kui erinevatesse klassidesse kuuluvad spiraalid.

5.2.2 Skaleeritud spiraalide jooniste sarnasus

Lisaks spiraalide roteerimisele arvutati spiraalide jooniste sarnasust ka nende skaleerimise korral. Sarnasused leiti Parkinsoni spiraalide ja skaleeritud Parkinsoni spiraalide, Parkinsoni spiraalide ja skaleeritud kontrollgrupi spiraalide, kontrollgrupi spiraalide ja skaleeritud Parkinsoni spiraalide ning kontrollgrupi spiraalide ja skaleeritud kontrollgrupi spiraalide vahel. Ka skaleeritud spiraalide jooniste sarnasus leiti kasutades kahte DTW meetodit, nagu ka esialgsete spiraalide ja roteeritud spiraalide korral.

Sarnaselt eelnevatele tulemustele on ka skaleerides jätkuvalt sarnaseimad kontrollgrupi spiraalid. Nagu ka roteerimisel, ei muutunud skaleerides kontrollgrupi spiraalid omavahel sarnasemaks, kui nad olid esialgsete aegridade korral. Kui roteerimisel 90 kraadi võrra muutusid Parkinsoni spiraalide omavahel ning kontrollgrupi spiraalid ja roteeritud Parkinsoni spiraalid sarnasemaks esialgsete spiraalide sarnasusest, siis skaleerimise korral ei muutunud mitte ühelgi juhul spiraalid sarnasemaks esialgsetest spiraalidest.

Samuti spiraale skaleerides muutusid erinevatesse klassidesse kuuluvad spiraalid omavahel sarnasemaks kui Parkinsoni spiraalid omavahel. Leidus ka üks erand, nagu roteerimise korral. Nimelt kasutades eukleidilisel kaugusel põhinevat sarnasuse leidmise meetodit ning jättes spiraalide jooniste x -koordinaadi skaleerimata ja skaleerides y -koordinaati 0.95 korda, oli Parkinsoni spiraalide keskmine sarnasus omavahel suurem kui esialgsete Parkinsoni spiraalide ja samade väärtustega skaleeritud kontrollgrupi spiraalide vaheline sarnasus. Spiraalide jooniseid enim sarnaseks ja ebasarnaseks muutvad skaleerimisfaktorid on toodud tabelis 3

Tabel 3. Spiraalide jooniseid sarnaseks ja ebasarnaseks muutvad skaleerimisfaktorid.

Sarnaseks muutvad skaleerimisfaktorid absoluutse kauguse põhjal		Sarnaseks muutvad skaleerimisfaktorid eukleidilise kauguse põhjal		Ebasarnaseks muut- vad skaleerimisfakto- rid	
0.90	0.95	0.95	1.00	2.90	2.95
0.95	1.00	1.00	1.05	2.95	3.00
1.00	1.05	0.95	0.90	2.90	2.85
1.00	0.95	1.00	0.95	2.95	2.90
1.05	1.00	1.05	1.00	3.00	2.95

5.2.3 Roteeritud ja skaleeritud spiraalide jooniste sarnasus

Kuna spiraalide roteerimine ja skaleerimine seni kõigi roteerimiseks kasutatud nurkade ning skaleerimiseks kasutatud väärtustega oleks olnud nii arvutuslikult kui ka ajaliselt kulukas, otsustati roteerimise ja skaleerimise korraga rakendamise mõju spiraalide jooniste sarnasusele vaadelda, tuginedes afinsetel teisendustel spiraale sarnaseks ja ebasarnaseks muutvatel väärtustel.

Esimesena arvutati roteerimise ja skaleerimise mõju spiraalide jooniste sarnasusele, kasutades afinsete teisenduste põhjal spiraale sarnaseks muutvaid väärtusi. Esimese sarnasuse arvutamise meetodi puhul rakendati roteerimiseks nurki 60, 75, 90, 105 ja 120 kraadi ning skaleerimiseks väärtuseid, mis on tabelis 3. Teise sarnasuse leidmise meetodi korral roteeriti spiraale 45, 60, 75, 90 ja 105 kraadi võrra ning skaleeriti väärtustega, mis on samuti tabelis 3.

Mõlema sarnasuse leidmise meetodi tulemused sarnanesid eraldi rakendatud afinsete teisenduste tulemustele: kõige sarnasemad ka kahe afinse teisenduse koos rakendamisel on kontrollgrupi spiraalid ning Parkinsoni spiraalid olid omavahel keskmiselt ebasarnasemad kui erinevatesse klassidesse kuuluvad spiraalid. Roteerimise ja skaleerimise koos rakendamisel ei olnud ühtegi erandit, mille puhul oleks Parkinsoni spiraalide joonised olnud omavahel keskmiselt sarnasemad kui keskmine Parkinsoni spiraalide ja kontrollgrupi spiraalide vaheline sarnasus. Samuti ei muutnud roteerimise ja skaleerimise koos rakendamine spiraalide jooniseid mitte ühegi väärtuse korral sarnasemaks, kui nad olid enne ega ei muutnud spiraale ka nii sarnaseks kui nad olid roteerimise või skaleerimise eraldi rakendamisel.

Lisaks spiraalide jooniste roteerimisele ja skaleerimisele väärtustega, mis muutsid afinsete teisenduste eraldi rakendamisel sarnaseimaks, leiti ka spiraalide jooniste sarnasus tuginedes spiraale kõige ebasarnasemaks muutvatele väärtustele. Selleks roteeriti mõlema spiraalide sarnasuse leidmise meetodi korral spiraale 225, 240, 255, 270 ja 285 kraadi võrra ning skaleeriti spiraale tabelis 3 toodud skaleerimisfaktoritega.

Ka ebasarnaseks muutvate roteerimisnurkade ja skaleerimisfaktorite koos rakendamisel ei muutunud trend: kontrollgrupi spiraalide joonised on jätkuvalt omavahel keskmiselt kõige sarnasemad, Parkinsoni spiraalid on omavahel keskmiselt ebasarnasemad kui erinevatesse klassidesse kuuluvad spiraalid ning afinsete teisenduste koos rakendamisel muutusid spiraalide joonised veelgi ebasarnasemaks, kui nad olid afinsete teisenduste eraldi rakendamisel.

5.3 Andmehulkade tasakaalustamise mõju sarnasusele

Andmehulkade tasakaalustamiseks rakendati töös SMOTE-meetodit, mis lisab andmeid vähemuses olevasse klassi, et saavutada tasakaalustatud klassijaotus. Selle mõju hindamiseks spiraalide jooniste sarnasusele rakendati SMOTE algoritmi spiraalide koordinaatidele, mida kasutati sarnasuse arvutamisel. Selle tulemusena saadi mõlemasse klassi 17 aegrida, enne oli Parkinsoni klassis 15 ja kontrollgrupis 17 spiraali, võimaldades klassidevahelist võrdlust võrdse suurusega andmestikel.

SMOTE'i rakendamisel ümardas Pythonis teegi imbalanced-learn [86] algoritm koordinaatide väärtused neljanda komakohani, samas kui varasemates arvutustes kasutati 14 komakoha täpsust. Seetõttu ei ole kontrollgrupi spiraalide jooniste sarnasus täpselt sama, mis oli varem, kuid jääb samasse suurusjärku ega mõjuta üldist järeldust.

Sarnasuse analüüs näitas, et SMOTE'iga genereeritud Parkinsoni spiraalide jooniste omavaheline keskmine sarnasus jäi madalamaks kui kontrollgrupi spiraalide keskmine sarnasus ning ka madalamaks kui klassidevaheline keskmine sarnasus (Parkinson vs kontrollgrupp).

SMOTE'i kombineerimisel roteerimisega kehtisid samad seaduspärad ja erandid, mida on kirjeldatud alapeatükis 5.2.1. Sama kehtis ka skaleerimise puhul – tulemused vastasid varem skaleerimisega saadud mustritele (vt 5.2.2). Siiski kadus ära eelnev erand, mille puhul skaleerimisel oli Parkinsoni spiraalide jooniste omavaheline sarnasus suurem kui klassidevaheline sarnasus. Roteerimis- ja skaleerimisväärtused, mis varasemalt muutsid spiraalide jooniseid sarnaseks või ebasarnaseks, jäid kehtima ka SMOTE'i kasutamisel. Kui SMOTE'ile lisati korraga nii roteerimine kui skaleerimine, ei ilmnunud samuti uusi seaduspärasid spiraalide jooniste sarnasuses võrreldes nendega, mis on juba kirjeldatud alapeatükis 5.2.3.

5.4 Klassifikaatorite tulemused

Järgnev peatükk käsitleb kuue klassifikaatori tulemusi erinevate andmehulkade klassifitseerimisel, kasutades tabelis 2 toodud tunnuste valikuid. Esmalt tuuakse välja esialgsete Archimedese spiraalide tulemused, seejärel roteerimise, skaleerimise ja nende kombinatsiooniga suurendatud andmestike tulemused, ning lõpuks SMOTE'i mõju. Suurendatud andmehulkade mahud on lisas 5, klassifitseerimistulemusi võrdlevad joonised on lisas 6 ja klassifitseerimistulemuste tabelid on kättesaadavad lõputöö GitLabi projektis².

²<https://gitlab.cs.taltech.ee/idjogi/iaib>

5.4.1 Esialgsete spiraalide jooniste põhjal klassifitseerimine

AdaBoosti klassifikaator ennustas kõige paremini esialgsel andmehulgal, kui kasutati tunnuste valikut 3 või 4. Neid tunnuse valikuid ennustamisel kasutades saadi mudeli õiguseks 75%, täpsuseks 75.71%, saagiseks 75%, spetsiifilisuseks 66.67% ja F1-skooriks 74.83%. Otsustuspuu klassifikaator ennustas kõige paremini samuti tunnuste valiku 3 põhjal, nagu AdaBoost, ning sai ka samasugused mudeli ennustamismõõdikute tulemused. K-lähim naaber sai parimad tulemused ennustamisel, kui kasutati tunnuste valikuid 6 ja 7. Saades nendel juhtudel mudeli iga meetrika väärtuseks 83.33%.

Logistiline regressioon ennustas kõige paremini esialgsel andmehulgal, kui kasutati tunnuste valikut 7. Neid tunnuseid ennustamisel kasutades saadi mudeli õiguseks 91.67%, täpsuseks 92.86%, saagiseks 91.67%, spetsiifilisuseks 100% ja F1-skooriks 91.61%. Otus-tusmets klassifitseeris esialgseid andmeid kõige paremini tunnuste valiku 6 korral, mil mudeli õigus oli 75%, täpsuseks 75.71%, saagis 75%, spetsiifilisuseks 83.33% ja F1-skooriks 74.83%. Tugivektor-masin klassifitseeris andmeid sarnase edukusega nii tunnuste valikute 4, 6 ja 7 korral, saades mudeli õiguseks 83.33%, täpsuseks 83.33%, saagiseks 83.33%, spetsiifilisuseks 83.33% ja F1-skooriks 83.33%.

Kõige kõrgemad klassifitseerimistulemused esialgsetel andmetel saavutas logistilise regressiooni mudel. Kuna andmete treenimiseks oli kasutusel väike andmehulk ja logistilise mudeli saavutatud spetsiifilisusskoor 100% viitavad, et mudel võis olla ületreenitud ning saavutatud ennustustulemused üloptimislikud. Kõige stabiilisem kasutatud klassifikaatorites oli tugivektor-masin, mis saavutas kõrge õigsuse ja F1-skoori kolme tunnuste valiku korral. Tunnuste valiku 5 korral, mis põhinevad ainult surve andmetele, oli kõigi mudelite ennustamistulemused madalaimad, jäädes alla või 50% juurde.

5.4.2 Roteeritud spiraalide jooniste põhjal klassifitseerimine

Roteeritud spiraalide jooniste klassifitseerimiseks loodi kuus andmehulka. Esmalt suurendati treeningandmestikku, roteerides spiraale iga 15 kraadi järel vahemikus 15–345 kraadi. Seejärel kasutati vaid neid roteerimisnurki, mis varasemate sarnasustulemuste põhjal muutsid spiraalide jooniseid sarnaseks või ebasarnaseks. Kõigile kolmele suurendatud andmehulgale lisati ka algsed treeningandmed. Nende andmestikega treenitud mudelite tulemusi võrreldi esialgse, suurendamata andmehulgaga saadud tulemustega.

Roteerimisnurkadega 15-345 kraadi, sammuga 15 kraadi suurendatud andmehulga tulemused olid järgmised (joonis 11).

Mudelitel AB, RF klassifitseerimistulemused paranesid suurendatud andmehulgaga treenides vastavalt kahe ja kolme erineva tunnuste valiku korral ning vähenesid kahe tunnuste valiku korral. Mudeli DT klassifitseerimistulemused paranesid viie erineva tunnuste valiku korral, kuid halvenesid ühe tunnuste valiku korral. Mudelite KNN, LR klassifitseerimistulemused paranesid roteerimisega suurendatud andmehulga treenimisel kahe tunnuste valiku korral. Siiski oli ka tunnuste valikuid, mille korral klassifitseerimistulemused nendel mudelitel halvenesid, KNN mudelil kolme erineva valiku korral ja LR mudelil ühe tunnuste valiku korral. Kõige halvemini mõjutas roteeritud andmehulgaga treenimine mudelit SVM, mille klassifitseerimistulemused halvenesid viie erineva tunnuste valiku korral.

Esimese meetodi põhjal sarnaseks muutvate nurkadega suurendatud andmehulga tulemused olid järgmised (joonis 11). Mudelite KNN ja SVM paranesid kahe erineva tunnuste valiku korral. Mudeli AB, DT ja RF klassifitseerimistulemused paranesid vastavalt ühe, kolme ning nelja erineva tunnuste valiku korral. Mudeli LR korral halvenesid tulemused ühe tunnuste valiku põhjal klassifitseerimisel ning paranesid kahe korral, mudeli RF klassifitseerimisvõimekus langes kahe erineva tunnuste valikuga, mudelite AB, DT ja KNN klassifitseerimistulemused halvenesid kolme tunnuste valiku korral. Taaskord mõjutas kõige halvemini sarnaseks muutvate nurkadega roteerimise abil suurendatud andmehulgaga treenimine mudelit SVM, mille klassifitseerimistulemused halvenesid nelja erineva tunnuste valiku korral.

Teise meetodi põhjal sarnaseks muutvate nurkadega suurendatud andmehulga tulemused olid järgmised (joonis 11). Mudelite KNN, SVM tulemused paranesid kahe erineva tunnuste valiku korral. Mudeli AB tulemused paranesid ühe tunnuste valiku korral. Mudelite DT, RF klassifitseerimistulemused paranesid vastavalt nelja ja viie erineva tunnuste valiku ning halvenesid ühe tunnuste valiku korral. Mudeli LR klassifitseerimisvõime langes kolme erineva tunnuste valikuga ja paranes ühe korral, mudelite AB, KNN klassifitseerimistulemused halvenesid kolme tunnuste valiku korral. Mudeli SVM klassifitseerimistulemused halvenesid, sarnaselt esimese meetodi järgi sarnaseks muutvate roteerimisnurkade abil suurendatud andmehulga treenimisega, nelja erineva tunnuste valiku korral.

Mudelitel treenimisel ebasarnaseks muutvate roteerimisnurkadega suurendatud andmehulgaga saadi järgmised klassifitseerimistulemused (joonis 11). Mudeli AB tulemused paranesid kahe erineva tunnuste valiku ning halvenesid kahe korral. Mudeli DT tulemused paranesid nelja erineva tunnuste valiku korral ning halvenesid ühe korral. Mudelite KNN, SVM tulemused paranesid kahe erineva tunnuste valiku korral ja halvenesid vastavalt kolme ja nelja erineva tunnuste valikuga. Mudeli LR tulemused paranesid ühe ja halvenesid kahe tunnuste valikute korral. Mudeli RF tulemused paranesid viie erineva tunnuste valiku korral ja halvenesid ainult ühe tunnuste valiku korral.

Juba roteeritud andmehulkadele lisati esialgsed andmehulgad, et näha kas ja kuidas need mõjutavad mudelite klassifitseerimistulemusi. Esialgsete andmete lisamine juba roteeritud andmehulkadesse mõjutas tulemusi võrreldes lihtsalt roteeritud andmehulkadega järgmiselt (joonis 12). Esialgsete spiraalide jooniste lisamine kõikide töös kasutatavate roteermisnurkadega suurendatud andmehulka parandas mudelite RF, SVM tulemusi. Mudeli KNN tulemused ei muutunud, mudeli DT tulemused halvenesid kahe erineva tunnuste valiku korral ja mudeli LR tulemused paranesid kahe tunnuste valiku korral. Mudelite AB tulemused paranesid ühe tunnuste valiku korral ja ka halvenesid ühe tunnuste valiku korral. Esialgsete spiraalide jooniste lisamine esimese meetodi põhjal sarnaseks muutvate roteermisnurkadega suurendatud andmehulka parandas mudelite DT, LR ja SVM klassifitseerimistulemusi, mudeli RF tulemused halvenesid kahe tunnuste valiku korral. Mudelite AB, KNN tulemused paranesid ühe tunnuste valiku korral ja halvenesid vastavalt kahe ning ühe tunnuste valiku korral. Esialgsete spiraalide jooniste lisamine teise meetodi põhjal sarnaseks muutvate roteermisnurkadega suurendatud andmehulka parandas üldiselt mudelite DT, LR klassifitseerimistulemusi, mudeli KNN tulemused halvenesid kahe tunnuste valiku korral. Mudeli RF tulemused paranesid ühe tunnuste valiku korral ja halvenesid kahe tunnuste valiku korral. Mudeli AB tulemused halvenesid ühe tunnuste valiku korral. Mudeli SVM tulemused ei muutunud. Esialgsete spiraalide jooniste lisamine ebasarnaseks muutvate roteermisnurkadega suurendatud andmehulka parandas mudeli LR klassifitseerimistulemusi, mudelite AB, DT, KNN tulemused halvenesid vastavalt ühe, nelja ning ühe tunnuste valiku korral. Mudeli RF tulemused paranesid ühe ja halvenesid kahe erineva tunnuste valiku korral. Mudeli SVM tulemused ei muutunud.

5.4.3 Skaleeritud spiraalide jooniste põhjal klassifitseerimine

Ka skaleeritud spiraalide klassifitseerimiseks loodi kuus andmehulka. Esmalt suurendati treeningandmestikku, skaleerides spiraalide jooniste koordinaate kahel viisil: x -koordinaate väärtustega 0.1–3.0 ja y -koordinaate väärtustega 0.05–2.95 ning vastupidi, sammuga 0.05. Seejärel kasutati ainult neid skaleerimisfaktoreid, mis sarnasustulemuste põhjal muutsid spiraalide joonised vastavalt sarnasemaks või ebasarnasemaks. Kõigile kolmele suurendatud andmehulgale lisati ka algsed treeningandmed. Mudelite klassifitseerimistulemusi võrreldi esialgse, suurendamata andmestikuga saadud tulemustega.

Esialgsete andmete mõlema koordinaadi skaleerimisel skaleerimisfaktoritega 0.05 - 3, sammuga 0.05 klassifitseerimistulemused olid järgmised (joonis 13). Mudelite LR, RF tulemused paranesid suurendatud andmehulgaga treenides vastaval kolme ja kahe erineva tunnuste valiku korral. Siiski vähenesid mudeli RF tulemused kahe tunnuste valiku korral. Mudelite DT, KNN, SVM tulemused paranesid vastavalt kolme, kahe ja kahe erineva tunnuste valiku korral, kuid halvenesid vastavalt kahe, nelja ja kolme tunnuste valiku korral.

Mudeli AB tulemused paranesid skaleerimisega suurendatud andmehulga treenimisel kahe tunnuste valiku korral ja halvenesid nelja tunnuste valikuga.

Esimese meetodi põhjal sarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulga tulemused olid järgmised (joonis 13). Mudelite AB ja DT tulemused paranesid ainult ühe tunnuste valiku korral ning halvenesid vastavalt kahe ning nelja erineva tunnuste valiku korral. Mudeli LR tulemused halvenesid ühe tunnuste valiku korral. Mudeli KNN klassifitseerimistulemused paranesid ja ka halvenesid kahe erineva tunnuste valiku korral. Mudelite RF ja SVM klassifitseerimistulemused paranesid kolme tunnuste valiku põhjal, kuid ka halvenesid vastavalt kolme ja kahe tunnuste valikuga.

Teise meetodi põhjal sarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulga klassifitseerimistulemused olid järgmised (joonis 13). Mudelite DT, RF ja SVM tulemused paranesid kolme erineva tunnuste valiku korral ning halvenesid vastavalt mudelitele kahe, kolme ja kahe tunnuste valiku korral. Mudeli KNN klassifitseerimistulemused paranesid kahe erineva tunnuste valiku korral ning klassifitseerimine halvenes kolme erineva tunnuste valiku põhjal. Mudeli AB tulemused paranesid ühe tunnuste valikuga, kuid halvenesid kahe tunnuste valikuga. Mudeli LR tulemused halvenesid ühe tunnuste valiku korral.

Ebasarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulga klassifitseerimistulemused olid järgmised (joonis 13). Mudelite AB ja DT tulemused paranesid ainult ühe tunnuste valiku korral ning halvenesid viie korral. Ka mudeli RF tulemused paranesid ainult ühe tunnuste valiku korral ja vähenesid nelja erineva tunnuste valiku korral. Mudeli LR tulemused halvenesid kolme erineva tunnuste valiku korral. Mudelite KNN ja SVM tulemused paranesid kahe erineva tunnuste valiku korral. Nende mudelite klassifitseerimistulemused halvenesid vastavalt nelja ja kolme erineva tunnuste valikuga.

Sarnaselt roteeritud andmehulkadele lisati ka juba skaleeritud andmehulkadele esialgsed andmehulgad, et näha, kas ja kuidas need mõjutavad mudelite klassifitseerimistulemusi. Esialgsete andmete lisamine mõlema koordinaadi skaleerimisel skaleerimisfaktoritega 0.05 - 3, sammuga 0.05 suurendatud andmehulka mõjutas tulemusi võrreldes lihtsalt niiviisi saadud andmehulga rakendamisel järgmiselt (joonis 14). Mudelite AB ja DT tulemused paranesid. Mudelite KNN, SVM klassifitseerimistulemused ei muutunud ning mudelite LR, RF tulemused halvenesid kahe tunnuste valiku korral. Esialgsete andmete lisamine esimese meetodi põhjal sarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulka parandas ning ka vähendas mudelite AB, LR tulemusi ühe tunnuste valiku korral. Mudeli DT tulemused paranesid nelja tunnuste valiku korral ja mudeli RF tulemused kahe tunnuste valikuga. Mudelite KNN ja SVM tulemused halvenesid ühe tunnuste valiku korral.

Esialgsete spiraalide jooniste andmete lisamine teise meetodi põhjal sarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulka parandas mudelite AB, KNN, RF tulemusi. Mudeli DT tulemused paranesid kahe ja halvenesid ühe tunnuste valiku korral. Mudelite SVM ja LR tulemused halvenesid ühe tunnuste valiku korral. Esialgsete treeningandmete lisamine ebasarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulka ainult parandas mudelite AB, DT, KNN ja RF klassifitseerimistulemusi. Mudeli SVM tulemused paranesid kahe tunnuste valiku korral ning halvenesid ühe tunnuste valikuga. Mudeli LR tulemused ei muutunud.

5.4.4 Roteeritud ja skaleeritud spiraalide jooniste põhjal klassifitseerimine

Roteerimise ja skaleerimise kombinatsiooniga loodi samuti kuus andmehulka. Esmalt suurendati treeningandmestikku, kasutades spiraalide jooniseid sarnasemaks muutvaid roteerimisnurki ja skaleerimisfaktoreid, seejärel ainult ebasarnasemaks muutvaid väärtusi. Lisaks koostati andmehulgad, kus kombineeritud teisendustega suurendatud andmetele lisati ka algsed treeningandmed. Kõigi andmehulkade põhjal treenitud mudelite klassifitseerimistulemusi võrreldi esialgse, suurendamata andmestikuga saadud tulemustega.

Esimese meetodi põhjal sarnaseks muutvate roteerimisnurkade ja skaleerimisfaktoritega suurendatud andmehulga klassifitseerimistulemused olid järgmised (joonis 15). Mudelite KNN, LR ja SVM tulemused paranesid kahe tunnuste valikute korral ning halvenesid vastavalt kahe, kolme ja nelja tunnuste valiku korral. Mudelite AB ja RF tulemused paranesid kolme erineva tunnuste valikuga ning halvenesid vastavalt mudelile kolme ja kahe erineva tunnuste valiku korral. Mudeli DT tulemused paranesid viie tunnuste valiku põhjal, kuid halvenesid ühe tunnuste valikuga.

Teise meetodi põhjal sarnaseks muutvate roteerimisnurkade ja skaleerimisfaktoritega suurendatud andmehulga tulemused olid järgmised (joonis 15). Mudelite KNN ja SVM tulemused paranesid kahe erineva tunnuste valiku korral ning halvenesid vastavalt kahe ja nelja tunnuste valiku korral. Mudelite AB, LR tulemused paranesid vastavalt kolme ja kahe erineva tunnuste valiku korral ning halvenesid kahe ja kolme valiku puhul. Mudeli DT tulemused paranesid nelja erineva tunnuste valiku korral ning klassifitseerimine halvenes ühe erineva tunnuste valiku põhjal. Mudeli RF tulemused paranesid viie tunnuste valikuga, kuid halvenesid ühe tunnuste valikuga.

Ebasarnaseks muutvate roteerimisnurkade ja skaleerimisfaktoritega suurendatud andmehulga tulemused olid järgmised (joonis 15).

Mudelite AB, LR, KNN ja SVM klassifitseerimistulemused paranesid kahe erineva tunnuste valiku korral ning halvenesid vastavalt nelja, kolme, nelja ja viie korral. Mudelite RF, DT tulemused paranesid kolme erineva tunnuste valiku korral ja halvenesid samuti kolme erineva tunnuste valikuga.

Sarnaselt nii roteeritud kui ka skaleeritud andmehulkadele lisati ka juba roteeritud ja skaleeritud andmehulkadele esialgne andmehulk. Esialgsete andmete lisamine juba roteeritud ja skaleeritud andmehulkadesse mõjutas klassifitseerimistulemusi võrreldes lihtsalt roteeritud ja skaleeritud andmehulkadega mudelite treenimisel järgmiselt (joonis 16). Esialgsete andmete lisamine esimese meetodi põhjal sarnaseks muutvate roteerimisnurkade ja skaleerimisfaktoritega suurendatud andmehulka parandas mudelite AB, LR, KNN ning RF tulemusi. Mudeli DT tulemused paranesid ühe tunnuste valiku puhul ja halvenesid samuti ühe tunnuste valiku korral. Mudeli SVM tulemused ei muutunud. Esialgsete spiraalide andmete lisamine teise meetodi põhjal sarnaseks muutvate roteerimisnurkade ja skaleerimisfaktoritega suurendatud andmehulka parandas mudeli AB tulemusi. Mudelite DT, KNN ja RF tulemused paranesid ühe tunnuste valiku korral ning halvenesid vastavalt kolme, ühe ja ühe tunnuste valiku korral. Mudeli LR tulemused nii vähenesid kui ka paranesid esialgsete andmete lisamisel ühe tunnuste valiku korral ning mudeli SVM tulemused ei muutunud. Esialgsete treeningandmete lisamine ebasarnaseks muutvate roteerimisnurkade ja skaleerimisfaktoritega suurendatud andmehulka parandas mudeli KNN tulemusi. Mudelite DT, RF tulemused paranesid kolme tunnuste valiku korral ning halvenesid ühe tunnuste valikuga. Mudeli AB tulemused paranesid nelja tunnuste valiku korral ja halvenesid ühe puhul. Mudeli LR tulemused ei paranenud, vaid halvenesid kahe tunnuste valiku korral. Mudeli SVM tulemused ei muutunud.

5.4.5 Tasakaalustatud andmehulkade põhjal klassifitseerimine

Tasakaalustatud andmehulkade klassifitseerimiseks tasakaalustati kõik eelnevalt mainitud andmehulgad: esialgne andmehulk, roteerimise abil suurendatud andmehulgad, skaleerimise abil suurendatud andmehulgad ning roteeritud ja skaleeritud andmehulgad. Kõigi andmehulkade põhjal treenitud mudelite klassifitseerimistulemusi võrreldi esialgse, suurendamata andmestikuga saadud tulemustega.

Esialgse treeningandmehulga tasakaalustamine SMOTE-iga tõi peamiselt positiivseid muudatusi klassifitseerimistulemustesse (joonis 17). Mudelite AB, DT, KNN, SVM tulemused paranesid tasakaalustatud andmehulgaga treenides vastavalt nelja, kolme, kolme ja ühe tunnuste valiku korral, võrreldes esialgse andmehulgaga. Mudelite LR ja RF tulemused paranesid andmehulga tasakaalustamisel vastavalt ühe ja viie tunnuste valiku korral ning halvenesid ühe tunnuste valiku korral.

Kõikide valikus olevate nurkadega roteeritud andmehulga tasakaalustamine ei muutnud eriliselt tulemusi võrreldes lihtsalt kõikide nurkadega roteerimisega (joonis 18). Kolme mudeli puhul olid klassifitseerimistulemused, võrreldes esialgsete tulemustega, samasugused, nagu olid roteeritud andmehulga tulemused, võrreldes esialgse andmehulgaga. Mudelite DT, LR, RF korral halvenes klassifitseerimine vastavalt kahe, ühe ja ühe tunnuste valikuga. Esimese meetodi põhjal sarnasemaks muutvate nurkadega andmehulga tasakaalustamine andis veidi erinevaid klassifitseerimistulemusi (joonis 18). Mudelite AB, DT klassifitseerimine muutus üldiselt paremaks. Mudelite LR ja SVM tulemused halvenesid ühe tunnuse valiku korral. Mudeli RF tulemused paranesid ning ka halvenesid ühe tunnuste valiku korral. Mudeli KNN tulemused jäid samasuguseks. Teise meetodi põhjal sarnasemaks muutvate nurkadega andmehulga tasakaalustamine andis suhteliselt samasuguseid tulemusi võrreldes tasakaalustamata andmehulgaga (joonis 18). Mudelite AB, KNN, RF, LR klassifitseerimine jäi samasuguseks, mudeli DT klassifitseerimine muutus paremaks ühe tunnuste valiku korral ning mudeli SVM tulemused halvenesid ühe tunnuse valiku korral. Ka ebasarnaseks muutvate roteerimisnurkadega andmehulga tasakaalustamine ei muutnud eriliselt tunnuste põhist klassifitseerimist (joonis 18): mudelite AB, KNN, LR ja RF tulemused jäid samasuguseks, võrreldes esialgsetega. Mudeli DT tulemused paranesid ühe tunnuste valiku korral ning mudeli SVM tulemused halvenesid ühe tunnuste valiku korral.

Skaleeritud andmehulga, mille loomiseks kasutati kõiki töös kasutatavaid skaleerimisfaktoreid, tasakaalustamine tõi võrreldes lihtsalt kõigi skaleerimisfaktoritega suurendatud andmehulgaga klassifitseerides paremaid tulemusi (joonis 20). Mudelite AB ja RF klassifitseerimistulemused paranesid, mudeli KNN klassifitseerimistulemused jäid samaks. Mudeli DT tulemused paranesid kahe tunnuste valiku korral ja halvenesid ühe valiku korral. Mudelite LR ja SVM tulemused ainult halvenesid. Esimese meetodi põhjal sarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulga tasakaalustamisel (joonis 20) paranesid mudelite AB, DT ja RF tulemused, võrreldes sama andmehulga tasakaalustamata variandiga. Mudeli KNN tulemused jäid samasuguseks. Mudeli LR tulemused paranesid ja ka halvenesid ühe tunnuste valiku korral. Mudeli SVM tulemused halvenesid. Teise meetodi põhjal sarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulga tasakaalustamine (joonis 20) parandas mudelite AB, DT, KNN ja RF tulemusi. Mudeli LR tulemused jällegi paranesid ja ka halvenesid ühe tunnuste valiku korral. Mudeli SVM tulemused halvenesid. Ebasarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulga tasakaalustamine (joonis 20) parandas klassifitseerimistulemusi mudelitel AB, DT ja RF. Mudeli KNN puhul jäid tulemused samasuguseks ning mudelite LR ja SVM puhul halvenesid.

Roteeritud ja skaleeritud andmehulga tasakaalustamine mõjus mudelitele erinevalt.

Esimese meetodi põhjal sarnaseks muutvate väärtustega suurendatud andmehulga tasakaalustamine ei parandanud mudelite tulemusi (joonis 22). Mudelite AB, KNN ja RF tulemused jäid samasuguseks ning mudelite DT, LR ja SVM tulemused halvenesid tasakaalustamisel. Teise meetodi järgi sarnaseks muutvate väärtustega suurendatud andmehulga tasakaalustamine tõi positiivsemaid tulemusi (joonis 22): mudelite AB, DT tulemused paranesid, võrreldes lihtsalt roteerimise ja skaleerimisega. Mudelite KNN, RF tulemused jäid samasuguseks ning mudelite LR, SVM tulemused vähenesid. Ebasarnaseks muutvate väärtustega suurendatud andmehulga tasakaalustamine mõjus järgmiselt (joonis 22). Mudelite AB, KNN tulemused ei muutunud. Mudeli RF tulemused nii paranesid kui ka halvenesid ühe tunnuste valiku korral. Mudelite DT, LR, SVM tulemused halvenesid.

Esialgsete andmete lisamine juba tasakaalustatud ja afiinsete teisendustega suurendatud andmehulkadesse mõjutas klassifitseerimistulemusi võrreldes lihtsalt tasakaalustatud ja afiinsete teisendustega suurendatud andmehulkadega mudelite treenimisel järgmiselt. Esialgsete andmete lisamine kõikide roteerimisnurkadega suurendatud andmehulgale ja andmehulga tasakaalustamine (joonis 19) parandas mudeli DT tulemusi, mudelite LR ja SVM tulemused ei muutunud. Mudelite AB, KNN ja RF tulemused paranesid vastavalt ühe, kahe ja kahe ning langesid ühe tunnuste valiku korral. Esialgsete andmete lisamine esimese meetodi järgi sarnaseks muutvate roteerimisnurkadega suurendatud andmehulgale ja andmehulga tasakaalustamine (joonis 19) parandas mudelite LR ja SVM tulemusi. Mudeli KNN tulemused halvenesid. Mudelite AB, DT ja RF tulemused paranesid vastavalt kolme, ühe ja ühe erineva tunnuste valiku korral ning halvenesid vastavalt ühe, kolme ja kahe korral. Esialgsete andmete lisamine teise meetodi järgi sarnaseks muutvate roteerimisnurkadega suurendatud andmehulgale ja andmehulga tasakaalustamine (joonis 19) parandas samuti mudelite LR ja SVM tulemusi. Mudelite AB, DT, KNN, RF tulemused paranesid vastavalt kolme, kahe, ühe ja ühe tunnuste valiku korral ning halvenesid vastavalt ühe, kahe, ühe ja ühe tunnuste valiku korral. Ebasarnaste väärtustega roteeritud andmehulgale esialgsete spiraalide jooniste lisamine (joonis 19) parandas mudelite AB, DT, LR, SVM tulemusi. Mudeli KNN tulemused halvenesid ühe tunnuste valiku korral. Mudeli RF tulemused paranesid kahe ning halvenesid ühe tunnuste valiku korral. Kõikide skaleerimisfaktoritega suurendatud andmestikku täiendades esialgsete spiraalidega ja tasakaalustades (joonis 21) paranesid mudeli KNN tulemused, võrreldes pelgalt tasakaalustatud andmestikuga. Mudelite AB, DT tulemused halvenesid ja mudelite LR, RF ja SVM tulemused ei muutunud. Esimese meetodi põhjal sarnaseks muutvate skaleerimisfaktoritega suurendatud andmehulga esialgsete andmete lisamine (joonis 21) parandas AB mudelit ühe tunnuste valiku puhul, kuid vähendas tulemusi kahe erineva tunnuste valikuga. Mudelite DT ja KNN tulemused halvenesid. Mudelite RF, LR, SVM tulemused paranesid ühe tunnuste valiku korral ja halvenesid vastavalt kahe, ühe ja ühe korral.

Teise meetodi järgi suurendatud andmehulka esialgsete andmete lisamine (joonis 21) parandas mudelite AB, RF tulemusi, kuid vähendas mudelite DT, LR, SVM tulemusi. Mudeli KNN tulemused paranesid kahe ja halvenesid ühe tunnuste valiku korral. Ebasarnaste skaaleerimisfaktoritega suurendatud andmestikku esialgsete lisamine (joonis 21) parandas kõigi mudelite täpsust, kuid tulemused ka halvenesid ühe tunnuste valiku korral mudelitel AB, DT, KNN, RF ja SVM. Esimese meetodi põhjal sarnaseks muutvate väärtustega roteeritud ja skaaleeritud andmestikku esialgseid spiraale lisades ja tasakaalustades (joonis 23) paranesid RF, SVM tulemused. Mudelite DT, KNN tulemused paranesid kahe ning halvenesid vastavalt ühe ja kahe tunnuste valiku korral. Mudeli AB tulemused halvenesid ja mudeli LR tulemused ei muutunud. Teise meetodi järgi halvenesid samas olukorras (joonis 23) mudeli AB tulemused. Mudelite LR, SVM tulemused ei muutunud. Mudelite DT, KNN ja RF tulemused paranesid vastavalt kahe, ühe ja kahe ning halvenesid vastavalt kahe, kahe ja kolme korral. Ebasarnaste väärtustega suurendatud andmestiku täiendamine esialgsete andmetega ja tasakaalustamine (joonis 23) parandas mudelite AB, SVM tulemusi, mudeli LR tulemused ei muutunud. DT, KNN, RF mudelite tulemusi paranesid vastavalt kolme, kahe ja kahe tunnuste valiku korral ning halvenesid vastavalt ühe, ühe ja kahe korral.

5.5 Tulemuste järeldused

Töö eesmärgiks oli uurida Archimedese spiraalide jooniste sarnasust Parkinsoni tõve kontekstis ning hinnata, kuidas erinevate andmehulkade suurendamise meetoditega treenitud mudelid suudavad eristada Parkinsoni põdejaid kontrollgrupist.

Kontrollgrupi spiraalide joonised osutusid kõige sarnasemateks nii enne kui ka pärast teisendusi. Neil oli madalaim keskmine sarnassuskoor ja väikseim varieeruvus, mis viitab joonistuste klassisisesele stabiilsusele. Parkinsoni patsientide spiraalide joonised olid omavahel vähem sarnased kui kontrollgrupi omad ning esialgsete andmete puhul olid Parkinsoni ja kontrollgrupi spiraalide joonised omavahel veel vähem sarnasemad, mis peegeldab haiguse mõju käekontrollile ja liikumismustritele.

Afinsete teisenduste rakendamisel muutusid Parkinsoni spiraalide joonised omavahel vähem sarnaseks kui erinevatesse klassidesse kuuluvad spiraalide joonised. See viitab sellele, et Parkinsoni tõvega seotud joonistuste varieeruvus on juba algselt väga kõrge ning teisendused, nagu roteerimine ja skaaleerimine, võivad olemasolevaid individuaalseid eripärasid veelgi võimendada. Lisaks võib Dynamic Time Warping meetod, mida kasutati sarnasuse hindamiseks, olla tundlik geomeetriliste moonutuste suhtes ega pruugi arvestada, et visuaalselt sarnased spiraalid võivad pärast teisendusi mõõdikute järgi tunduda erinevad. Seega ei pruugi afinsete teisenduste rakendamine alati parandada Parkinsoni spiraalide jooniste klassisisest ühtlust, vaid võib vastupidi suurendada nende vahelisi kaugusi.

Andmehulga tasakaalustamine avaldas sarnasusele mõju sarnaselt afiinselt teisendustele: Parkinsoni klassi spiraalide jooniste omavaheline sarnasus jäi madalamaks kui teiste arvutatud sarnasuste puhul. Kuna SMOTE-i ja afiinsete teisenduste koosrakendamisel ei muutunud sarnasusskoorid oluliselt, võrreldes ainult afiinsete teisendustega, võib järeldada, et SMOTE-il oli andmehulkade sarnasusele väiksem mõju kui afiinsetel teisendustel.

Afiinsete teisenduste rakendamine avaldas klassifitseerimistulemustele erinevat mõju. Roteerimise puhul ei ilmnunud selget vahet, kas spiraalide jooniste hulka suurendati sarnaseks või ebasarnaseks muutvate roteerimisnurkade abil – tulemused jäid kõigi suurendatud andmehulkadega mudelite treenimisel sarnasele tasemele. See viitab, et spiraali orientatsioon ei ole mudelite jaoks määrav tunnus ning roteerimine üksi ei lisa sisuliselt eristavat infot.

Skaleerimise mõju oli selgelt sõltuv kasutatud skaleerimisfaktoritest. Kui suurendamine toimus sarnaseks muutvate skaleerimisväärtustega, paranesid klassifitseerimistulemused märgatavalt. Ebasarnaseks muutvate skaleerimisväärtustega suurendatud andmehulgad aga halvendasid klassifitseerimist. Eriti tundlikumate mudelite, nagu logistiline regressioon ja tugivektor-masin korral. Parimad tulemused saavutati, kui andmehulka suurendati kombineeritult nii roteerimise kui ka skaleerimise abil, kasutades üksnes sarnaseks muutvaid teisendusi. Afiinsete teisenduste koor rakendamine säilitas andmete loomuliku struktuuri ning toetas mudelite üldistusvõimet.

SMOTE-põhine andmehulga tasakaalustamine mõjus positiivselt klassifitseerimistulemustele eelkõige esialgse andmestiku korral. Mitmete mudelite, eriti AdaBoosti ja otsustusmetsa, täpsus ja stabiilsus paranesid, kui SMOTE-i abil loodi tasakaalustatud treeningandmestik. Kui aga SMOTE-i kombineeriti juba teisendatud (roteeritud ja/või skaleeritud) andmetega, sõltus mõju tugevalt kasutatud teisendustest. Sarnaseks muutvate väärtustega suurendatud andmehulkade puhul oli SMOTE-i mõju mõnel mudelil positiivne, ent ebasarnaseks muutvate väärtustega suurendatud andmestike puhul ei parandanud SMOTE-i tulemusi ning sageli halvenesid tulemused: eriti just mudelite SVM, LR ja DT puhul. See viitab, et SMOTE-i on kasulik tasakaalustamiseks juhul, kui algandmete kvaliteet ja struktuur säilib, kuid kui andmed on juba tugevalt moonutatud või ebaloomulikud, võib SMOTE pigem süvendada mudeli segadust andmete klassifitseerimisel.

Lisaks selgus, et suurendatud andmestike täiendamine esialgsete treeningandmetega parandas mõnel juhul klassifitseerimistulemusi. Kui mudel sai õppida nii algsetelt kui ka suurendatud näidetelt, säilis andmete algne struktuur paremini ning suurenes mudeli võime eristada Parkinsoni ja kontrollgrupi spiraalide jooniseid. Eriti oli see nähtav sarnaseks muutvate teisenduste korral, kus algandmete kaasamine aitas vähendada ebaloomulike teisenduste võimalikku mõju.

See viitab, et suurendatud andmestike kombineerimine algandmetega võib olla tõhus strateegia mudeli üldistusvõime parandamiseks.

Teema edasiarenduseks soovitab autor uurida andmehulkade suurendamist ka teiste meetodite abil. Näiteks võiks katsetada, kuidas mõjub sarnasusele ja klassifitseerimisele andmete suurendamine generatiivsete meetoditega. Samuti võiks kasutada teisi sarnasuse leidmise viise, et võrrelda, kas mõni muu meetod annab stabiilsemaid või paremini eristuvaid tulemusi kui kasutatud DTW-põhised lähenemised. Lisaks võiks masinõppe meetodina kasutada tehismärgivõrke, näiteks konvolutsioonilisi või rekurrentseid märgivõrke, mis võivad suuta iseseisvalt tuvastada Parkinsonile iseloomulikke joonemustreid ilma eelnevalt arvutatud tunnusteta. Tööd võiks ka laiendada teistele joonistusülesannetele, et uurida, milline on andmehulkade suurendamismeetodite rakendamise mõju neile. Kuna töös ei suurendatud andmeid surveandmete põhjal, võiks ka see olla eraldi suund edasiarenduseks. Tulevikus võiks katsetada, kuidas andmehulkade suurendamine lisaks aja- ja trajektoori põhiste tunnustele ka survepõhiste tunnuste alusel mõjutab spiraalide jooniste sarnasust ja klassifitseerimistulemusi.

6 Kokkuvõte

Lõputöös uuriti Parkinsoni tõve tuvastamise võimalusi joonistusülesande abil, keskendudes Archimedese spiraalide analüüsile ning andmehulkade suurendamise meetodite mõjule. Töö eesmärk oli hinnata, kuidas mõjutavad suurendatud andmehulgad spiraalide jooniste sarnasust ning milline on nende mõju masinõppemudelite klassifitseerimisvõimele. Kasutati DraWritePD andmestikku, mis sisaldas Parkinsoni põdejate ja kontrollisikute joonistatud spiraale. Spiraalidevahelist sarnasust mõõdeti DTW-põhiste meetoditega, andmehulkade suurendamiseks rakendati afinseid teisendusi: roteerimine ja skaleerimine, ning andmehulkade tasakaalustamiseks kasutati SMOTE meetodit. Klassifitseerimisel kasutati kuut mudelit ja 19 spiraalide jooniste põhjal arvutatud tunnust. Mudelite puhul hinnati nende täpsust, tundlikkust, spetsiifilisust, F1-skoori ja üldist õigsust.

Sarnasusanalüüsi tulemused näitasid, et kõige sarnasemad olid kontrollgrupi spiraalide joonised, mille varieeruvus oli madal. Parkinsoni spiraalide joonised olid varieeruvamad ning afinsete teisenduste ja tasakaalustamise järel muutusid erinevatesse klassidesse kuuluvad spiraalide joonised keskmiselt sarnasemaks kui Parkinsoni spiraalide joonised omavahel. Töö käigus kasutatud roteerimisnurgad ja skaleerimisfaktorid ei suurendanud spiraalide jooniste omavahelist sarnasust – teisendused pigem tõid esile olemasolevaid erinevusi.

Klassifitseerimistulemustes ilmnas, et sarnaseks muutvad teisendused parandasid ennustusõigsust, eriti skaleerimise puhul. Roteerimise mõju jäi üldiselt tagasihoidlikuks. Parimad tulemused saavutati sarnaseks muutvate teisenduste kombineerimisel. Juba teisendatud andmestike täiendamine esialgsete treeningandmetega parandas mitmel juhul mudelite üldistusvõimet. SMOTE meetod mõjus hästi esialgse andmestiku puhul, kuid teisendatud andmete korral sõltus selle mõju tugevalt teisenduse tüübist.

Tulevikus võiks uurida ka teisi suurendusmeetodeid, sügavõppemudeleid, erinevaid sarnasuse mõõdikuid ning teisi joonistusülesandeid. Samuti võiks katsetada surveandmete põhjal andmehulga suurendamist ning uurida, sellisel viisil suurendatud andmehulga mõju sarnasusele ja klassifitseerimisele.

Kasutatud kirjandus

- [1] M. Sokolova ja G. Lapalme, „A systematic analysis of performance measures for classification tasks,“ *Information Processing & Management*, köide 45, nr 4, lk. 427–437, 2009. DOI: 10.1016/j.ipm.2009.03.002.
- [2] L. M. L. de Lau ja M. M. B. Breteler, „Epidemiology of Parkinson’s disease,“ *The Lancet Neurology*, köide 5, nr 6, lk. 525–535, 2006. DOI: 10.1016/S1474-4422(06)70471-9.
- [3] V. L. Feigin, E. Nichols, T. Alam *et al.*, „Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016,“ *The Lancet Neurology*, köide 18, nr 5, lk. 459–480, 2019.
- [4] E. Tolosa, A. Garrido, S. W. Scholz ja W. Poewe, „Challenges in the diagnosis of Parkinson’s disease,“ *The Lancet Neurology*, köide 20, nr 5, lk. 385–397, 2021. DOI: 10.1016/S1474-4422(21)00030-2.
- [5] L. Parnetti, L. Gaetani, P. Eusebi *et al.*, „CSF and blood biomarkers for Parkinson’s disease,“ *The Lancet Neurology*, köide 18, nr 6, lk. 573–586, 2019. DOI: [https://doi.org/10.1016/S1474-4422\(19\)30024-9](https://doi.org/10.1016/S1474-4422(19)30024-9).
- [6] W. Xu, L. Chen, G. Cai *et al.*, „Diagnosis of Parkinson’s Disease via the Metabolic Fingerprint in Saliva by Deep Learning,“ *Small Methods*, köide 7, 2023. DOI: 10.1002/smt.202300285.
- [7] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal ja M. Faundez-Zanuy, „Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson’s disease,“ *Artificial Intelligence in Medicine*, köide 67, lk. 39–46, 2016. DOI: 10.1016/j.artmed.2016.01.004.
- [8] P. Drotar, J. Mekyska, I. Rektorova, L. Masarova, Z. Smekal ja M. Faundez-Zanuy, „A new modality for quantitative evaluation of Parkinson’s disease: In-air movement,“ teoses *13th IEEE international conference on bioinformatics and bioengineering*, IEEE, 2013, lk. 1–4.
- [9] S. Rosenblum, M. Samuel, S. Zlotnik, I. Erikh ja I. Schlesinger, „Handwriting as an objective tool for Parkinson’s disease diagnosis,“ *Journal of Neurology*, köide 260, nr 9, lk. 2357–2361, 2013. DOI: 10.1007/s00415-013-6996-x.
- [10] A. Ünlü, R. Brause ja K. Krakow, „Handwriting Analysis for Diagnosis and Prognosis of Parkinson’s Disease,“ teoses *Biological and Medical Data Analysis*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, lk. 441–450.

- [11] A. Mumuni ja F. Mumuni, „Data augmentation: A comprehensive survey of modern approaches,“ *Array*, köide 16, lk. 100258, 2022. DOI: <https://doi.org/10.1016/j.array.2022.100258>.
- [12] S. Nõmm, K. Bardõš, A. Toomela, K. Medijainen ja P. Taba, „Detailed Analysis of the Luria’s Alternating Series Tests for Parkinson’s Disease Diagnostics,“ teoses *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, lk. 1347–1352.
- [13] J. Westin, S. Ghiamati, M. Memedi *et al.*, „A new computer method for assessing drawing impairment in Parkinson’s disease,“ *Journal of Neuroscience Methods*, köide 190, nr 1, lk. 143–148, 2010. DOI: <https://doi.org/10.1016/j.jneumeth.2010.04.027>.
- [14] S. Nõmm, K. Bardõš, I. Mašarov, J. Kozhenkina, A. Toomela ja T. Toomsoo, „Recognition and Analysis of the Contours Drawn during the Poppelreuter’s Test,“ teoses *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2016, lk. 170–175. DOI: [10.1109/ICMLA.2016.0036](https://doi.org/10.1109/ICMLA.2016.0036).
- [15] K. Bardõš, „Luria vaheldutavate seeriaste tõlgendatavate anomaaliaste ja kinemaatiliste parameetrite analüüs Parkinsoni tõve modelleerimisel,“ M.S. thesis, School of Information Technologies, Tallinn, Estonia, 2018. aadress: <https://digikogu.taltech.ee/et/Item/90256117-b2ae-4c14-8b5b-5b00a4d98522>.
- [16] C. Taleb, M. Khachab, C. Mokbel ja L. Likforman-Sulem, „Feature selection for an improved Parkinson’s disease identification based on handwriting,“ teoses *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, 2017, lk. 52–56. DOI: [10.1109/ASAR.2017.8067759](https://doi.org/10.1109/ASAR.2017.8067759).
- [17] E. Valla, S. Nõmm, K. Medijainen, P. Taba ja A. Toomela, „Tremor-related feature engineering for machine learning based Parkinson’s disease diagnostics,“ *Bio-medical Signal Processing and Control*, köide 75, Art. no. 103551, 2022. DOI: [10.1016/j.bspc.2022.103551](https://doi.org/10.1016/j.bspc.2022.103551).
- [18] M. Shaban, „Deep Convolutional Neural Network for Parkinson’s Disease Based Handwriting Screening,“ teoses *2020 IEEE 17th International Symposium on Bio-medical Imaging Workshops (ISBI Workshops)*, 2020, lk. 1–4. DOI: [10.1109/ISBIWorkshops50223.2020.9153407](https://doi.org/10.1109/ISBIWorkshops50223.2020.9153407).
- [19] P. Zham, D. K. Kumar, P. Dabnichki, S. Poosapadi Arjunan ja S. Raghav, „Distinguishing Different Stages of Parkinson’s Disease Using Composite Index of Speed and Pen-Pressure of Sketching a Spiral,“ *Frontiers in Neurology*, köide 8, 2017. DOI: [10.3389/fneur.2017.00435](https://doi.org/10.3389/fneur.2017.00435).

- [20] S. Kalash, M. Ajaj, N. Abbas ja S. Taleb, „Automated Early Detection of Parkinson’s Disease Through Augmented Handwritten Patterns,” teoses *2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME)*, 2023, lk. 01–06. DOI: 10.1109/ICABME59496.2023.10293078.
- [21] S. Agrawal ja S. Sahu, „Image-based Parkinson disease detection using deep transfer learning and optimization algorithm,” *International Journal of Information Technology*, köide 16, lk. 871–879, 2024. DOI: 10.1007/s41870-023-01601-3.
- [22] T. B. Shiraj, M. Shafiul Islam Dipta ja N. Akhter, „A Hybrid Transfer Learning Approach to Detect Parkinson’s Disease,” teoses *2023 26th International Conference on Computer and Information Technology (ICCIT)*, 2023, lk. 1–6. DOI: 10.1109/ICCIT60459.2023.10441255.
- [23] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza *et al.*, „Generative adversarial nets,” *Advances in neural information processing systems*, köide 27, 2014.
- [24] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger ja H. Greenspan, „GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification,” *Neurocomputing*, köide 321, lk. 321–331, 2018. DOI: <https://doi.org/10.1016/j.neucom.2018.09.013>.
- [25] E. Dzotsenidze, E. Valla, S. Nömm, K. Medijainen, P. Taba ja A. Toomela, „Generative Adversarial Networks as a Data Augmentation Tool for CNN-Based Parkinson’s Disease Diagnostics,” *IFAC-PapersOnLine*, köide 55, nr 29, lk. 108–113, 2022. DOI: <https://doi.org/10.1016/j.ifacol.2022.10.240>.
- [26] Q. Dao, M. A. El-Yacoubi ja A.-S. Rigaud, „Detection of Alzheimer Disease on Online Handwriting Using 1D Convolutional Neural Network,” *IEEE Access*, köide 11, lk. 2148–2155, 2023. DOI: 10.1109/ACCESS.2022.3232396.
- [27] C. Taleb, L. Likforman-Sulem, C. Mokbel *et al.*, „Detection of Parkinson’s disease from handwriting using deep learning: a comparative study,” *Evolutionary Intelligence*, köide 16, lk. 1813–1824, 2023. DOI: 10.1007/s12065-020-00470-0.
- [28] F. J. Moreno-Barea, J. M. Jerez ja L. Franco, „Improving classification accuracy using data augmentation on small data sets,” *Expert Systems with Applications*, köide 161, lk. 113 696, 2020. DOI: <https://doi.org/10.1016/j.eswa.2020.113696>.
- [29] D. A. Dablain ja N. V. Chawla, „Towards Understanding How Data Augmentation Works with Imbalanced Data,” *arXiv preprint arXiv:2304.05895*, 2023. DOI: 10.48550/arXiv.2304.05895.

- [30] A. A. Khan, O. Chaudhari ja R. Chandra, „A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation,” *Expert Systems with Applications*, köide 244, lk. 122 778, 2024. DOI: <https://doi.org/10.1016/j.eswa.2023.122778>.
- [31] L. Perez ja J. Wang, „The Effectiveness of Data Augmentation in Image Classification using Deep Learning,” *arXiv preprint arXiv:1712.04621*, 2017. DOI: [10.48550/arXiv.1712.04621](https://doi.org/10.48550/arXiv.1712.04621).
- [32] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk ja F. Herrera, *Learning from Imbalanced Data Sets* (Studies in Computational Intelligence). Springer, 2018, köide 811. DOI: [10.1007/978-3-319-98074-4](https://doi.org/10.1007/978-3-319-98074-4).
- [33] E. Abel, M. Abel ja Ü. Kaasik, *Koolimatemaatika entsüklopeedia*. Tartu: Ilmamaa, 2001, lk. 8, 29–30, 126.
- [34] C. Shorten ja T. M. Khoshgoftaar, „A survey on Image Data Augmentation for Deep Learning,” *Journal of Big Data*, köide 6, nr 1, lk. 60, 2019. DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
- [35] B. K. Iwana ja S. Uchida, „An empirical survey of data augmentation for time series classification with neural networks,” *PLOS ONE*, köide 16, nr 7, e0254841, 2021. DOI: [10.1371/journal.pone.0254841](https://doi.org/10.1371/journal.pone.0254841).
- [36] N. V. Chawla, K. W. Bowyer, L. O. Hall ja W. P. Kegelmeyer, „SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, köide 16, lk. 321–357, 2002. DOI: [10.1613/jair.953](https://doi.org/10.1613/jair.953).
- [37] A. Sawicki ja S. Zieliński, „Augmentation of Segmented Motion Capture Data for Improving Generalization of Deep Neural Networks,” teoses *Computer Information Systems and Industrial Management*, Cham: Springer International Publishing, 2020, lk. 278–290.
- [38] J. Yeomans, S. Thwaites, W. S. P. Robertson, D. Booth, B. Ng ja D. Thewlis, „Simulating Time-Series Data for Improved Deep Neural Network Performance,” *IEEE Access*, köide 7, lk. 131 248–131 255, 2019. DOI: [10.1109/ACCESS.2019.2940701](https://doi.org/10.1109/ACCESS.2019.2940701).
- [39] B. Iwana ja S. Uchida, „Time Series Data Augmentation for Neural Networks by Time Warping with a Discriminative Teacher,” teoses *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2021, lk. 3558–3565. DOI: [10.1109/ICPR48806.2021.9412812](https://doi.org/10.1109/ICPR48806.2021.9412812).
- [40] H. Sakoe ja S. Chiba, „Dynamic Programming Algorithm Optimization for Spoken Word Recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, köide 26, nr 1, lk. 43–49, 1978. DOI: [10.1109/TASSP.1978.1163055](https://doi.org/10.1109/TASSP.1978.1163055).

- [41] N. Takahashi, M. Gygli, B. Pfister ja L. Van Gool, „Deep Convolutional Neural Networks and Data Augmentation for Acoustic Event Recognition,” teoses *Proceedings of Interspeech*, 2016, lk. 2982–2986. DOI: 10.21437/Interspeech.2016-805.
- [42] X. Cui, V. Goel ja B. Kingsbury, „Data Augmentation for Deep Neural Network Acoustic Modeling,” teoses *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, köide 23, 2014, lk. 5582–5586. DOI: 10.1109/ICASSP.2014.6854671.
- [43] K. Kamycki, T. Kapuscinski ja M. Oszust, „Data Augmentation with Suboptimal Warping for Time-Series Classification,” *Sensors*, köide 20, nr 1, 2020. DOI: 10.3390/s20010098.
- [44] F. Petitjean, A. Ketterlin ja P. Gançarski, „A global averaging method for dynamic time warping, with applications to clustering,” *Pattern Recognition*, köide 44, nr 3, lk. 678–693, 2011. DOI: <https://doi.org/10.1016/j.patcog.2010.09.013>.
- [45] S. Smyl ja K. Kuber, „Data Preprocessing and Augmentation for Multiple Short Time Series Forecasting with Recurrent Neural Networks,” teoses *Proceedings of the International Symposium on Forecasting (ISF)*, 2016.
- [46] Y. Kang, R. J. Hyndman ja F. Li, „GRATIS: GeneRAtIng TIme Series with diverse and controllable characteristics,” *Statistical Analysis and Data Mining: An ASA Data Science Journal*, köide 13, nr 4, lk. 354–376, 2020. DOI: <https://doi.org/10.1002/sam.11461>.
- [47] N. Huang, Z. Shen, S. Long *et al.*, „The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, köide 454, lk. 903–995, 1998. DOI: 10.1098/rspa.1998.0193.
- [48] P. Comon, „Independent component analysis, A new concept?” *Signal Processing*, köide 36, nr 3, lk. 287–314, 1994, Higher Order Statistics. DOI: [https://doi.org/10.1016/0165-1684\(94\)90029-9](https://doi.org/10.1016/0165-1684(94)90029-9).
- [49] R. B. Cleveland, W. S. Cleveland, J. E. McRae ja I. Terpenning, „STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion),” *Journal of Official Statistics*, köide 6, lk. 3–73, 1990.
- [50] H. Han, W.-Y. Wang ja B.-H. Mao, „Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning,” teoses *Proceedings of the 2005 International Conference on Intelligent Computing (ICIC'05)*, seeria Lecture Notes in Computer Science, köide 3644, Springer, 2005, lk. 878–887. DOI: 10.1007/11538059_91.

- [51] H. He, Y. Bai, E. A. Garcia ja S. Li, „ADASYN: Adaptive synthetic sampling approach for imbalanced learning,“ teoses *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, lk. 1322–1328. DOI: 10.1109/IJCNN.2008.4633969.
- [52] A. Fernández, S. Garcia, F. Herrera ja N. Chawla, „SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary,“ *Journal of Artificial Intelligence Research*, köide 61, lk. 863–905, 2018. DOI: 10.1613/jair.1.11192.
- [53] L. Wang ja D. Rajan, „An image similarity descriptor for classification tasks,“ *Journal of Visual Communication and Image Representation*, köide 71, lk. 102 847, 2020. DOI: <https://doi.org/10.1016/j.jvcir.2020.102847>.
- [54] J. Zujovic, T. N. Pappas ja D. L. Neuhoff, „Structural Texture Similarity Metrics for Image Analysis and Retrieval,“ *IEEE Transactions on Image Processing*, köide 22, nr 7, lk. 2545–2558, 2013. DOI: 10.1109/TIP.2013.2251645.
- [55] B. Girod, „Psychovisual Aspects Of Image Processing: What’s Wrong With Mean Squared Error?“ Teoses *Proceedings of the Seventh Workshop on Multidimensional Signal Processing*, 1991, P.2–P.2. DOI: 10.1109/MDSP.1991.639240.
- [56] Z. Wang, G. Feng, L. Shen ja X. Zhang, „Cover Selection for Steganography Using Image Similarity,“ *IEEE Transactions on Dependable and Secure Computing*, köide 20, nr 3, lk. 2328–2340, 2023. DOI: 10.1109/TDSC.2022.3181039.
- [57] Z. Wang, A. Bovik, H. Sheikh ja E. Simoncelli, „Image Quality Assessment: From Error Visibility to Structural Similarity,“ *Image Processing, IEEE Transactions on*, köide 13, lk. 600–612, 2004. DOI: 10.1109/TIP.2003.819861.
- [58] Z. Wang ja E. Simoncelli, „Translation insensitive image similarity in complex wavelet domain,“ teoses *Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, köide 2, 2005, ii/573–ii/576 Vol. 2. DOI: 10.1109/ICASSP.2005.1415469.
- [59] OpenCV, *Contour Features in OpenCV (Python)*, OpenCV 3.4 Documentation, Accessed: October 16, 2024, 2024. aadress: https://docs.opencv.org/3.4/d5/d45/tutorial_py_contours_more_functions.html.
- [60] D. Lowe, „Object recognition from local scale-invariant features,“ teoses *Proceedings of the Seventh IEEE International Conference on Computer Vision*, köide 2, 1999, 1150–1157 vol.2. DOI: 10.1109/ICCV.1999.790410.
- [61] D. Lowe, „Distinctive Image Features from Scale-Invariant Keypoints,“ *International Journal of Computer Vision*, köide 60, lk. 91–110, 2004. DOI: 10.1023/B%3AVISI.0000029664.99615.94.

- [62] H. Bay, T. Tuytelaars ja L. Van Gool, „SURF: Speeded Up Robust Features,“ teoses *Computer Vision – ECCV 2006*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, lk. 404–417.
- [63] E. Rublee, V. Rabaud, K. Konolige ja G. Bradski, „ORB: An efficient alternative to SIFT or SURF,“ teoses *2011 International Conference on Computer Vision*, 2011, lk. 2564–2571. DOI: 10.1109/ICCV.2011.6126544.
- [64] N. Dalal ja B. Triggs, „Histograms of oriented gradients for human detection,“ teoses *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, köide 1, 2005, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.
- [65] E. Rosten ja T. Drummond, „Machine Learning for High-Speed Corner Detection,“ teoses *Computer Vision – ECCV 2006*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, lk. 430–443.
- [66] M. Calonder, V. Lepetit, C. Strecha ja P. Fua, „BRIEF: Binary Robust Independent Elementary Features,“ teoses *Computer Vision – ECCV 2010*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, lk. 778–792.
- [67] M. Li, X. Chen, X. Li, B. Ma ja P. M. B. Vitányi, „The similarity metric,“ *IEEE Transactions on Information Theory*, köide 50, nr 12, lk. 3250–3264, 2004. DOI: 10.1109/TIT.2004.838101.
- [68] T. Watanabe, K. Sugawara ja H. Sugihara, „A new pattern representation scheme using data compression,“ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, köide 24, nr 5, lk. 579–590, 2002. DOI: 10.1109/34.1000234.
- [69] G. Koch, R. Zemel ja R. Salakhutdinov, „Siamese Neural Networks for One-shot Image Recognition,“ teoses *Proceedings of the 32nd International Conference on Machine Learning (ICML) Deep Learning Workshop*, 2015. aadress: <https://www.cs.toronto.edu/~rsalakhu/papers/kochmetzgerzemel.pdf>.
- [70] J. Lee. „Image Similarity using VGG16, Transfer Learning and Cosine Similarity.“ (2020), aadress: <https://medium.com/@jeff.lee.1990710/image-similarity-using-vgg16-transfer-learning-and-cosine-similarity-98571d8055e3> (vaadatud 03.05.2024).
- [71] J. Serrà ja J. L. Arcos, „An empirical evaluation of similarity measures for time series classification,“ *Knowledge-Based Systems*, köide 67, lk. 305–314, 2014. DOI: <https://doi.org/10.1016/j.knosys.2014.04.035>.

- [72] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann ja E. Keogh, „Experimental Comparison of Representation Methods and Distance Measures for Time Series Data,“ *Data Mining and Knowledge Discovery*, köide 26, nr 2, lk. 275–309, 2013. DOI: 10.1007/s10618-012-0250-5.
- [73] M. Müller, *Information Retrieval for Music and Motion*. Springer, 2007. DOI: 10.1007/978-3-540-74048-3.
- [74] GeeksforGeeks, *Dynamic Time Warping (DTW) in Time Series*, Accessed: May 4, 2025, 2025. address: <https://www.geeksforgeeks.org/dynamic-time-warping-dtw-in-time-series/>.
- [75] X. Xi, E. Keogh, C. Shelton, L. Wei ja C. A. Ratanamahatana, „Fast Time Series Classification Using Numerosity Reduction,“ teoses *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, ACM, 2006, lk. 1033–1040. DOI: 10.1145/1143844.1143974.
- [76] Indian Institute of Technology Madras, *CS6464: Concepts in Statistical Learning Theory*, http://www.cse.iitm.ac.in/~vplab/statistical_learning_theory.html, 2023. address: http://www.cse.iitm.ac.in/~vplab/statistical_learning_theory.html.
- [77] O. Theobald, *Machine Learning for Absolute Beginners*. London, UK: Scatterplot Press, 2017.
- [78] N. H. Ali, M. Emaduldeen ja A. E. Ali, „Learning Evolution: a Survey,“ *Iraqi Journal of Science*, köide 62, nr 12, lk. 4978–4987, 2021.
- [79] A. F. A. H. Alnuaimi ja T. H. K. Albaldawi, „An overview of machine learning classification techniques,“ *BIO Web Conf.*, köide 97, lk. 00 133, 2024. DOI: 10.1051/bioconf/20249700133.
- [80] J. P. Mueller ja L. Massaron, *Machine Learning For Dummies*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2016.
- [81] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, 2. väljaanne. Elsevier Ltd, 2020.
- [82] R. Praba, D. G. R. T ja S. B., „Study On Machine Learning Algorithms,“ *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, lk. 67–72, 2021. DOI: 10.32628/CSEIT2173105.
- [83] P. Xia, L. Zhang ja F. Li, „Learning similarity with cosine similarity ensemble,“ *Information Sciences*, köide 307, lk. 39–52, 2015. DOI: <https://doi.org/10.1016/j.ins.2015.02.024>.
- [84] S. Salvador ja P. Chan, „Toward accurate dynamic time warping in linear time and space,“ teoses *Intelligent Data Analysis*, köide 11, Springer, 2007, lk. 561–580.

- [85] J. Wainer ja G. Cawley, „Nested cross-validation when selecting classifiers is overzealous for most practical applications,“ *Expert Systems with Applications*, köide 182, lk. 115 222, 2021. DOI: <https://doi.org/10.1016/j.eswa.2021.115222>.
- [86] The imbalanced-learn developers, *imbalanced-learn: SMOTE implementation in Python*, Accessed: May 5, 2025, 2024. address: %7Bhttps://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html%7D.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

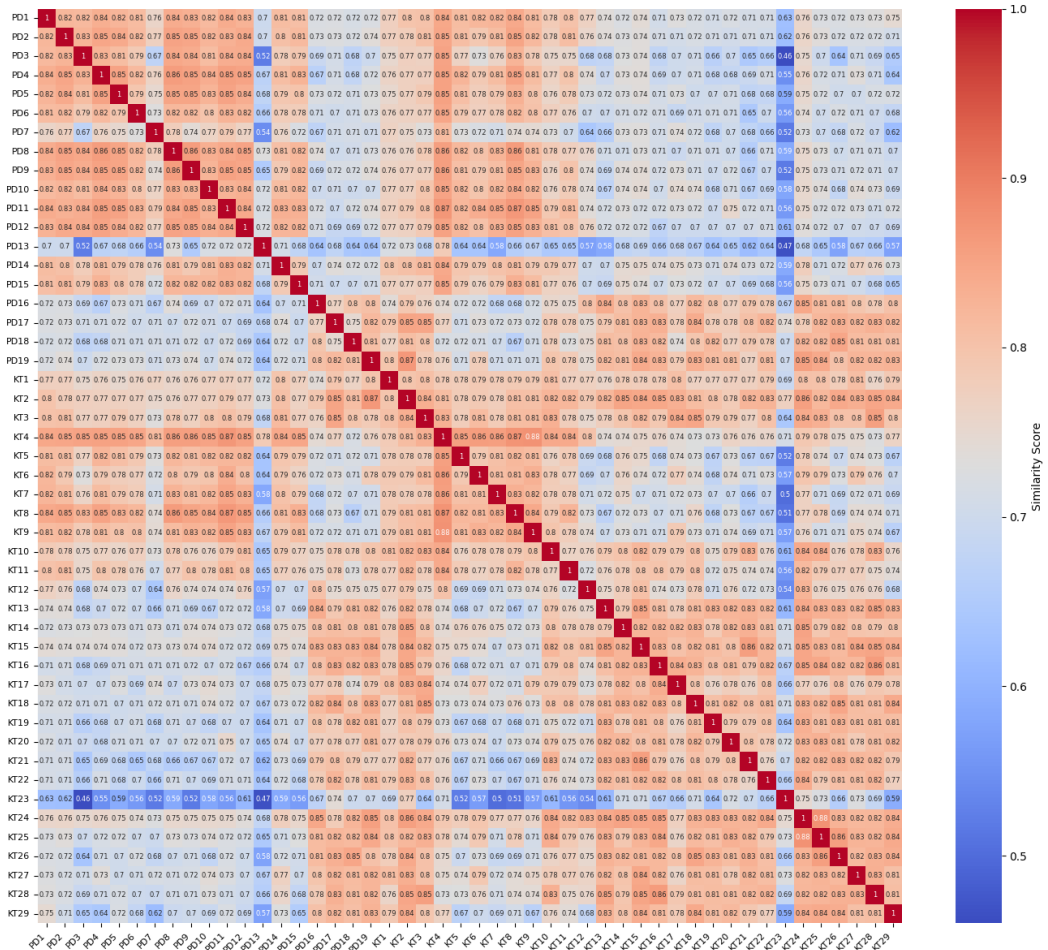
Ma, Ida Tuule Jõgi

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Parkinsoni diagnoosimiseks kasutatavate Archimedese spiraalide hulga suurendamise analüüs”, mille juhendaja on Sven Nõmm
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

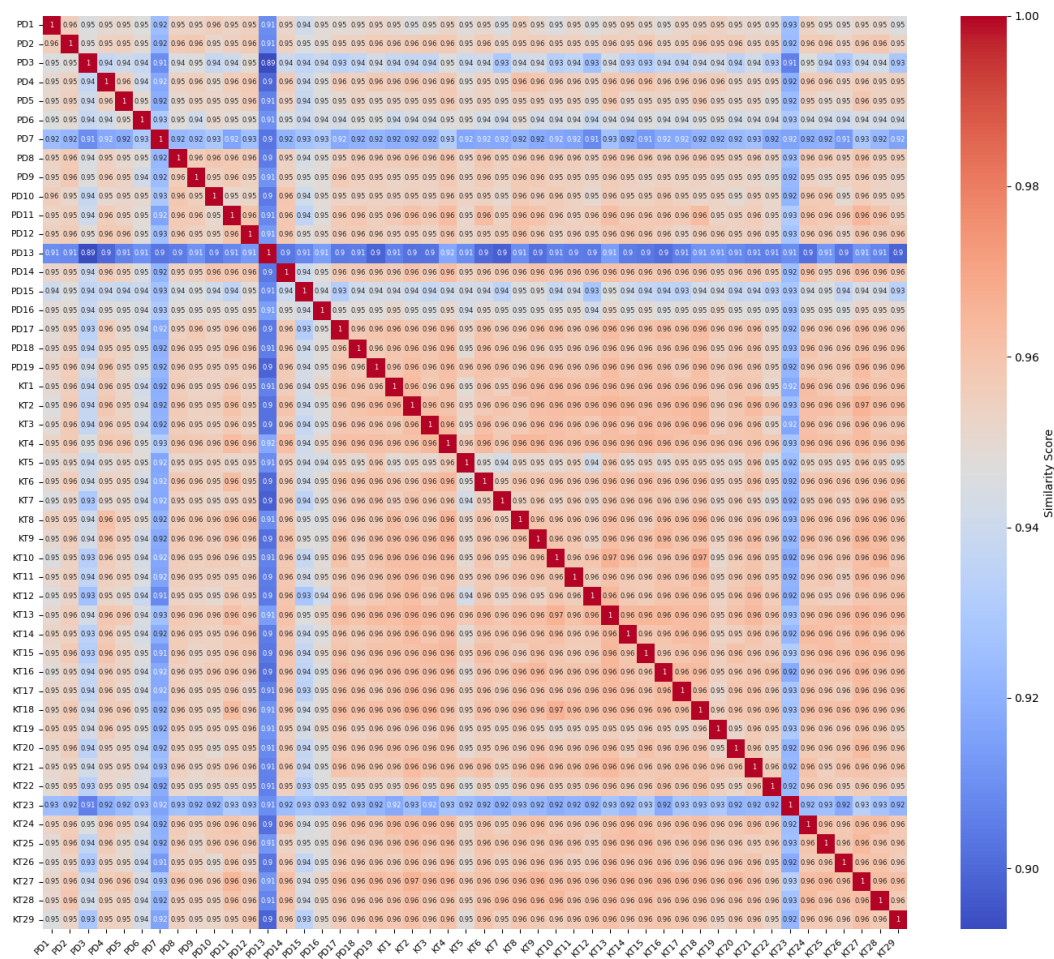
01.06.2025

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

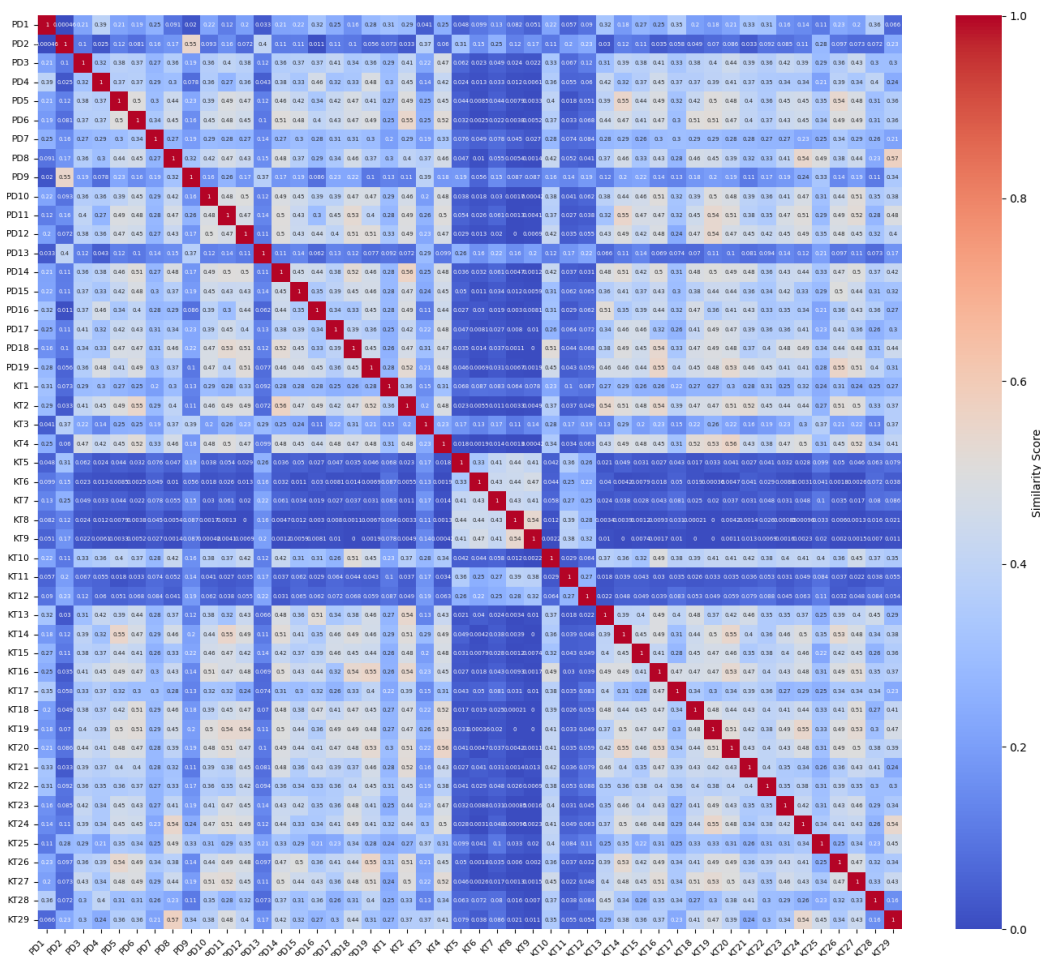
Lisa 2 – Spiraalide sarnasust hindavate meetodite soojus- kaardid



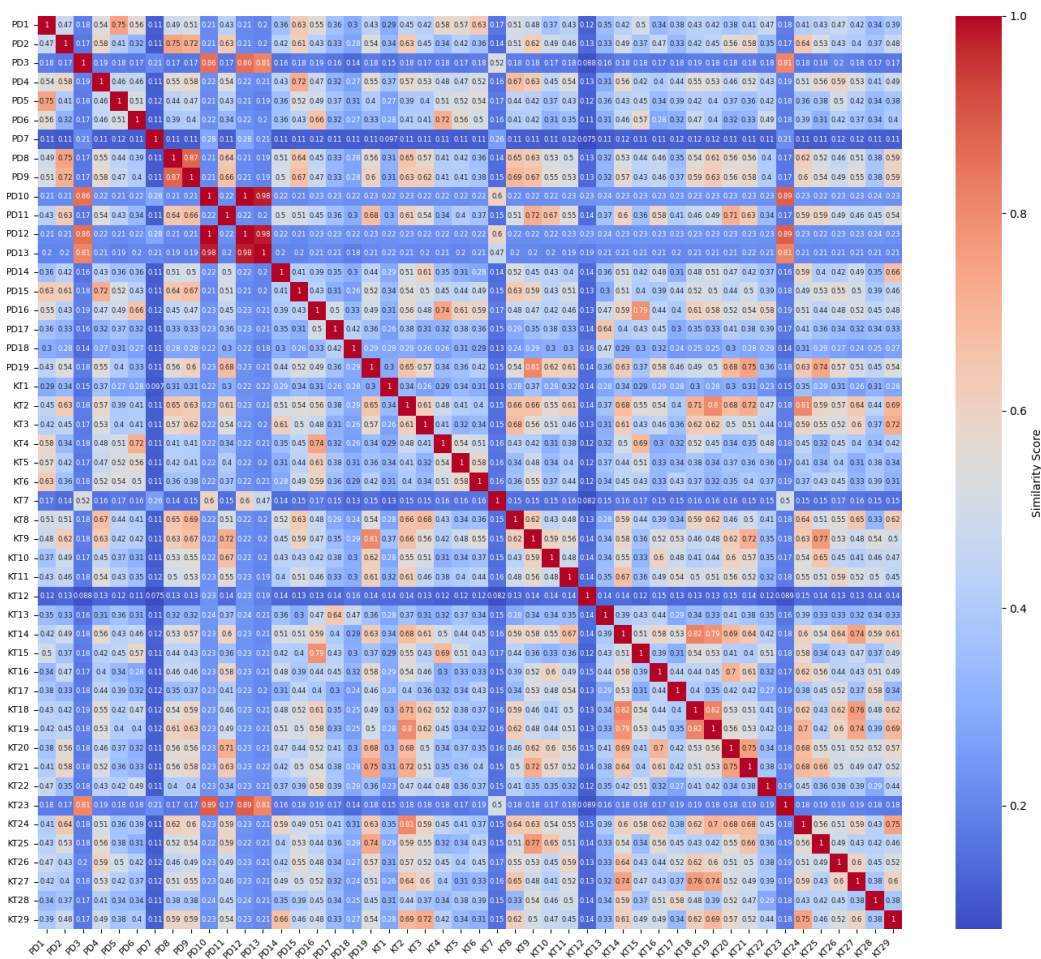
Joonis 6. Spiraalidevaheline sarnasus SIFT-meetodi põhjal.



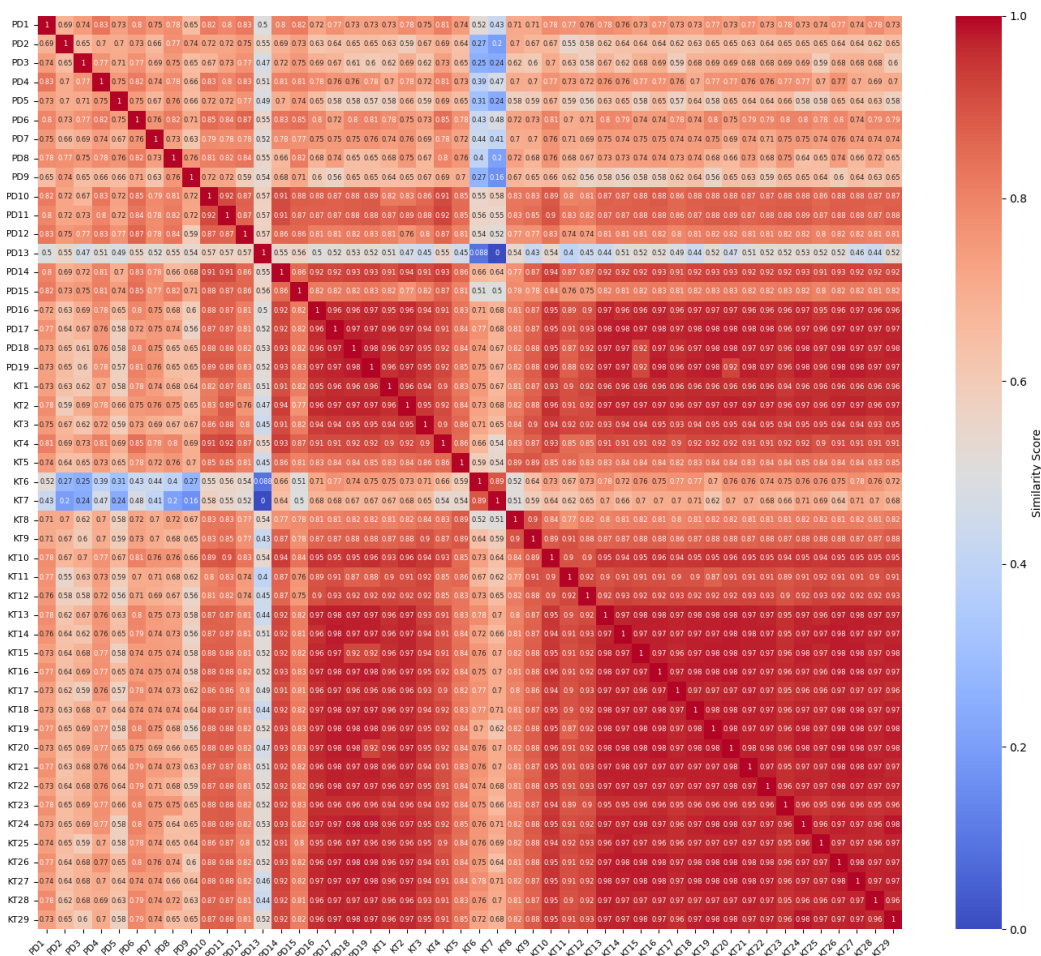
Joonis 7. Spiraalidevaheline sarnasus ORB-meetodi põhjal.



Joonis 8. Spiraalidevaheline sarnasus HOG-meetodi põhjal.



Joonis 9. Spiraalidevaheline sarnasus OpenCV matchShapes() meetodi põhjal.



Joonis 10. Spiraalidevaheline sarnasus DTW-meetodi põhjal.

Lisa 3 – Klassifitseerimiseks kasutatud spiraalide tunnuste valemid

Tabel 4. Andmehulkade põhiseks klassifitseerimiseks kasutatud tunnuste arvutamiseks vajalikud kinemaatilised ja geomeetrilised tunnused [17].

Tunnus	Tunnuse selgitus	Tunnuse valem
<i>Slope</i>	Tõus	$k = \frac{y_i - y_{i-1}}{x_i - x_{i-1}}$, kus (x_i, y_i) ja (x_{i-1}, y_{i-1}) on vastavalt hetke- ja eelneva ajahetke koordinaadid
α	Tõusunurk	$\alpha = \arctan(k)$
ϕ_i	Pöördenurk	$\phi_i = \pi + \alpha_{i-1} - \alpha_i$
$\alpha_velocity$	Tõusunurga muutumise kiirus ehk nurkkiirus	$\omega_i = \frac{\alpha_i - \alpha_{i-1}}{\Delta t}$
α_accel	Tõusunurga kiirendus (nurga 2. tuletis)	$\alpha''_i = \frac{\omega_i - \omega_{i-1}}{\Delta t}$
<i>Displacement</i>	Siire	$d_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$
<i>Velocity</i>	Kiirus	$v_i = \frac{d_i}{t_i - t_{i-1}}$
<i>Acceleration</i>	Kiirendus	$a_i = \frac{v_i - v_{i-1}}{t_i - t_{i-1}}$
<i>Velocity_x</i>	Kiirus X-suunas	$v_x = \frac{x_i - x_{i-1}}{\Delta t}$
<i>Accel_x</i>	Kiirendus X-suunas	$a_x = \frac{v_{x,i} - v_{x,i-1}}{\Delta t}$
<i>Jerk</i>	Kiirenduse muutumise kiirus aja suhtes	$j(t) = \frac{d^3}{dt^3}x(t)$
<i>Snap</i>	<i>Jerk</i> 'i muutumise kiirus aja suhtes	$s(t) = \frac{d^4}{dt^4}x(t)$
<i>Crackle</i>	<i>Snap</i> 'i muutumise kiirus aja suhtes	$c(t) = \frac{d^5}{dt^5}x(t)$
<i>Pop</i>	<i>Crackle</i> 'i muutumise kiirus aja suhtes	$p(t) = \frac{d^6}{dt^6}x(t)$

Tabel 5. Andmehulkade põhiseks klassifitseerimiseks kasutatud tunnuste arvutamiseks vajalikud surge tunnused [17].

Tunnus	Tunnuse selgitus	Tunnuse valem
<i>Pressure_diff</i>	Surve erinevus punktide p_i ja p_{i+1} vahel	$\Delta p = p_{i+1} - p_i$
<i>Yank</i>	Surve muutumise kiirus aja suhtes (1. tuletis)	$y(t) = \frac{d}{dt}F(t)$
<i>Tug</i>	<i>Yank</i> 'i muutumise kiirus aja suhtes (2. tuletis)	$t(t) = \frac{d^2}{dt^2}F(t)$
<i>Snatch</i>	<i>Tug</i> 'i muutumise kiirus aja suhtes (3. tuletis)	$sn(t) = \frac{d^3}{dt^3}F(t)$
<i>Shake</i>	<i>Snatch</i> 'i muutumise kiirus aja suhtes (4. tuletis)	$sh(t) = \frac{d^4}{dt^4}F(t)$

Tabel 6. Andmehulkade põhiseks klassifitseerimiseks kasutatud kinemaatilised, geomeetrised ja ajalised tunnused [17].

Tunnus	Tunnuse selgitus	Tunnuse valem
<i>Duration</i>	Ajavahemik esimese ja viimase spiraali ajatempli vahel	$D = t_n - t_0$, kus t_0 on esimene ja t_n viimane ajatempel
ϕ_{mass}	Spiraali pöördenurkade kogusumma (radiaanides)	$\phi_N = \sum_{k=1}^N \phi_k $
<i>Velocity_median</i>	Kiiruse mediaan	$v_{med} = \text{median}(v_i)$
α_{accel_max}	Tõusunurga maksimaalne kiirendus	$\alpha_{accel_max} = \max(\alpha''_i)$
<i>Snap_mass</i>	Siirde neljanda ajatuletise mass	$s(t)_{mass} = \sum_{k=1}^N s(t)_k $
<i>Crackle_mass</i>	Siirde viienda ajatuletise mass	$c(t)_{mass} = \sum_{k=1}^N c(t)_k $
<i>Pop_mass</i>	Siirde kuuenda ajatuletise mass	$p(t)_{mass} = \sum_{k=1}^N p(t)_k $
<i>Accel_x_mass</i>	X-suunalise kiiruse muutumise määra mass	$a_{x_mass} = \sum_{k=1}^N a_{x,k} $
<i>Accel_x_min</i>	X-suunalise kiiruse muutumise määra mass	$a_{x_min} = \min(a_x)$
α_{accel_min}	Tõusunurga minimaalne kiirendus	$\alpha_{accel_min} = \min(\alpha''_i)$
$\alpha_{velocity_max}$	Tõusunurga maksimaalne kiirus	$\alpha_{v_max} = \max(\omega_i)$

Tabel 7. Andmehulkade põhiseks klassifitseerimiseks kasutatud surge tunnuste [17].

Tunnus	Tunnuse selgitus	Tunnuse valem
<i>Pressure_median</i>	Survesignaali me- diaan	$p_{\text{med}} = \text{median}(p)$
<i>Shake_mass</i>	Pinnale rakendatud jõu neljanda ajatuletise mass	$sh(t)_{\text{mass}} = \sum_{k=1}^N sh(t)_k $
<i>Shake_max</i>	Pinnale rakendatud jõu neljanda ajatuletise maksimaalne väärtus	$sh(t)_{\text{max}} = \max(sh(t))$
<i>Shake_mean</i>	Pinnale rakendatud jõu neljanda ajatuletise keskmine väärtus	$sh(t)_{\text{mean}} = \frac{1}{N} \sum_{k=1}^N sh(t)_k$
<i>Shake_median</i>	Pinnale rakendatud jõu neljanda ajatuletise me- diaan	$sh(t)_{\text{med}} = \text{median}(sh(t))$
<i>Pressure_diff_- min</i>	Surve minimaalne muu- tus punktide $[p_i, p_{i+1}]$ vahel	$\Delta p_{\text{min}} = \min(\Delta p)$
<i>Pressure_diff_- max</i>	Surve maksimaal- ne muutus punktide $[p_i, p_{i+1}]$ vahel	$\Delta p_{\text{max}} = \max(\Delta p)$
<i>Snatch_mean</i>	Pinnale rakendatud jõu kolmanda ajatuletise keskmine väärtus	$sn(t)_{\text{mean}} = \frac{1}{N} \sum_{k=1}^N sn(t)_k$

Lisa 4 – Spiraalide klassidesisesed ja -vahelised keskmised sarnasusskoorid arvutatud Dynamic Time Warpingu'ga

Tabel 8. Esialgsete archimedese spiraalide keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

Sarnasus	Keskmine	Mediaan	Standardhälve
PD vs PD	1331655	1317178	639748
PD vs KT	1415304	1376639	619654
KT vs KT	148515	144505	52608

Tabel 9. Esialgsete archimedese spiraalide keskmine sarnasus, arvutatuna spiraalide aegridade punktide eukleidilise kaugusena.

Sarnasus	Keskmine	Mediaan	Standardhälve
PD vs PD	160310	137964	152704
PD vs KT	168473	83643	212726
KT vs KT	18040	18058	6555

Tabel 10. Tasakaalustatud archimedese spiraalide andmehulga keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

Sarnasus	Keskmine	Mediaan	Standardhälve
PD vs PD	1460072	1468272	639748
PD vs KT	1415304	1376639	619654
KT vs KT	148515	144505	52608

Tabel 11. Tasakaalustatud archimedese spiraalide andmehulga keskmine sarnasus, arvutatuna spiraalide aegridade punktide eukleidilise kaugusena.

Sarnasus	Keskmine	Mediaan	Standardhälve
PD vs PD	160310	137964	152704
PD vs KT	168473	83643	212726
KT vs KT	18040	18058	6555

Tabel 12. Roteeritud archimedese spiraalide keskmine sarnasus, arvatatuna spiraalide aegridade absoluutse kaugusena.

Roteerimis- nurk	PD vs PD ro- teeritud	PD vs KT ro- teeritud	KT vs PD ro- teeritud	KT vs KT ro- teeritud
15	8644395	5549400	7417350	2031349
30	8566912	5623455	6520916	2091569
45	6923383	4817549	5082689	1588617
60	4738851	3413322	3510543	1011144
75	2773521	2171561	2164411	537057
90	1333198	1420730	1402548	149386
105	3884785	2859709	3133147	819077
120	6897826	4184078	5173465	1607073
135	8560262	5669988	6515657	2085179
150	8417341	5528736	7464529	1967397
165	8724932	4786255	8185934	2039084
180	9792439	4261227	8710301	2410774
195	11349634	4528411	9429140	2877742
210	13262085	6238507	10543552	3341360
225	15270532	8855661	11939678	3749950
240	16895146	11393129	12986192	4046374
255	17911183	13519561	13513864	4203593
270	17914686	13508078	13486732	4195208
285	16836092	11378338	12906646	4015600
300	15190106	8859619	11813324	3703658
315	13079463	6334396	10347487	3286737
330	11134316	4608181	9119703	2814624
345	9551669	4243302	8384773	2323370

Tabel 13. Roteeritud archimedese spiraalide keskmine sarnasus, arvatatuna spiraalide aegridade eukleidilise kaugusena.

Roteerimis- nurk	PD vs PD ro- teeritud	PD vs KT ro- teeritud	KT vs PD ro- teeritud	KT vs KT ro- teeritud
15	6178357	4396894	4384396	1452560
30	4592866	3108736	3097248	1096361
45	3207846	2065449	2055351	774402
60	2282577	1535028	1535032	565433
75	1889628	1192080	1196871	504248
90	2400298	1628433	1635495	610318
105	3426980	2226813	2200772	853542
120	4883201	3325274	3278920	1193125
135	6461189	4600817	4554378	1546976
150	7912297	5804615	5781379	1874855
165	9234778	6930972	6905591	2170687
180	10401535	7901984	7877782	2429031
195	11385715	8751513	8724722	2645191
210	12168982	9439530	9423168	2815681
225	12731600	9930766	9917001	2937607
240	13081463	10216695	10202231	3008847
255	13163440	10302757	10177780	3028222
270	13024077	10191508	9863047	2995230
285	12653810	9875401	9863047	2910421
300	12064932	9363494	9350765	2775213
315	11252398	8677156	8665456	2591910
330	10233049	7821428	7810232	2363478
345	9030401	6786758	6775884	2094098

Tabel 14. Tasakaalustatud ja roteeritud archimedese spiraalide keskmine sarnasus, arvutuna spiraalide aegridade absoluutse kaugusena.

Roteerimis- nurk	PD vs PD ro- teeritud	PD vs KT ro- teeritud	KT vs PD ro- teeritud	KT vs KT ro- teeritud
15	8514378	5549719	7319648	2031539
30	8427017	5623670	6449410	2091524
45	6830871	4818219	5016776	1588703
60	4655772	3415978	3464536	1011108
75	2756015	2172164	216410	537224
90	1412817	1420873	1413106	149336
105	3844936	2859108	3105064	818470
120	6815948	4183342	5104893	1606285
135	8443957	5666965	6443556	2084229
150	8282597	5528119	7372027	1966455
165	8623515	4785416	8084738	2037688
180	9685894	4260101	8590539	2409387
195	11234422	4527082	9283829	2876436
210	13167640	6237237	10394485	3340077
225	15134088	8854635	11775288	3748776
240	16729211	11392416	12804384	4045352
255	17721217	13519149	13323411	4202746
270	17694596	13507772	13296999	4194458
285	16634021	11377874	12726948	4014839
300	15032099	8858938	11652270	3702897
315	12988228	6333808	10209386	3286028
330	11066504	4607824	8995578	2814154
345	9440203	4243238	8283812	2323106

Tabel 15. Tasakaalustatud ja roteeritud archimedese spiraalide keskmine sarnasus, arvutatuna spiraalide aegridade eukleidilise kaugusena.

Roteerimis- nurk	PD vs PD ro- teeritud	PD vs KT ro- teeritud	KT vs PD ro- teeritud	KT vs KT ro- teeritud
15	6149338	4397059	4336287	1452518
30	4586849	3108721	3070703	1096318
45	3201634	2065432	2033293	774298
60	2269394	1534943	1515904	565255
75	1876769	1191822	1185949	503992
90	2376973	1628130	1610699	609990
105	3407558	2226463	2169036	853146
120	4852586	3324908	3242281	1192669
135	6408659	4600461	4498801	1546470
150	7842580	5804279	5707146	1874309
165	9149018	6930663	6813458	2170109
180	10301054	7901701	7770013	2428431
195	11269907	8751260	8603582	2644579
210	12038711	9439307	9290145	2815067
225	12590131	9930573	9775754	2937001
240	12930461	10216535	10056350	3008260
255	13009146	10302632	10141033	3027663
270	12870763	10191418	10032198	2994709
285	12504647	9875352	9721779	2909946
300	11918749	9363492	9217429	2774796
315	11118144	8677226	8544080	2591561
330	10117665	7821572	7702666	2363208
345	8940769	6786960	6686058	2093909

Tabel 16. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05 ja y -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05, saadud keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.10	0.05	8058377	6981339	4734881	1920862
0.15	0.10	7532423	6416392	4329183	1811033
0.20	0.15	7004769	5850319	3923492	1701171
0.25	0.20	6479982	5295362	3517862	1591294
0.30	0.25	5967996	4748616	3114269	1481378
0.35	0.30	5469605	4259796	2741897	1372332
0.40	0.35	5002599	3788537	2417118	1264413
0.45	0.40	4581204	3381634	2160429	1158728
0.50	0.45	4148805	3036280	1990003	1030021
0.55	0.50	3792392	2727905	1861213	948333
0.60	0.55	3449198	2462515	1780885	836355
0.65	0.60	3173480	2247978	1775498	769864
0.70	0.65	3000048	2133026	1856402	714587
0.75	0.70	2779837	1955440	1938666	663831
0.80	0.75	2554317	1836285	2014228	592436
0.85	0.80	2324353	1720146	1856402	517949
0.90	0.85	2125956	1636008	1816568	451077
0.95	0.90	1871895	1556896	1626286	356435
1.00	0.95	1556315	1495070	1515506	220563
1.05	1.00	1680736	1587014	1550839	257382
1.10	1.05	2076011	1751117	1716479	404359
1.15	1.10	2393745	1976748	1885227	512058
1.20	1.15	2674982	2272373	2036866	599045
1.25	1.20	2972411	2491806	2232846	678818
1.30	1.25	3289031	2543974	2439805	770063
1.35	1.30	3586341	2596513	2614311	856287
1.40	1.25	3881918	2628135	2813300	929443
1.45	1.30	4172617	2640581	3022087	1002200
1.50	1.45	4447064	2689278	3251939	1058951
1.55	1.50	4753071	2723139	3461663	1140585

Tabel 17. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05 ja y -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05, saadud keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

X- skaleerimis- faktor	Y- skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.60	1.55	5009528	2756980	3652138	1220413
1.65	1.60	5331801	2814875	3872098	1290990
1.70	1.65	5651222	2915214	4127360	1365010
1.75	1.70	6015052	3021118	4395846	1468108
1.80	1.75	6391831	3145559	4678470	1582569
1.85	1.80	6770339	3274013	4970615	1688508
1.90	1.90	7134808	3406534	5268004	1783206
1.95	1.95	7502415	3558027	5570864	1871707
2.00	2.00	7877591	3712512	5881430	1953613
2.05	2.05	8265440	3870432	6198257	2043654
2.10	2.10	8663494	4034175	6520742	2145126
2.15	2.15	9064935	4206386	6845874	2261549
2.20	2.20	9491866	4383795	7175127	2378771
2.25	2.25	9905594	4565486	7501542	2493243
2.30	2.30	10309179	4750375	7838099	2598417
2.35	2.35	10714792	4942384	8182256	2701292
2.40	2.40	11122487	5141742	8530916	2804340
2.45	2.45	11533492	5348024	8884273	2907556
2.50	2.50	11945887	5561158	9242944	3010945
2.55	2.55	12358521	5779917	9602977	3114511
2.60	2.60	12779307	6004339	9965977	3218248
2.65	2.65	13204330	6234287	10335761	3322100
2.70	2.70	13631444	6469187	10709900	3426096
2.75	2.75	14060428	6708045	11085136	3530167
2.80	2.80	14491313	6950219	11462997	3634331
2.85	2.85	14923895	7195495	11841937	3738557
2.95	2.90	15357885	7443865	12223216	3842897
3.00	2.95	15793194	7695053	12606227	3947283

Tabel 18. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05 ja y -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05, saadud keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.05	0.10	8164145	7191985	4793534	1939250
0.10	0.15	7659323	6721202	4387835	1831135
0.15	0.20	7155506	6253116	3982137	1722954
0.20	0.25	6653533	5771417	3576446	1614143
0.25	0.30	6146112	5272982	3170822	1504945
0.30	0.35	5642291	4750657	2768441	1395538
0.35	0.40	5152068	4245860	2416960	1287081
0.40	0.45	4706953	3771254	2148624	1180237
0.45	0.50	4291766	3346272	1946313	1065351
0.50	0.55	3944549	2986291	1807659	985954
0.55	0.60	3598955	2676485	1728518	884204
0.60	0.65	3311509	2441505	1713878	809622
0.65	0.70	3137181	2274222	1795889	756490
0.70	0.75	2905115	2059333	1881901	695564
0.75	0.80	2636510	1907483	1960685	619486
0.80	0.85	2366644	1743696	1999077	531158
0.85	0.90	2101801	1620158	1802963	452766
0.90	0.95	1845820	1518848	1586995	360248
0.95	1.00	1541415	1433595	1465269	233439
1.00	1.05	1474997	1465592	1406361	202599
1.05	1.10	1828424	1633509	1538426	335598
1.10	1.15	2136347	1836934	1684241	443410
1.15	1.20	2454943	2135063	1865879	535835
1.20	1.25	2754631	2371941	2063812	622259
1.25	1.30	3116210	2444564	2292253	721663
1.30	1.35	3436214	2527821	2472727	813664
1.35	1.40	3766514	2575640	2669446	900121
1.40	1.45	4067843	2606014	2888723	986487
1.45	1.50	4376522	2663665	3129717	1056528
1.50	1.55	4692405	2708631	3367168	1132237

Tabel 19. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05 ja y -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05, saadud keskmine sarnasus, arvatatuna spiraalide aegridade absoluutse kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.55	1.60	4963362	2745240	3588177	1207273
1.60	1.65	5266001	2804590	3790713	1279852
1.65	1.70	5592269	2901024	4032927	1351356
1.70	1.75	5943495	3007869	4303193	1450865
1.75	1.80	6324159	3134709	4577560	1563114
1.80	1.85	6699678	3265942	4868247	1677036
1.85	1.90	7071329	3402421	5165938	1769322
1.90	1.95	7442235	3551642	5468985	1859155
1.95	2.00	7822566	3713483	5779339	1941410
2.00	2.05	8204922	3874017	6096145	2029610
2.05	2.10	8602500	4063868	6419671	2130632
2.10	2.15	8998066	4244050	6745531	2244239
2.15	2.20	9424794	4425577	7075775	2362789
2.20	2.25	9843681	4613426	7406484	2476786
2.25	2.30	10247218	4803831	7741415	2584008
2.30	2.35	10655003	5004906	8085612	2686942
2.35	2.40	11063083	5209644	8436077	2790034
2.40	2.45	11475232	5421028	8790006	2893282
2.45	2.50	11888406	5638209	9149534	2996705
2.50	2.55	12301046	5861124	9510543	3100297
2.55	2.60	12720726	6089711	9874059	3204079
2.60	2.65	13146452	6323260	10246325	3308031
2.65	2.70	13574052	6560962	10620788	3412092
2.70	2.75	14003993	6802005	10998552	3516234
2.75	2.80	14435391	7046169	11377888	3620481
2.80	2.85	14868420	7293386	11759150	3724804
2.85	2.90	15303005	7543547	12141797	3829194
2.90	2.95	15738874	7796187	12526195	3933665
2.95	3.00	16175962	8051009	12911060	4038191

Tabel 20. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05 ja y -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05, saadud keskmine sarnasus, arvutatuna spiraalide aegridade eukleidilise kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.10	0.05	6068409	5324575	3996608	1411694
0.15	0.10	5716454	4970533	3711883	1333553
0.20	0.15	5355959	4621316	3424329	1255539
0.25	0.20	4997057	4264324	3149655	1177579
0.30	0.25	4934458	3895897	2898297	1099693
0.35	0.30	4287525	2680289	2068002	1021683
0.40	0.35	3964195	2337119	1739159	943773
0.45	0.40	3950040	2081199	1482297	868841
0.50	0.45	3245697	2586773	1822497	787696
0.55	0.50	2912596	1625376	1625376	708965
0.60	0.55	2578015	2036592	1432511	627609
0.65	0.60	2244366	1278255	1278255	544201
0.70	0.65	1906186	1154811	1154811	459988
0.75	0.70	1572355	765244	935476	376488
0.80	0.75	1254886	396663	376488	295427
0.85	0.80	956175	161652	217892	211782
0.90	0.85	685143	90036	146612	146612
0.95	0.90	411530	50543	90538	83496
1.00	0.95	223325	30105	46304	46304
1.05	1.00	223073	22734	36077	36077
1.10	1.05	205007	19822	46304	46304
1.15	1.10	456455	388394	98768	98768
1.20	1.15	712495	611694	162155	162155
1.25	1.20	988850	816872	231286	231286
1.30	1.25	1191852	1053941	309793	309793
1.35	1.30	1509982	1396801	380047	380047
1.40	1.35	2237980	1856658	457654	457654
1.45	1.40	2561058	1607481	536658	536658
1.50	1.45	2892467	1741561	616823	616823
1.55	1.50	3217996	1881307	697059	697059

Tabel 21. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05 ja y -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05, saadud keskmine sarnasus, arvatatuna spiraalide aegridade eukleidilise kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.60	1.55	3545022	2022343	2808281	857431
1.65	1.60	3864285	2177721	3061764	937776
1.70	1.65	4186277	2340308	3304683	1017705
1.75	1.70	4511876	2511319	3557319	1097725
1.80	1.75	4838389	2691891	3794670	1176076
1.85	1.80	5165201	2876328	4038268	1252646
1.90	1.85	5479924	3069368	4290092	1343865
1.95	1.90	5804400	3268706	4545028	1410734
2.00	1.95	6121454	3467201	4804603	1483681
2.05	2.00	6451488	3669252	5067751	1565769
2.10	2.05	6770933	3878025	5335909	1643005
2.15	2.10	7102499	4080250	5607317	1720147
2.20	2.15	7429893	4280997	5882451	1797095
2.25	2.20	7758645	4501179	6160130	1874192
2.30	2.25	8087747	4712486	6439284	1951204
2.35	2.30	8414939	4928413	6721823	2025184
2.40	2.35	8743784	5142638	7005603	2102514
2.45	2.40	9077733	5359985	7290963	2182072
2.50	2.45	9408285	5574488	7577373	2259963
2.55	2.50	9732301	5794227	7864790	2356128
2.60	2.55	10063348	6013846	8152952	2451004
2.65	2.60	10395668	6231503	8441724	2548990
2.70	2.65	10726965	6450360	8730795	2631862
2.75	2.70	11057948	6688191	9019897	2720740
2.80	2.75	11389286	6889848	9308612	2797678
2.85	2.80	11724456	7118803	9598739	2874162
2.90	2.85	12055356	7328536	9888963	2951530
2.95	2.90	12387342	7549234	10188272	3028467
3.00	2.95	12716272	7770114	10472394	3105995

Tabel 22. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05 ja y -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05, saadud keskmine sarnasus, arvutatuna spiraalide aegridade eukleidilise kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.05	0.10	6177908	5423113	4077960	1433184
0.10	0.15	5816203	5068231	3792429	1355050
0.15	0.20	5455535	4725044	3511029	1276995
0.20	0.25	5099645	4373254	3235357	1198940
0.25	0.30	4741993	4011025	2961915	1121063
0.30	0.35	4386221	3664437	2686682	1043205
0.35	0.40	4031983	3321863	2418040	963425
0.40	0.45	3680958	2988268	2157624	887053
0.45	0.50	3335485	2686784	1913046	808892
0.50	0.55	2994692	2373416	1686546	729996
0.55	0.60	2659956	2107769	1483011	649429
0.60	0.65	2327418	1850551	1317910	566579
0.65	0.70	1988771	1580531	1186048	482487
0.70	0.75	1655530	1312052	1086858	398786
0.75	0.80	1325459	1028757	996578	317016
0.80	0.85	1014664	847627	815880	238768
0.85	0.90	730068	626993	571441	165778
0.90	0.95	459899	402759	362928	100153
0.95	1.00	255887	241150	210512	46752
1.00	1.05	225639	199047	193996	35876
1.05	1.10	408432	328004	342236	83001
1.10	1.15	649067	511121	597832	144520
1.15	1.20	930175	767512	824280	212385
1.20	1.25	1218353	1037749	1042916	284505
1.25	1.30	1525731	1204198	1224187	359850
1.30	1.35	1839996	1409762	1519063	437276
1.35	1.40	2159103	1467508	1759471	516110
1.40	1.45	2486865	1574040	1998745	596325
1.45	1.50	2812851	1713414	2242119	675915
1.50	1.55	3143880	1857607	2498559	756262

Tabel 23. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05 ja y -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05, saadud keskmine sarnasus, arvatatuna spiraalide aegridade eukleidilise kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.55	1.60	3466405	1993186	2757686	836704
1.60	1.65	3788861	2143490	3001507	917406
1.65	1.70	4112695	2302567	3252996	997504
1.70	1.75	4443453	2459739	3479543	1075963
1.75	1.80	4757943	2648191	3793576	1115506
1.80	1.85	5085833	2834861	3982487	1235486
1.85	1.90	5404179	2997495	4238437	1321922
1.90	1.95	5728029	3219710	4483543	1390181
1.95	2.00	6051616	3420887	4748450	1467799
2.00	2.05	6377704	3620833	5005802	1545260
2.05	2.10	6701813	3820502	5270002	1624035
2.10	2.15	7028470	4025069	5540787	1699616
2.15	2.20	7354479	4217500	5814531	1776628
2.20	2.25	7680590	4444159	6099055	1853638
2.25	2.30	8009518	4649634	6370588	1931274
2.30	2.35	8337792	4861404	6652680	2007841
2.35	2.40	8668780	5071364	6936279	2084804
2.40	2.45	8998570	5289757	7221040	2161739
2.45	2.50	9328923	5503954	7509594	2238661
2.50	2.55	9658344	5727046	7792761	2315808
2.55	2.60	9988375	5934193	8080367	2395208
2.60	2.65	10317314	6154836	8368639	2469431
2.65	2.70	10645619	6371536	8657529	2546825
2.70	2.75	10975745	6593576	8941596	2623351
2.75	2.80	11304444	6811844	9234306	2700711
2.80	2.85	11636014	7031810	9525333	2777050
2.85	2.90	11968926	7250979	9810567	2854029
2.90	2.95	12300573	7473714	10100654	2930953
2.95	3.00	12632051	7692724	10397466	3007881

Tabel 24. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05 ja y -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.10	0.05	8018377	6981259	4692604	1920639
0.15	0.10	7500000	6416279	4293739	1810810
0.20	0.15	6979626	5850176	3894882	1700949
0.25	0.20	6462430	5295201	3496078	1591073
0.30	0.25	5957585	4746642	3099157	1481156
0.35	0.30	5467021	4259623	2732345	1372113
0.40	0.35	5004305	3788429	2413269	1264194
0.45	0.40	4584662	3381490	2158863	1158509
0.50	0.45	4162056	3036149	1900832	1029692
0.55	0.50	3800667	2727831	1863174	947974
0.60	0.55	3440311	2464275	1777435	859376
0.65	0.60	3184975	2274814	1777069	769476
0.70	0.65	3001855	2137214	1850164	714228
0.75	0.70	2765878	1951524	1937084	664359
0.80	0.75	2539577	1835969	2011298	592066
0.85	0.80	2317538	1718919	2017381	571558
0.90	0.85	2128327	1653757	1807740	450689
0.95	0.90	1900996	1556888	1620713	356089
1.00	0.95	1614574	1495131	1521063	220399
1.05	1.00	1712944	1587219	1522645	257523
1.10	1.05	2070169	1751336	1693777	404619
1.15	1.10	2377488	1976930	1854947	512382
1.20	1.15	2649507	2272505	1998748	599340
1.25	1.20	2937514	2491884	2182661	816266
1.30	1.25	3239921	2544099	2384462	770352
1.35	1.30	3521591	2628321	2463186	856563
1.40	1.35	3791131	2688231	2749996	929678
1.45	1.40	4068720	2604833	2937742	1002412
1.50	1.45	4344721	2689521	3143900	1059153
1.55	1.50	4654627	2723424	3345396	1140778

Tabel 25. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05 ja y -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.55	1.50	4654627	2723424	3345396	1140778
1.60	1.55	4938656	2753749	3537993	1220689
1.65	1.60	5253695	2815008	3778834	1291247
1.70	1.65	5563960	2915318	4028015	1365126
1.75	1.70	5920193	3021177	4293245	1468182
1.80	1.75	6293172	3145576	4569803	1582618
1.85	1.80	6668141	3274024	4855572	1688499
1.90	1.85	7035376	3406501	5147084	1783155
1.95	1.90	7407335	3557977	5445084	1871615
2.00	1.95	7781040	3712425	5751293	1953482
2.05	2.00	8160652	3870230	6064531	2043482
2.10	2.05	8558854	4034034	6383666	2144927
2.15	2.10	8955335	4206217	6704761	2261244
2.20	2.15	9373734	4383587	7031290	2378367
2.25	2.20	9781389	4565242	7355927	2492748
2.30	2.25	10180454	4750084	7688465	2597878
2.35	2.30	10580873	4942050	8028201	2700725
2.40	2.35	10980237	5141366	8372344	2803746
2.45	2.40	11384687	5347786	8720999	2906934
2.50	2.45	11789841	5560698	9074837	3010297
2.55	2.50	12196190	5779418	9430133	3113835
2.60	2.55	12609619	6003799	9789093	3217545
2.65	2.60	13028103	6233708	10154075	3321371
2.70	2.65	13447653	6468751	10522862	3425340
2.75	2.70	13867621	6707391	10892507	3529535
2.80	2.75	14296011	6949529	11264110	3633522
2.85	2.80	14714637	7194769	11636784	3737722
2.90	2.85	15139691	7443104	12011554	3842035
2.95	2.90	15566408	7694258	12387996	3946396
3.00	2.95	15994463	7947805	12765245	4050853

Tabel 26. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05 ja y -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

X- skaleerimis- faktor	Y- skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.05	0.10	8124059	7191911	4750218	1939020
0.10	0.15	7626297	6721112	4351353	1830913
0.15	0.20	7130605	6253008	3952488	1722741
0.20	0.25	6635662	5771292	3553631	1613937
0.25	0.30	6136762	5272843	3154833	1504742
0.30	0.35	5639720	4750506	2758984	1395336
0.35	0.40	5155342	4245711	2411113	1286879
0.40	0.45	4670051	3741115	2144708	1180030
0.45	0.50	4310095	3361052	1944463	1065066
0.50	0.55	3949997	2986242	1806651	985646
0.55	0.60	3660772	2766455	1722872	883872
0.60	0.65	3325206	2471389	1708706	809292
0.65	0.70	3139483	2273932	1790856	756171
0.70	0.75	2896533	2059047	1882422	695279
0.75	0.80	2625460	1907174	1964077	619132
0.80	0.85	2364149	1743372	1989617	530800
0.85	0.90	2108432	1619884	1796608	452393
0.90	0.95	1873711	1518670	1582388	359901
0.95	1.00	1592885	1433590	1465248	233173
1.00	1.05	1527593	1465794	1412730	202596
1.05	1.10	1829330	1633788	1523430	358805
1.10	1.15	2124631	1837216	1656513	443705
1.15	1.20	2430698	2155309	1826595	536170
1.2+	1.25	2728372	2372166	2015308	622586
1.25	1.30	3067995	2442765	2231304	721976
1.30	1.35	3387835	2527993	2497451	813962
1.35	1.40	3684752	2575844	2595087	900392
1.40	1.45	3965939	2606197	2800862	986757
1.45	1.50	4273194	2663924	3023993	1056743
1.50	1.55	4594704	2708859	3249378	1132419

Tabel 27. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05 ja y -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvutatuna spiraalide aegridade absoluutse kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.55	1.6	4887815	2745668	3469310	1207482
1.6	1.65	5191136	2804820	3680405	1280183
1.65	1.7	5511950	2901133	3930126	1351523
1.7	1.75	5851570	3007921	4199041	1451026
1.75	1.8	6230526	3134776	4468876	1563171
1.8	1.85	6603536	3251932	4753675	1677061
1.85	1.9	6947701	3402401	5045841	1769297
1.9	1.95	7348405	3551498	5343637	1859986
1.95	2.0	7725186	3713403	5649645	1941303
2.0	2.05	8103042	3873911	5962767	2029462
2.05	2.1	8493738	4063745	6282689	2130444
2.1	2.15	8883752	4243892	6605174	2243980
2.15	2.2	9301513	4425394	6937723	2362422
2.2	2.25	9711716	4613195	7260969	2476335
2.25	2.3	10108729	4803561	7592908	2583502
2.3	2.35	10590736	5004593	7933251	2686408
2.35	2.4	10911323	5209289	8279455	2789473
2.4	2.45	11315517	5420632	8629067	2892693
2.45	2.5	11786371	5634772	8984433	2996089
2.5	2.55	12142799	5860646	9340932	3099654
2.55	2.6	12554712	6089194	9700204	3203410
2.6	2.65	12973713	6322705	10067730	3307335
2.65	2.7	13393759	6560369	10437139	3411369
2.7	2.75	13815585	6801376	10809070	3515485
2.75	2.8	14237517	7045504	11182645	3619706
2.8	2.85	14663153	7292686	11557225	3724003
2.85	2.9	15089398	7542811	11935399	3828367
2.9	2.95	15516527	7795418	12310866	3932812
2.95	3.0	15944378	8050208	12688972	4037311

Tabel 28. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05 ja y -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvatatuna spiraalide aeGRIDade eukleidilise kaugusena.

X- skaleerimis- faktor	Y- skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.1	0.05	6032114	5324557	3957160	1411510
0.15	0.1	5685297	4979053	3677509	1333376
0.2	0.15	5330170	4616471	3399450	1255332
0.25	0.2	4976783	4261341	3123892	1177442
0.3	0.25	4625614	3914871	2857768	1099484
0.35	0.3	4279158	3559358	2581528	1021528
0.4	0.35	3933835	3250103	2324125	943692
0.45	0.4	3593321	2900282	2041757	864561
0.5	0.45	3254984	2583628	1834870	787551
0.55	0.5	2926163	2299643	1619632	708543
0.6	0.55	2596063	2096360	1427557	627445
0.65	0.6	2265540	1781241	1273417	544026
0.7	0.65	1931440	1510151	1152215	459899
0.75	0.7	1599429	1265337	1057419	376309
0.8	0.75	1287002	1020643	967576	294487
0.85	0.8	981026	807196	819225	217712
0.9	0.85	704985	593332	539332	146530
0.95	0.9	472201	366759	339512	83444
1.0	0.95	292835	227769	214344	36044
1.05	1.0	310522	289917	252943	46295
1.1	1.05	492325	388527	410889	98776
1.15	1.1	734326	574904	622190	162177
1.2	1.15	1012701	816917	835238	213136
1.25	1.2	1298131	1056687	1062944	304313
1.3	1.25	1593969	1259369	1307803	380076
1.35	1.3	1914564	1387911	1544653	457669
1.4	1.35	2238777	1490033	1777167	535664
1.45	1.4	2553474	1609322	2007188	616793
1.5	1.45	2878508	1741947	2258600	697000
1.55	1.5	3199711	1881654	2504960	776815

Tabel 29. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05 ja y -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvatatuna spiraalide aegridade eukleidilise kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.60	1.55	3519521	2022668	2749738	857310
1.65	1.60	3838262	2177992	2995296	937624
1.70	1.65	4148503	2341053	3231423	1017523
1.75	1.70	4467522	2511486	3474484	1099612
1.80	1.75	4783680	2691000	3710825	1175832
1.85	1.80	5101071	2853991	3956191	1254338
1.90	1.85	5417380	3026935	4197675	1332556
1.95	1.90	5735630	3266661	4444882	1409964
2.00	1.95	6053623	3467045	4703666	1487992
2.05	2.00	6372204	3669056	4960843	1565371
2.10	2.05	6691927	3862573	5229424	1642580
2.15	2.10	7011675	4079994	5494010	1719696
2.20	2.15	7332555	4280755	5760357	1796730
2.25	2.20	7654967	4508786	6032677	1873694
2.30	2.25	7978267	4713963	6305606	1950864
2.35	2.30	8299528	4928072	6583239	2027562
2.40	2.35	8622225	5139947	6861246	2104561
2.45	2.40	8945044	5358811	7140728	2181487
2.50	2.45	9268881	5577982	7421196	2258598
2.55	2.50	9591274	5793822	7702604	2335501
2.60	2.55	9914479	5987407	7987407	2412398
2.65	2.60	10288783	6231094	8267410	2489910
2.70	2.65	10651999	6449913	8505351	2566210
2.75	2.70	11088457	6688864	8833437	2643107
2.80	2.75	11201344	6888864	9115800	2720099
2.85	2.80	11553332	7190834	9401573	2796928
2.90	2.85	11856400	7328030	9684911	2873832
2.95	2.90	12180313	7548713	9969963	2950735
3.00	2.95	12501747	7769581	10255297	3027651

Tabel 30. Spiraalide skaleerimisel x -skaleerimisfakoriga 0.05 - 1.50, sammuga 0.05 ja y -skaleerimisfakoriga 0.10 - 1.55, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvutatuna spiraalide aegridade eukleidilise kaugusena.

X- skaleerimis- faktor	Y- skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
0.05	0.10	6138886	5423098	4037483	1432997
0.10	0.15	5782933	5068206	3756617	1358470
0.15	0.20	5427660	4712470	3479497	1276822
0.20	0.25	5074866	4360662	3208990	1198880
0.25	0.30	4725888	4001152	2938185	1120985
0.30	0.35	4375777	3664284	2668597	1042784
0.35	0.40	4027344	3390312	2408324	964901
0.40	0.45	3681878	2988161	2147983	886915
0.45	0.50	3342083	2668683	1905184	808746
0.50	0.55	3006895	2374231	1681903	729855
0.55	0.60	2676665	2107657	1478565	649277
0.60	0.65	2346901	1850443	1318203	566417
0.65	0.70	2011653	1585023	1183190	482248
0.70	0.75	1681384	1311923	1090451	398616
0.75	0.80	1356360	1040876	1002648	315686
0.80	0.85	1050930	807963	843225	238632
0.85	0.90	771458	627595	599995	165670
0.90	0.95	516932	407519	397990	100073
0.95	1.00	320828	225914	245042	46702
1.00	1.05	284724	193595	236899	35589
1.05	1.10	444127	328901	368310	82993
1.10	1.15	668799	516283	601618	145439
1.15	1.20	940493	767969	816408	212419
1.20	1.25	1230204	1037967	1031587	284951
1.25	1.30	1532281	1270940	1270074	359501
1.30	1.35	1842610	1347331	1498081	437395
1.35	1.40	2161393	1467885	1734351	516148
1.40	1.45	2482991	1694544	2007618	596344
1.45	1.50	2802248	1713767	2202434	675908
1.50	1.55	3126337	1857935	2452097	756223

Tabel 31. Spiraalide skaleerimisel x -skaleerimisfakoriga 1.55 - 2.95, sammuga 0.05 ja y -skaleerimisfakoriga 1.60 - 3.00, sammuga 0.05, ning tasakaalustamisel saadud keskmine sarnasus, arvutatuna spiraalide aegridade eukleidilise kaugusena.

X - skaleerimis- faktor	Y - skaleerimis- faktor	PD vs PD skaleeri- tud	PD vs KT skaleeri- tud	KT vs PD skaleeri- tud	KT vs KT skaleeri- tud
1.55	1.60	3442331	1993502	2701465	836832
1.60	1.65	3758366	2143780	2936983	916941
1.65	1.70	4075887	2302801	3181053	997361
1.70	1.75	4391461	2461438	3419188	1076976
1.75	1.80	4708613	2648330	3665187	1155349
1.80	1.85	5025277	2834947	3898686	1235893
1.85	1.90	5342876	3021798	4140725	1318222
1.90	1.95	5660813	3219647	4387557	1389828
1.95	2.00	5978700	3410621	4642477	1464469
2.00	2.05	6298771	3620681	4900497	1544901
2.05	2.10	6617308	3821628	5162103	1622100
2.10	2.15	6937940	4024845	5426168	1699205
2.15	2.20	7258647	4232155	5691491	1776047
2.20	2.25	7578251	4438992	5965933	1853179
2.25	2.30	7901103	4649895	6230077	1930705
2.30	2.35	8223045	4861903	6516452	2007340
2.35	2.40	8548685	5077300	6794252	2084873
2.40	2.45	8871172	5289443	7073161	2161191
2.45	2.50	9195177	5502948	7352270	2238092
2.50	2.55	9518693	5702004	7633111	2315501
2.55	2.60	9840491	5968608	7914074	2391896
2.60	2.65	10161513	6154294	8199626	2468788
2.65	2.70	10482106	6372549	8479515	2545681
2.70	2.75	10804100	6601362	8760596	2622575
2.75	2.80	11126312	6811862	9044033	2699473
2.80	2.85	11446410	7031640	9328434	2776369
2.85	2.90	11774414	7250501	9612991	2853282
2.90	2.95	12095704	7471313	9897795	2930192
2.95	3.00	12419333	7692219	10183005	3007099

Tabel 32. Esimese meetodi järgi sarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD roteeritud ja skaleeritud	PD vs KT roteeritud ja skaleeritud	KT vs PD roteeritud ja skaleeritud	KT vs KT roteeritud ja skaleeritud
0.9	0.95	60	4320937	2959618	3238962	983498
0.9	0.95	75	3496741	2476487	2638321	777618
0.9	0.95	90	2955077	2169808	2301005	636948
0.9	0.95	105	3198792	2298688	2523746	697230
0.9	0.95	120	3860656	2635315	2998405	866045
0.95	1.0	60	4505168	3178122	3344440	990412
0.95	1.0	75	3577159	2611695	2705541	757129
0.95	1.0	90	2902610	2241253	2308762	578471
0.95	1.0	105	3194297	2470412	2546491	651281
0.95	1.0	120	3951191	2780021	3078199	846607
1.0	1.05	60	4610480	3348384	3419664	971684
1.0	1.05	75	3631882	2735321	2729514	749242
1.0	1.05	90	2988172	2353191	2336159	585758
1.0	1.05	105	3261157	2563798	2581777	644078
1.0	1.05	120	4070540	2969455	3156285	853095
1.0	0.95	60	4846715	3477964	3593825	1046488
1.0	0.95	75	3899659	2860034	2943406	817440
1.0	0.95	90	3113619	2385458	2450484	622747
1.0	0.95	105	3233355	2419795	2573704	662916
1.0	0.95	120	3860109	2687068	3024632	827692
1.05	1.0	60	4860313	3639421	3643542	988410
1.05	1.0	75	3936550	2970824	2963801	802533
1.05	1.0	90	3116696	2469299	2445598	602931
1.05	1.0	105	3265652	2535525	2587984	643798
1.05	1.0	120	3973618	2866807	3094708	831723

Tabel 33. Teise meetodi järgi sarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD roteeritud ja skaleeritud	PD vs KT roteeritud ja skaleeritud	KT vs PD roteeritud ja skaleeritud	KT vs KT roteeritud ja skaleeritud
0.95	0.9	45	3162175	2102905	1972908	762611
0.95	0.9	60	2714052	1819992	1739968	661037
0.95	0.9	75	2428512	1618507	1549261	605541
0.95	0.9	90	2390909	1606232	1540464	599652
0.95	0.9	105	2559892	1709683	1636179	641275
1.0	0.95	45	3224388	2107235	2049528	777304
1.0	0.95	60	2763599	1829899	1792154	671997
1.0	0.95	75	2462404	1641830	1588723	613124
1.0	0.95	90	2422234	1604003	1581036	606782
1.0	0.95	105	2595718	1709998	1682909	649554
1.0	1.05	45	3317844	2128539	2144397	800239
1.05	1.0	60	2844457	1862193	1865689	692466
1.05	1.0	75	2533007	1641382	1654791	631071
1.05	1.0	90	2288497	1629371	1646733	623599
1.05	1.0	105	2664617	1735965	1753140	666875
1.05	1.0	105	3127767	2014261	1986392	756680
0.95	1.0	60	2674827	1747365	1737381	655288
0.95	1.0	75	2421256	1592727	1578785	608591
0.95	1.0	105	2631673	1734351	1708189	660548
1.0	1.05	45	3119042	2209962	2063648	772268
1.0	1.05	60	2729555	1769068	1792428	668635
1.0	1.05	75	2460503	1585688	1607683	617852
1.0	1.05	90	2477229	1611247	1632786	622703
1.0	1.05	105	2699453	1753041	1770436	676401
1.0	1.05	45	3119042	2209962	2063648	772268
1.0	1.05	105	2664617	1735965	1753140	666875

Tabel 34. Tasakaalustatud ning esimese meetodi järgi sarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD rotee- ritud ja skaleeri- tud	PD vs KT rotee- ritud ja skaleeri- tud	KT vs PD rotee- ritud ja skaleeri- tud	KT vs KT rotee- ritud ja skaleeri- tud
0.9	0.95	60	4243464	2959642	3197062	983428
0.9	0.95	75	3447981	2476586	2607126	777601
0.9	0.95	90	2932135	2169849	2278355	636824
0.9	0.95	105	3172128	2298597	2496584	696986
0.9	0.95	120	3821035	2635094	2967064	865705
0.95	1.0	60	4427045	3178171	3298252	990378
0.95	1.0	75	3527492	2613810	2673867	757168
0.95	1.0	90	2890177	2241354	2289611	578445
0.95	1.0	105	3175503	2407355	2523286	651111
0.95	1.0	120	3918256	2779823	3046116	846315
1.0	1.05	60	4535811	3364426	3374098	971693
1.0	1.05	75	3585869	2735446	2698622	749344
1.0	1.05	90	2964259	2353349	2316248	585876
1.0	1.05	105	3249086	2563773	2559735	644010
1.0	1.05	120	4038684	2969289	3142255	852877
1.0	0.95	60	4767193	3477967	3546625	1046412
1.0	0.95	75	3847913	2860111	2909501	817448
1.0	0.95	90	3096349	2385509	2428233	622667
1.0	0.95	105	3212238	2419704	2549293	662704
1.0	0.95	120	3825942	2686843	2995377	827372
1.05	1.0	60	4788637	3639439	3590494	988386
1.05	1.0	75	3886857	2970926	2928326	802607
1.05	1.0	90	3100929	2469437	2423999	602983
1.05	1.0	105	3243860	2553506	2564205	643683
1.05	1.0	120	3939323	2866643	3061856	831472

Tabel 35. Tasakaalustatud ning teise meetodi järgi sarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD roteeritud ja skaleeritud	PD vs KT roteeritud ja skaleeritud	KT vs PD roteeritud ja skaleeritud	KT vs KT roteeritud ja skaleeritud
0.95	0.9	45	3160841	2102863	1953283	762456
0.95	0.9	60	2709817	1819917	1720450	660846
0.95	0.9	75	2423177	1618380	1530495	690538
0.95	0.9	90	2382969	1606070	1524270	599419
0.95	0.9	105	2549723	1709488	1617280	641020
1.0	0.95	45	2719943	2107221	2027910	771196
1.0	0.95	60	2755944	1828943	1772741	671848
1.0	0.95	75	2544098	1614781	1578435	617935
1.0	0.95	90	2411290	1603854	1564347	565062
1.0	0.95	105	2382349	2179086	1662928	649302
1.05	1.0	45	3124829	2016262	1963765	776957
1.05	1.0	60	2833386	2188538	2122303	692624
1.05	1.0	75	2521185	1841275	1638542	623939
1.05	1.0	90	2474455	1629912	1626625	623939
1.05	1.0	105	2648526	1735765	1730420	666626
1.05	1.0	105	2621697	1739411	1687882	660296
0.95	1.0	60	2665993	1747282	1718654	718654
0.95	1.0	75	2407497	1570773	1538903	604351
0.95	1.0	105	2621697	1739411	1687882	660296
1.0	1.05	45	3183759	2020972	2039845	772816
1.0	1.05	60	2717377	1769004	1772143	668502
1.0	1.05	75	2451532	1585954	1589975	617678
1.0	1.05	90	2460789	1611068	1617909	622487
1.0	1.05	105	2682077	1752827	1747739	676145
1.0	1.05	90	2407884	1595243	1560913	603686
1.0	1.05	105	2682077	1752827	1747739	676145
1.0	1.05	105	2682077	1752827	1747739	676145

Tabel 36. Esimese meetodi järgi ebasarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD roteeritud ja skaleeritud	PD vs KT roteeritud ja skaleeritud	KT vs PD roteeritud ja skaleeritud	KT vs KT roteeritud ja skaleeritud
2.9	2.85	225	27330248	10347947	24547725	6888353
2.9	2.85	240	29638040	13974465	25942568	7301172
2.9	2.85	255	31330894	17180131	26895917	7586670
2.9	2.85	270	32170289	18805545	27339715	7720414
2.9	2.85	285	32041877	18538818	27276215	7694562
2.95	2.9	225	27652806	10390336	24885808	6966392
2.95	2.9	240	30374901	14078289	26301946	7391143
2.95	2.9	255	31720646	17339523	27721187	7681933
2.95	2.9	270	31735444	18992446	27722593	7818081
2.95	2.9	285	32439850	18720843	27643786	7791874
3.0	2.95	225	30715487	10430390	25224111	7744166
3.0	2.95	240	30359446	14182296	26661433	7481666
3.0	2.95	255	30599410	17490938	27646551	7597136
3.0	2.95	270	31957789	19179439	28105525	7985756
3.0	2.95	285	31385955	18009241	28052471	7971535
3.05	3.0	225	32500838	10476250	25562636	7135266
3.05	3.0	240	33150321	14285605	27010322	7559059
3.05	3.0	255	33784905	17586887	28021948	7872490
3.05	3.0	270	33344051	19365673	28488513	8013458
3.05	3.0	285	33740871	19085152	28407020	7968517
2.9	2.95	225	27917530	10096871	24605366	7493368
2.9	2.95	240	29171638	13896540	26108198	7340198
2.9	2.95	255	29773684	16162385	27620896	7544426
2.9	2.95	270	32436607	18906211	27662295	7793257
2.9	2.95	285	32738868	19620656	27668079	7798379
2.95	3.0	225	32100157	10138707	27326768	6904383
2.95	3.0	240	31952044	13999233	26489733	6948882
2.95	3.0	255	31957789	17424931	27501376	7473121
2.95	3.0	270	32841452	19092163	28011282	7890955
2.95	3.0	285	32769063	18845099	27982408	7877634

Tabel 37. Teise meetodi järgi ebasarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD roteeritud ja skaleeritud	PD vs KT roteeritud ja skaleeritud	KT vs PD roteeritud ja skaleeritud	KT vs KT roteeritud ja skaleeritud
2.9	2.85	225	25139492	18186085	20893347	5791659
2.9	2.85	240	25367413	18401207	21089760	5838710
2.9	2.85	255	25454392	19489635	21189506	5858862
2.9	2.85	270	25428686	18483894	21179473	5852562
2.9	2.85	285	25290613	18378027	21066735	5821063
2.95	2.9	225	25478041	18360170	21186631	5867780
2.95	2.9	240	25698989	18622584	21392501	5915042
2.95	2.9	255	25788606	18702377	21483699	5935295
2.95	2.9	270	25760331	18705988	21473732	5929047
2.95	2.9	285	25621091	18599665	21360422	5918934
3.0	2.95	225	25580731	18662291	21479700	5943925
3.0	2.95	240	26033733	20337733	21865861	5994355
3.0	2.95	255	26119995	18948229	21777964	5980704
3.0	2.95	270	26095571	18929791	21676704	6005489
3.0	2.95	285	25950644	18958044	21654531	5983855
3.05	3.0	225	26138704	18847536	21769223	6037090
3.05	3.0	240	26365267	19066433	21798964	6070754
3.05	3.0	255	26425711	19165380	22071521	6067202
3.05	3.0	270	26286507	19043095	21947752	6050979
3.05	3.0	285	26040995	19200402	20934357	5800846
2.9	2.95	225	25175294	18200402	21176997	5858122
2.9	2.95	240	25435139	18455237	21318423	5891649
2.9	2.95	255	25595375	18588529	21360435	5905397
2.9	2.95	270	25735512	18682881	21307648	5889347
2.9	2.95	285	25566530	18655364	21232765	5867968
2.95	3.0	225	25660227	18616053	21489009	5934446
2.95	3.0	240	25782266	18810570	21617778	5968085
2.95	3.0	255	25928781	18581532	21658977	5976844
2.95	3.0	270	26055884	18787284	21601801	5960107
2.95	3.0	285	25894152	18787284	21601801	5960107

Tabel 38. Tasakaalustatud ning esimese meetodi järgi ebasarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD rotee- ritud ja skaleeri- tud	PD vs KT rotee- ritud ja skaleeri- tud	KT vs PD rotee- ritud ja skaleeri- tud	KT vs KT rotee- ritud ja skaleeri- tud
225.0	2.9	2.85	27330248	10347947	24547725	6888353
240.0	2.9	2.85	29638040	13974465	25942568	7301172
255.0	2.9	2.85	31330894	17180131	26895917	7586670
270.0	2.9	2.85	32170289	18805545	27339715	7720414
285.0	2.9	2.85	32037742	18538818	27262145	7694562
225.0	2.95	2.9	27652806	10390336	24885808	6966392
240.0	2.95	2.9	29998680	14078289	26301946	7391413
255.0	2.95	2.9	31720656	17339523	27721187	7681933
270.0	2.95	2.9	32574544	18992446	27722593	7818081
285.0	2.95	2.9	32439850	18720843	27643786	7791874
225.0	3.0	2.95	27975617	10430390	25224111	7049458
240.0	3.0	2.95	30359446	14182296	26661433	7481666
255.0	3.0	2.95	32110500	17490938	27646531	7777206
270.0	3.0	2.95	32798859	19179439	28105525	7915756
285.0	3.0	2.95	32842007	18902941	28052471	7888193
225.0	3.05	3.0	28298667	10476250	25562636	7132556
240.0	3.05	3.0	30719926	14285605	27021032	7571936
255.0	3.05	3.0	32500157	17658687	28021948	7872490
270.0	3.05	3.0	33380333	19365673	28488513	8013438
285.0	3.05	3.0	33244051	19085152	28407020	7986517
225.0	2.9	2.95	27318905	10096871	24650366	6901125
240.0	2.9	2.95	29717638	13896540	26108198	7340198
255.0	2.9	2.95	31522530	17262977	27125906	7644426
270.0	2.9	2.95	32436884	18906215	27662895	7793257
285.0	2.9	2.95	32366607	18663711	27600679	7780299
225.0	2.95	3.0	27642262	10138707	24988891	6984293
240.0	2.95	3.0	30077878	13999233	26467796	7430501
255.0	2.95	3.0	31915521	17424931	27501323	7739728
270.0	2.95	3.0	32841452	19092613	28011282	7890955
285.0	2.95	3.0	32769063	18845099	27982408	7877634

Tabel 39. Tasakaalustatud ning teise meetodi järgi ebasarnaseks muutvate väärtustega roteeritud ja skaleeritud Archimedese spiraalide sarnasus.

X- skaleeri- misfaktor	Y- skaleeri- misfaktor	Roteeri- misnurk	PD vs PD rotee- ritud ja skaleeri- tud	PD vs KT rotee- ritud ja skaleeri- tud	KT vs PD rotee- ritud ja skaleeri- tud	KT vs KT rotee- ritud ja skaleeri- tud
2.9	2.85	225	24784844	18185489	20559337	5790235
2.9	2.85	240	25006939	18400641	20761256	5837312
2.9	2.85	255	25091115	18498995	20850453	5857495
2.9	2.85	270	25063340	18438380	20840290	5851287
2.9	2.85	285	25076953	18377531	20728701	5828701
2.95	2.9	225	25109764	18405564	20848693	5866336
2.95	2.9	240	25337860	18622006	21046994	5913623
2.95	2.9	255	25741795	18702187	21139379	5933906
2.95	2.9	270	25389208	18705464	21129394	5927693
2.95	2.9	285	25249566	18599165	21007152	5896068
3.0	2.95	225	25440517	18625673	21137351	5942458
3.0	2.95	240	25656401	18845333	21339744	5990392
3.0	2.95	255	25745160	18947268	21428227	6001436
3.0	2.95	270	25711366	18942736	21418117	6004114
3.0	2.95	285	25596085	18880271	21305702	5988603
3.05	3.0	225	25765603	18846908	21420712	6067923
3.05	3.0	240	25987520	19065743	21625964	6086772
3.05	3.0	255	26027060	19164809	21716928	6080548
3.05	3.0	270	26043944	19249761	21706838	6048721
3.05	3.0	285	25902986	19042573	21594167	5995401
2.9	2.95	225	24800429	18917992	20599856	5797404
2.9	2.95	240	25002865	18444651	20837860	5856701
2.9	2.95	255	25322164	18852967	20976762	5907524
2.9	2.95	270	25265933	18682843	21061287	5899031
2.9	2.95	285	25193136	18564849	20965280	5882417
2.95	3.0	225	25170989	18419890	20886637	5875505
2.95	3.0	240	25414681	18666907	21126531	5933004
2.95	3.0	255	25578514	18809998	21265898	5966688
2.95	3.0	270	25589329	18850684	21310980	5975458
2.95	3.0	285	25518444	18786758	21254199	5958756

Lisa 5 – Suurendatud andmehulkade suurused

Tabel 40. Roteerimisega suurendatud andmehulkade suurused.

Andmehulga suurendamise meetod	Andmehulga suurus
Roteeritud nurkade 15-345, sammuga 15 võrra	828
Roteeritud spiraale sarnasemaks muutvate nurkadega, absoluutse kauguse sarnasusmeetodi põhjal	180
Roteeritud spiraale sarnasemaks muutvate nurkadega, Eukleidese kauguse sarnasusmeetodi põhjal	180
Roteeritud spiraale ebasarnasemaks muutvate nurkadega	180
Roteeritud nurkade 15-345, sammuga 15 võrra ja esialgsed treeningandmed	864
Roteeritud spiraale sarnasemaks muutvate nurkadega ja esialgsed treeningandmed	216
Roteeritud spiraale ebasarnasemaks muutvate nurkadega ja esialgsed treeningandmed	216

Tabel 41. Skaleerimisega suurendatud andmehulkade suurused.

Andmehulga suurendamise meetod	Andmehulga suurus
Skaleeritud mõlemat koordinaati skaleerimisfaktoritega 0.05-3, sammuga 0.05	4248
Skaleeritud spiraale sarnasemaks muutvate skaleerimisfaktoritega, absoluutse kauguse sarnasusmeetodi põhjal	180
Skaleeritud spiraale sarnasemaks muutvate skaleerimisfaktoritega, Eukleidese kauguse sarnasusmeetodi põhjal	180
Skaleeritud spiraale ebasarnasemaks muutvate skaleerimisfaktoritega	180
Skaleeritud mõlemat koordinaati skaleerimisfaktoritega 0.05-3, sammuga 0.05, lisatud esialgsed treeningandmed	4284
Skaleeritud spiraale sarnasemaks muutvate skaleerimisfaktoritega, lisatud esialgsed treeningandmed	216
Skaleeritud spiraale ebasarnasemaks muutvate skaleerimisfaktoritega, lisatud esialgsed treeningandmed	216

Tabel 42. Roteerimise ja skaleerimisega suurendatud andmehulkade suurused.

Andmehulga suurendamise meetod	Andmehulga suurus
Roteeritud ja skaleeritud spiraale sarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega, absoluutse kauguse sarnasusmeetodi põhjal	900
Roteeritud ja skaleeritud spiraale sarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega, Eukleidese kauguse sarnasusmeetodi põhjal	900
Roteeritud ja skaleeritud spiraale ebasarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega	900
Roteeritud ja skaleeritud spiraale sarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega , lisatud esialgsed treeningandmed	936
Roteeritud ja skaleeritud spiraale ebasarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega , lisatud esialgsed treeningandmed	936

Tabel 43. Roteerimisega suurendatud ja SMOTE-iga tasakaalustatud andmehulkade suurused.

Andmehulga suurendamise meetod	Andmehulga suurus
Roteeritud nurkade 15-345, sammuga 15 võrra, tasakaalustatud SMOTE-iga	1058
Roteeritud spiraale sarnasemaks muutvate nurkadega, absoluutse kauguse sarnasusmeetodi põhjal, ja tasakaalustatud SMOTE-iga	230
Roteeritud spiraale sarnasemaks muutvate nurkadega, Eukleidese kauguse sarnasusmeetodi põhjal, ja tasakaalustatud SMOTE-iga	230
Roteeritud spiraale ebasarnasemaks muutvate nurkadega ja tasakaalustatud SMOTE-iga	230
Roteeritud nurkade 15-345, sammuga 15 võrra, tasakaalustatud SMOTE-iga, lisatud ka esialgsed treeningandmed	1104
Roteeritud spiraale sarnasemaks muutvate nurkadega, tasakaalustatud SMOTE-iga, lisatud ka esialgsed treeningandmed	276
Roteeritud spiraale ebasarnasemaks muutvate nurkadega, tasakaalustatud SMOTE-iga, lisatud ka esialgsed treeningandmed	276

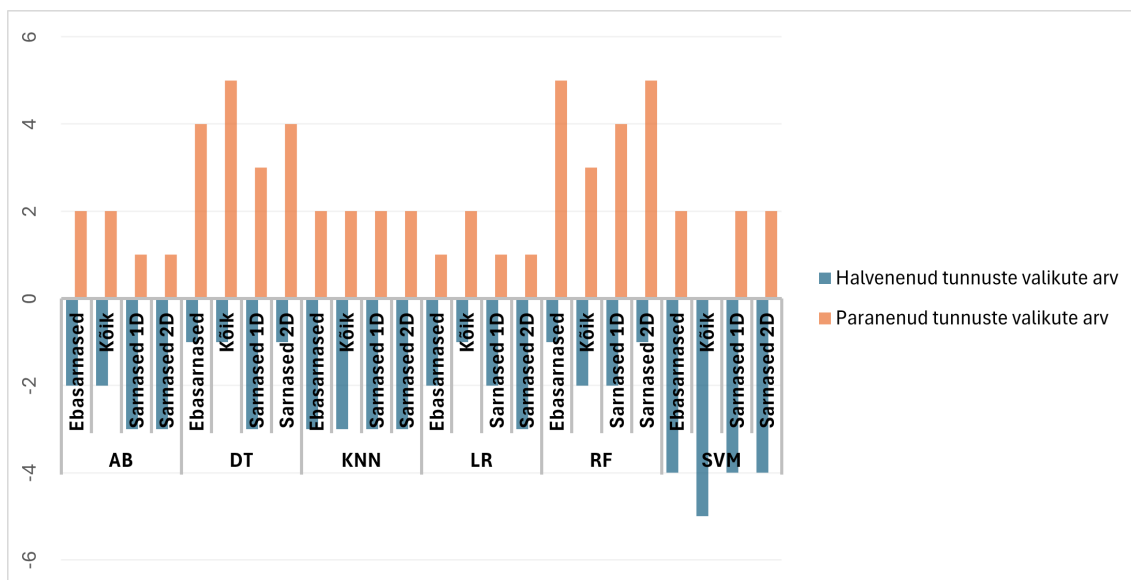
Tabel 44. Skaleerimisega suurendatud ja SMOTE-iga tasakaalustatud andmehulkade suurus.

Andmehulga suurendamise meetod	Andmehulga suurus
Skaleeritud mõlemat koordinaati skaleerimisfaktoritega 0.05-3, sammuga 0.05, tasakaalustatud SMOTE-iga	5428
Skaleeritud spiraale sarnasemaks muutvate skaleerimisfaktoritega, absoluutse kauguse sarnasusmeetodi põhjal, tasakaalustatud SMOTE-iga	230
Skaleeritud spiraale sarnasemaks muutvate skaleerimisfaktoritega, Eukleidese kauguse sarnasusmeetodi põhjal, tasakaalustatud SMOTE-iga	230
Skaleeritud spiraale ebasarnasemaks muutvate skaleerimisfaktoritega, tasakaalustatud SMOTE-iga	230
Skaleeritud mõlemat koordinaati skaleerimisfaktoritega 0.05-3, sammuga 0.05, tasakaalustatud SMOTE-iga ja lisatud esialgsed treeningandmed	5474
Skaleeritud spiraale sarnasemaks muutvate skaleerimisfaktoritega, tasakaalustatud SMOTE-iga ja lisatud esialgsed treeningandmed	276
Skaleeritud spiraale ebasarnasemaks muutvate skaleerimisfaktoritega, tasakaalustatud SMOTE-iga ja lisatud esialgsed treeningandmed	276

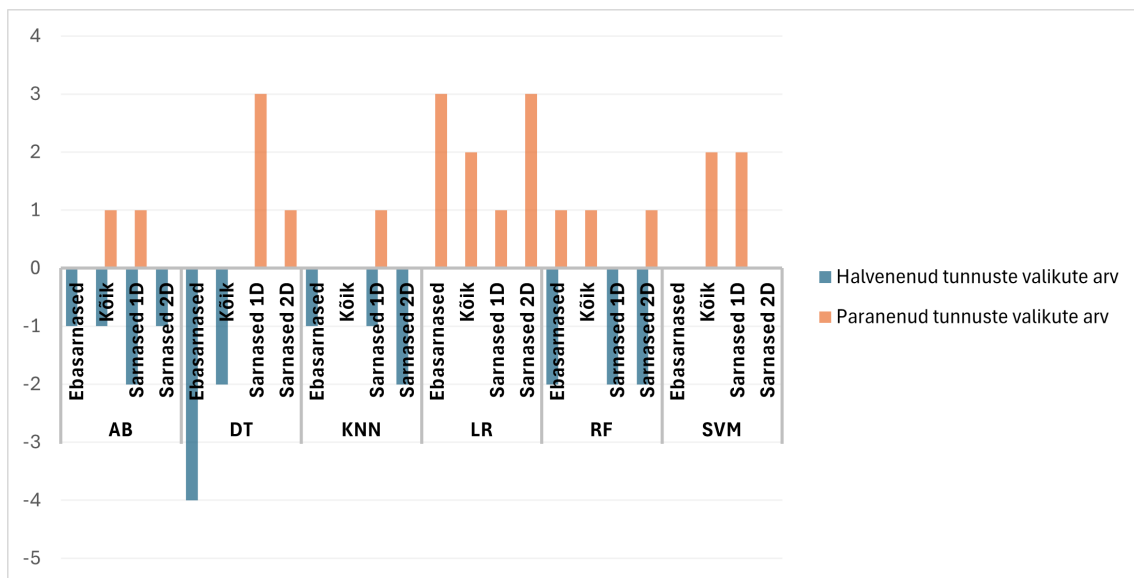
Tabel 45. Roteerimise ja skaleerimisega suurendatud, SMOTE-iga tasakaalustatud andmehulkade suurused.

Andmehulga suurendamise meetod	Andmehulga suurus
Roteeritud ja skaleeritud spiraale sarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega, absoluutse kauguse sarnasusmeetodi põhjal, tasakaalustatud SMOTE-iga	1150
Roteeritud ja skaleeritud spiraale sarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega, Eukleidese kauguse sarnasusmeetodi põhjal, tasakaalustatud SMOTE-iga	1150
Roteeritud ja skaleeritud spiraale ebasarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega, tasakaalustatud SMOTE-iga	1150
Roteeritud ja skaleeritud spiraale sarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega, tasakaalustatud SMOTE-iga ning lisatud esialgsed treeningandmed	1196
Roteeritud ja skaleeritud spiraale ebasarnasemaks muutvate roteerimisnurkadega ja skaleerimisfaktoritega, tasakaalustatud SMOTE-iga ning lisatud esialgsed treeningandmed	1196

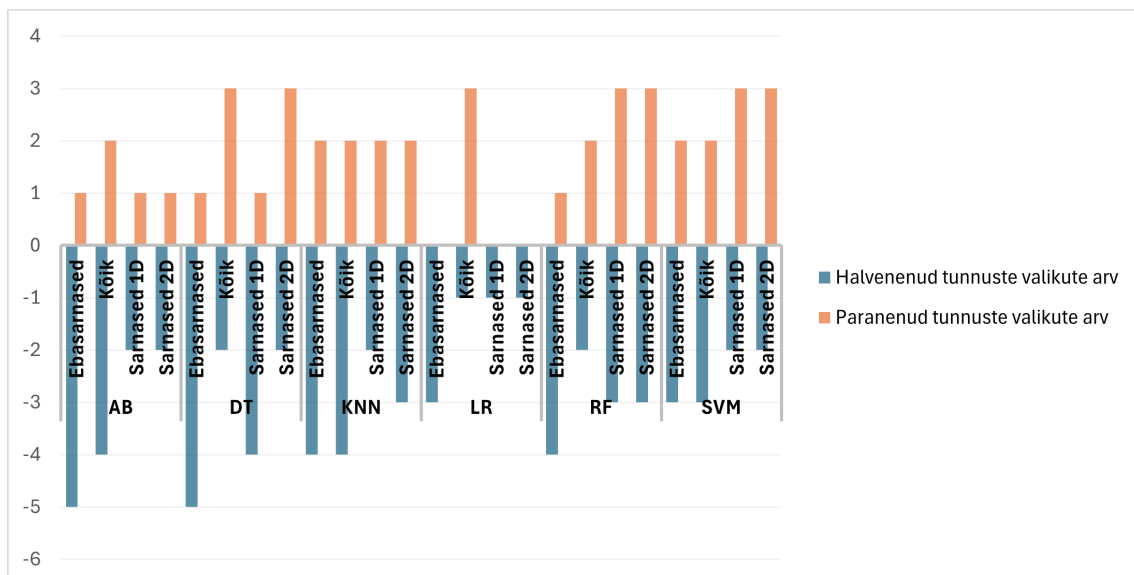
Lisa 6 – Suurendatud andmehulkade ja tunnuste valikute põhjal saadud klassifitseerimise tulemused



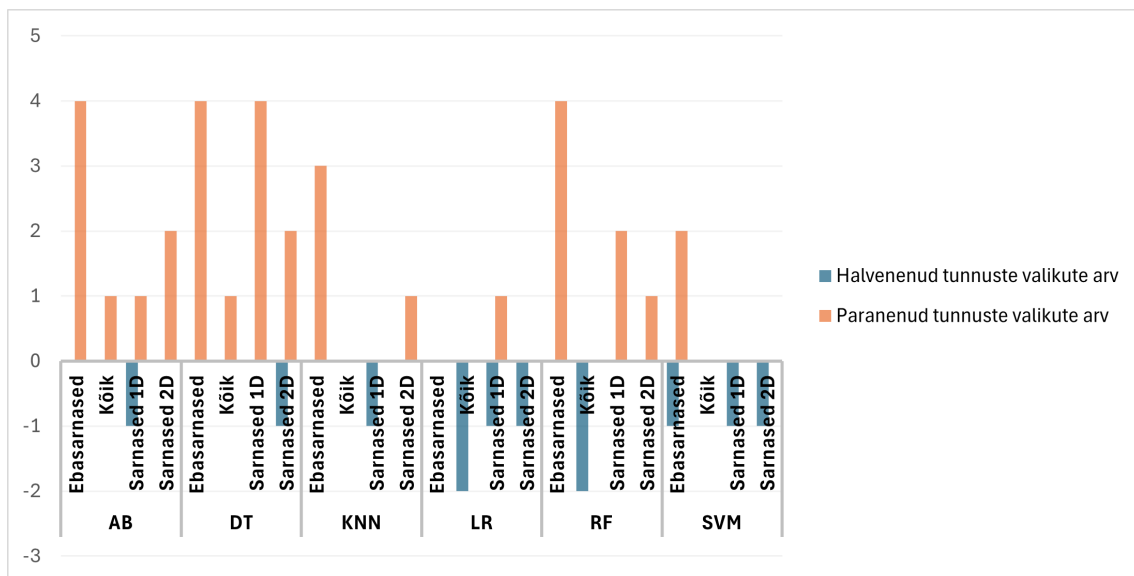
Joonis 11. Roteerimise rakendamise suurendatud andmehulkade põhjal saadud klassifitseerimistulemuste võrdlus esialgse andmehulga põhjal klassifitseerimisega.



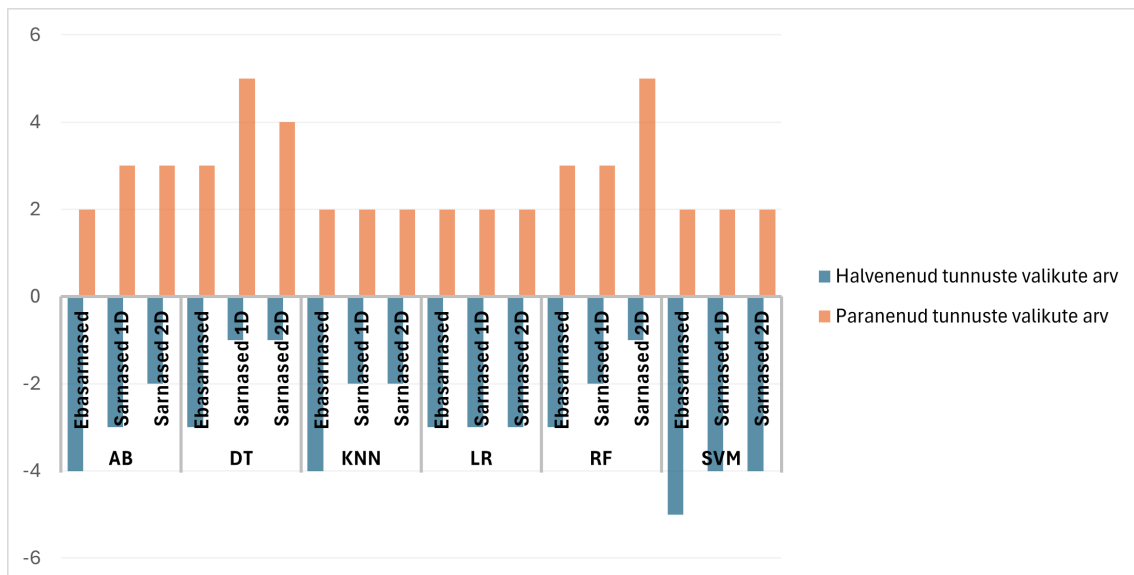
Joonis 12. Esialgse andmehulga lisamisel roteeritud andmehulkadesse saadud klassifitseerimistulemuste võrdlus ainult roteerimise rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemustega.



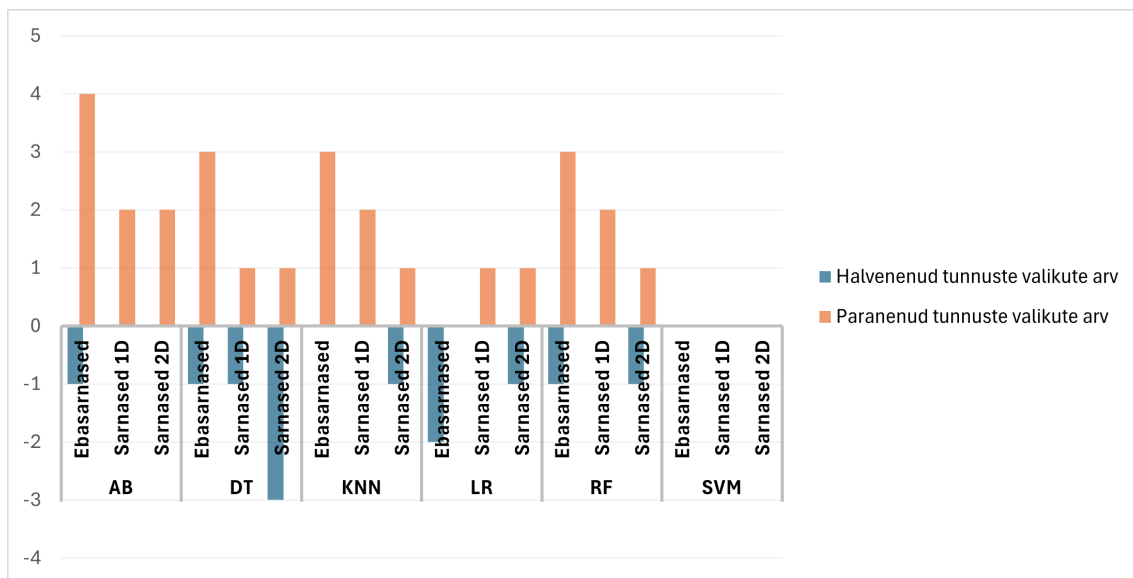
Joonis 13. Skaleerimise rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemuste võrdlus esialgse andmehulga põhjal klassifitseerimisega.



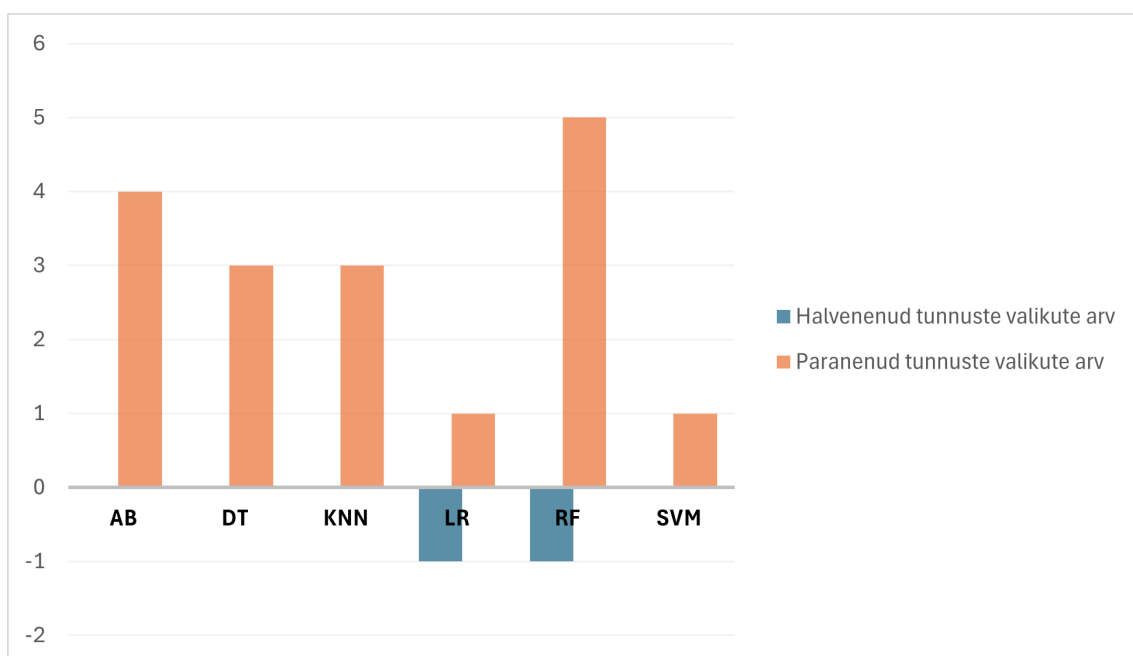
Joonis 14. Esialgse andmehulga lisamisel skaleeritud andmehulkadesse saadud klassifitseerimistulemuste võrdlus ainult skaleerimisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemustega.



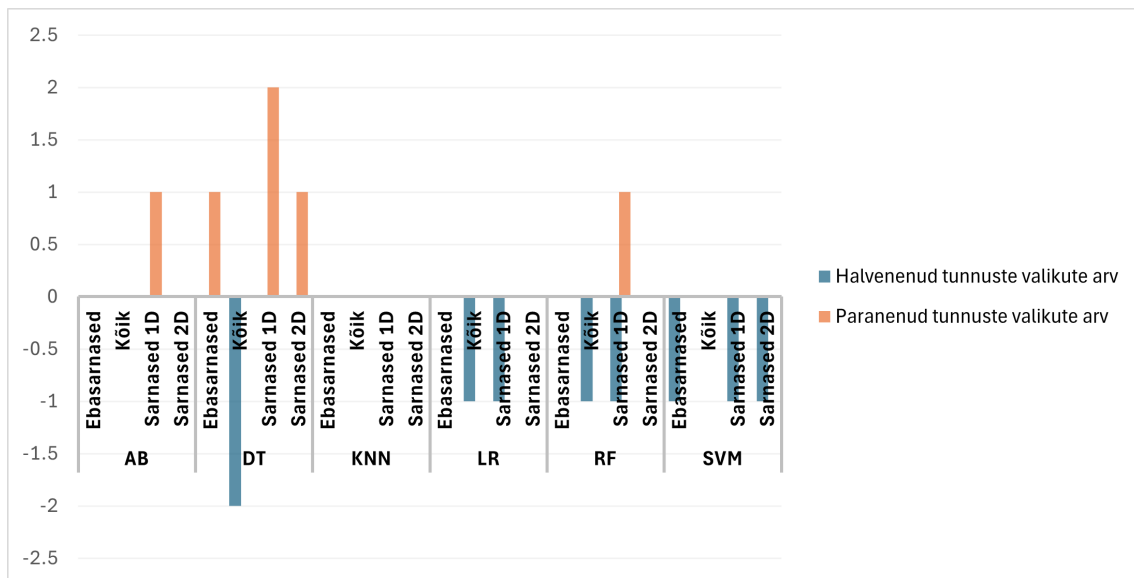
Joonis 15. Roteerimise ja skaleerimisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemuste võrdlus esialgse andmehulga põhjal klassifitseerimisega.



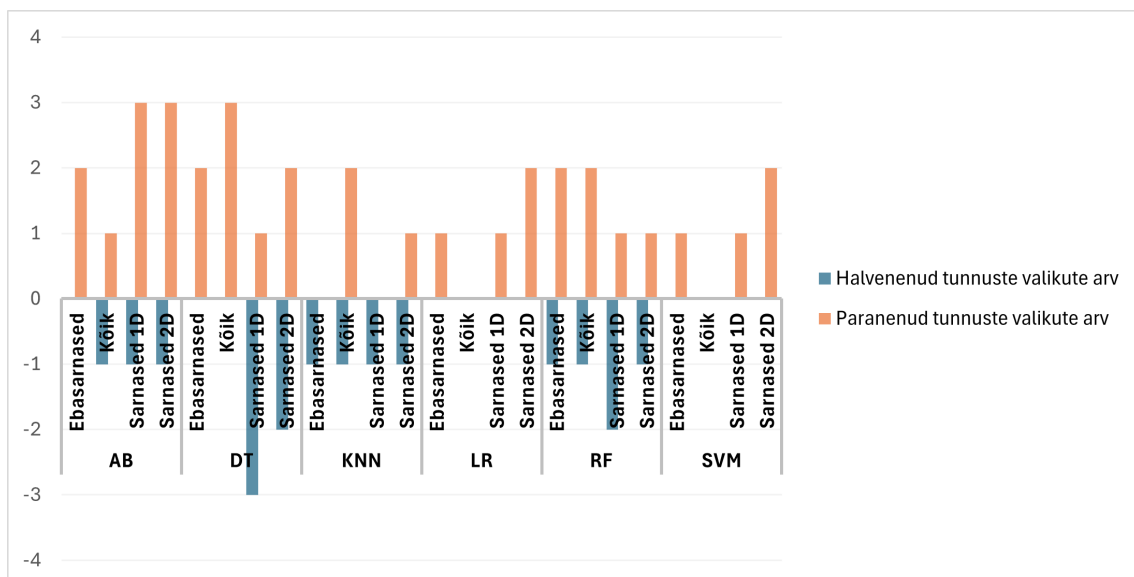
Joonis 16. Esialgse andmehulga lisamisel roteeritud ja skaleeritud andmehulkadesse saadud klassifitseerimistulemuste võrdlus ainult roteerimise ja skaleerimise rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemustega.



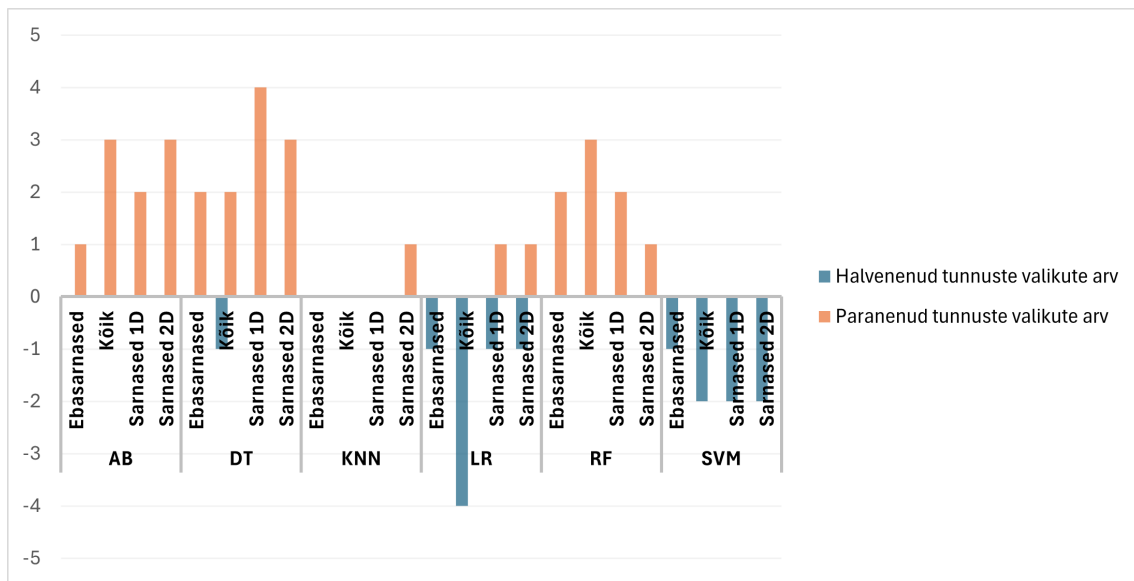
Joonis 17. Esialgse andmehulga tasakaalustamise rakendamisega suurendatud andmehulga põhjal saadud klassifitseerimistulemuste võrdlus esialgse andmehulga põhjal klassifitseerimisega.



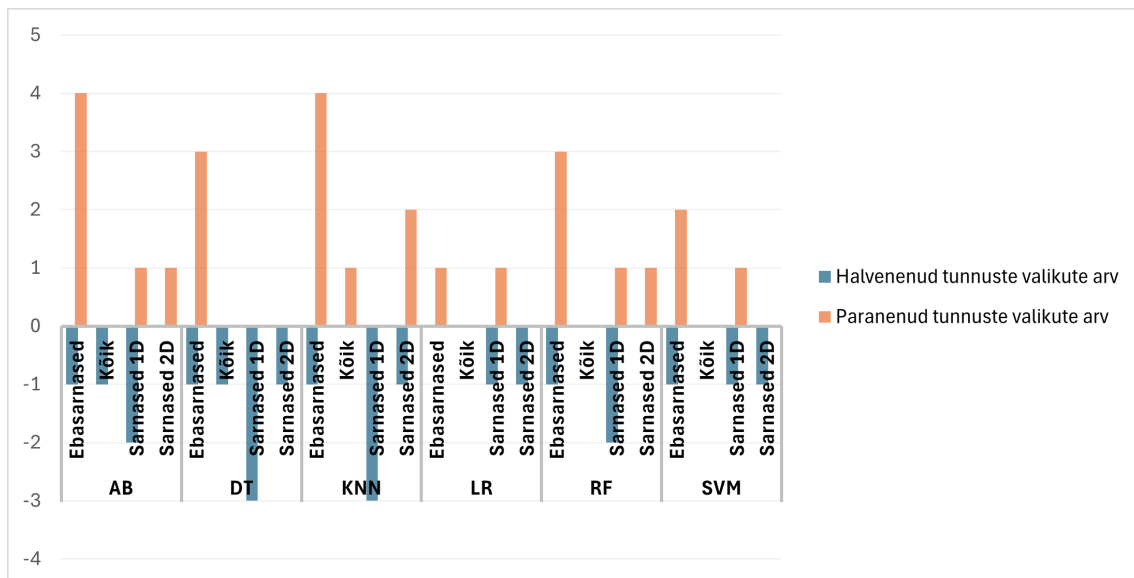
Joonis 18. Roteerimise ja SMOTE-i rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemuste võrdlus esialgse andmehulga põhjal klassifitseerimisega.



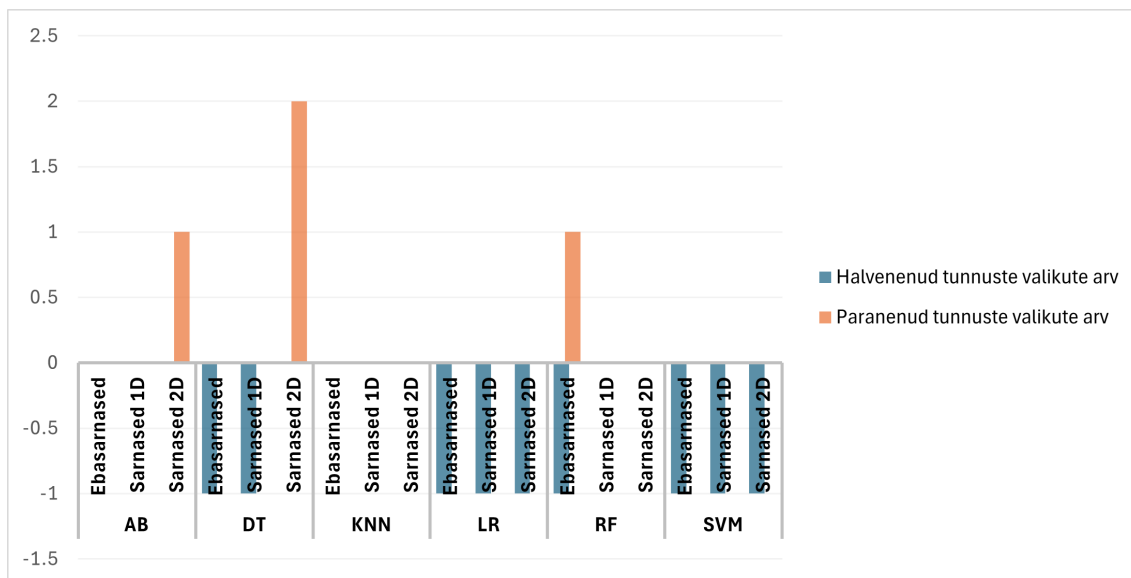
Joonis 19. Esialgse andmehulga lisamisel roteeritud ja tasakaalustatud andmehulkadesse saadud klassifitseerimistulemuste võrdlus ainult roteerimise ja SMOTE-i rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemustega.



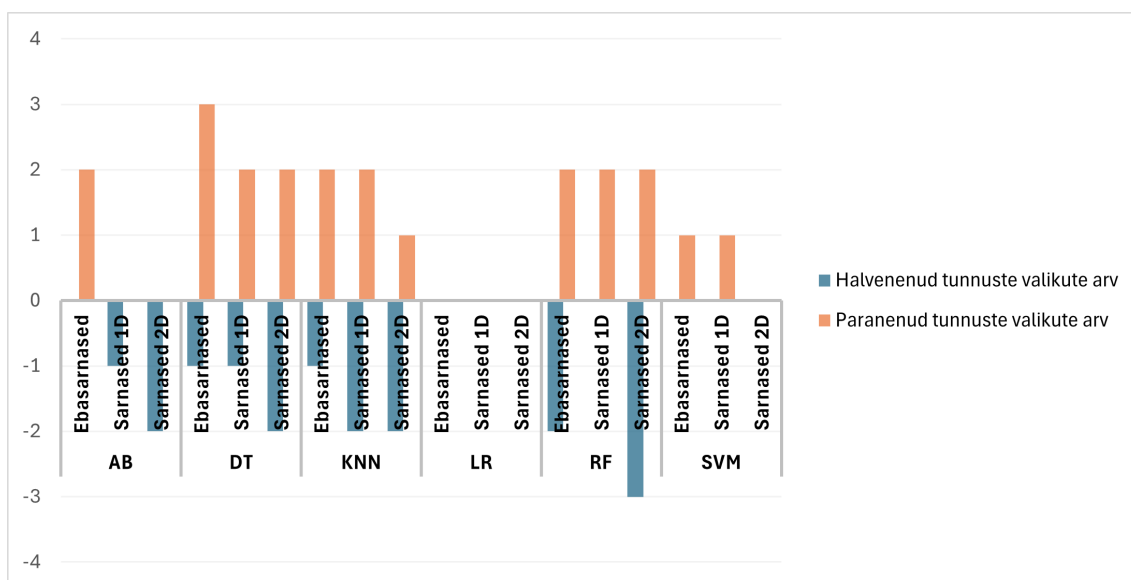
Joonis 20. Skaleerimise ja SMOTE-i rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemuste võrdlus esialgse andmehulga põhjal klassifitseerimisega.



Joonis 21. Esialgse andmehulga lisamisel skaleeritud ja tasakaalustatud andmehulkadesse saadud klassifitseerimistulemuste võrdlus ainult skaleerimise ja SMOTE-i rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemustega.



Joonis 22. Roteerimise, skaleerimise ja SMOTE-i rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemuste võrdlus esialgse andmehulga põhjal klassifitseerimisega.



Joonis 23. Esialgse andmehulga lisamisel roteeritud, skaleeritud ja tasakaalustatud andmehulkadesse saadud klassifitseerimistulemuste võrdlus ainult roteerimise, skaleerimise ja SMOTE-i rakendamisega suurendatud andmehulkade põhjal saadud klassifitseerimistulemustega.