TARTU ÜLIKOOL
Arvutiteaduse instituut
Andmeteaduse õppekava

Rimmo Rõõm

Bakterite eristamine fluoromeetri spektrist masinõppe abil

Magistritöö (15 EAP)

Juhendajad:
Ott Rebane (PhD)
Anna Aljanaki (PhD)

Tartu 2024

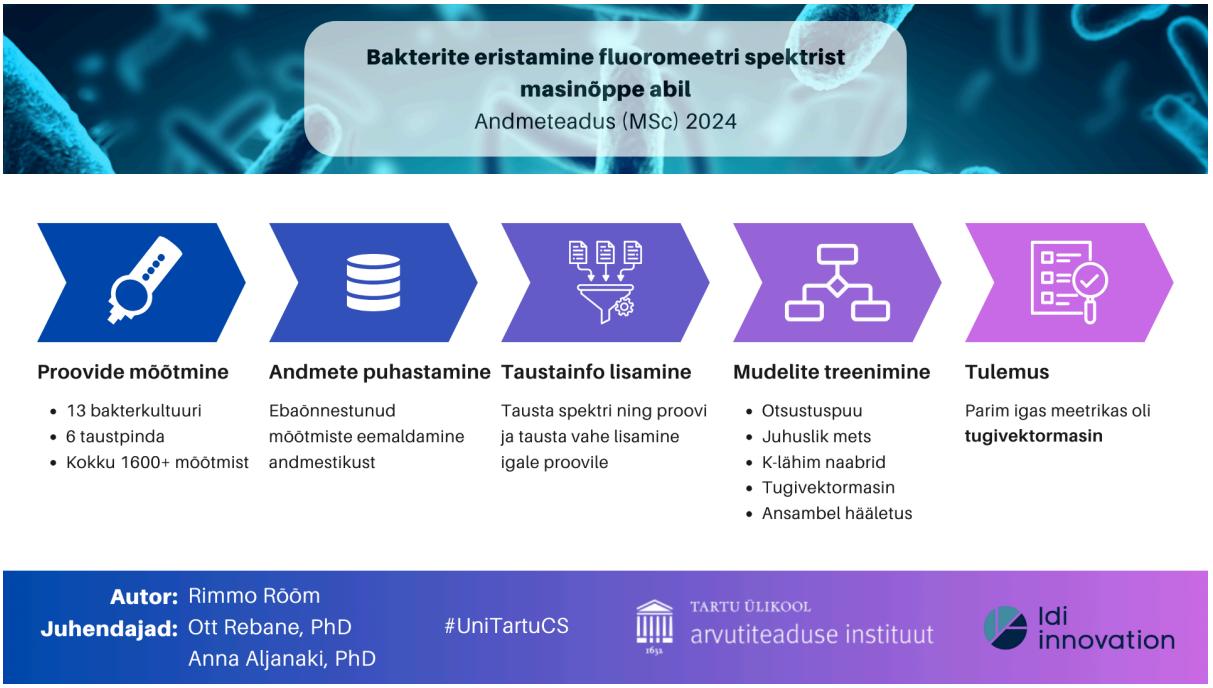
Bakterite eristamine fluoromeetri spektrist masinõppe abil

Lühikokkuvõte:

Magistritöö käigus leitakse sobivaim masinõppe lahendus LDI Innovation OÜ poolt arendatud seadmele H2B-Spectral, mis võimaldaks paremini eristada erinevaid mikroorganisme valitud tahkete pindade peal. Seade töötab mitmekanalilise fluoromeetrina, ergastades mõõtepinda kolme erineva ultravioletse lainepikkusega ning mõõdab emiteeruvat optilist fluorestsents-signaali kolmel erineva lainepikkuste vahemikuga emissiooni kanalil. Saadud kaheksa numbrilise (üks kanal on peegelduskanal ja ei anna infot) hulga põhjal peab sensori tarkvara klassifitseerima mõõtepunkti eelnevalt õpitud klassidesse. Käesoleva töö käigus mõõdetakse üle kolmteist klassi erinevaid mikroorganisme ning võrreldakse erinevaid masinõppemeetodeid (s.h. otsustuspuud, juhuslikud metsad, K-lähimad naabrid, tugivektormasin, ansambel hääletus) nende klassifitseerimiseks. Töö käigus valitud efektiivseim klassifitseerimismeetod leiab kasutust H2B-Spectral standardse masinõppesüsteemi juurutamisel tarkvaras.

Võtmesõnad: Fluoromeeter, fluorestsents, bakterid, masinõpe.

CERCS: P176 Tehisintellekt, P180 Metroloogia, instrumentatsioon, P200 Elektromagnetism, optika, akustika



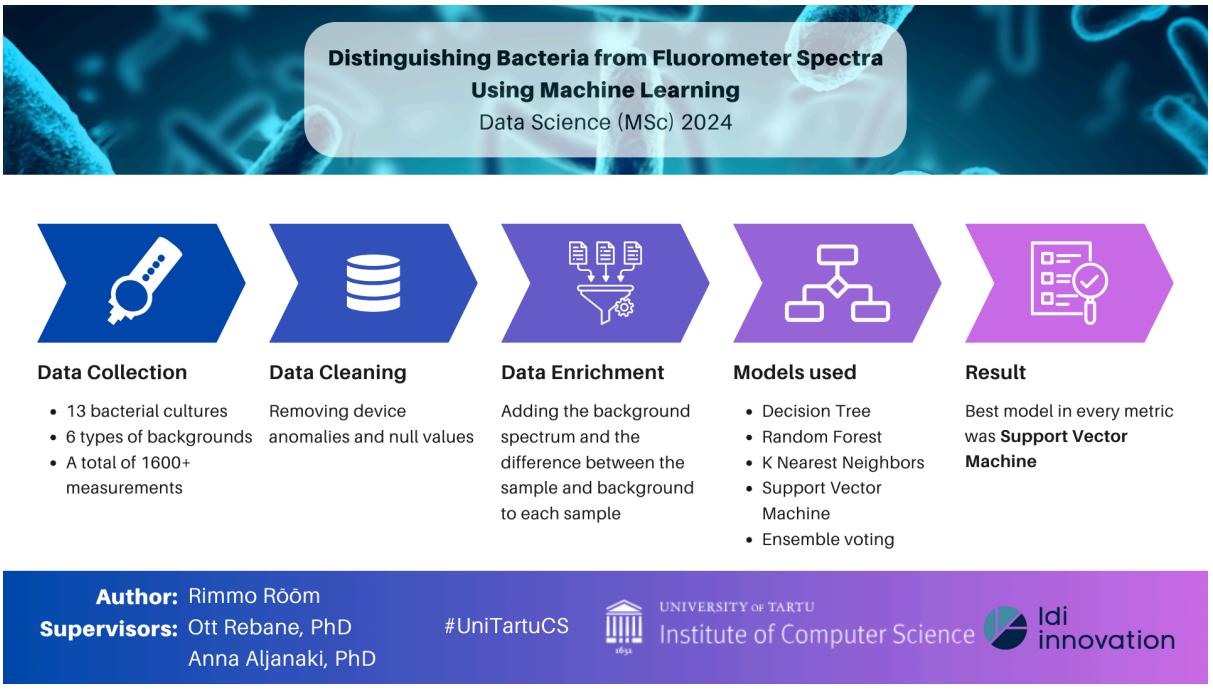
Distinguishing Bacteria from Fluorometer Spectra Using Machine Learning

Abstract:

In this master thesis, the most suitable machine learning solution is found for the fluorometer device H2B-Spectral developed by LDI Innovation OÜ. The machine learning methods tested in this thesis aim to improve the differentiation of various microorganisms on selected solid surfaces. The device functions as a multi-channel fluorometer, exciting the measured sample surface with three different ultraviolet wavelengths and reading the emitted optical fluorescence signal on three different wavelength channels. Based on the obtained eight number data (one channel provides no information), the sensor's software must classify the measurement point into pre-learned classes. In this study, over thirteen classes of various microorganisms are measured, and different machine learning methods (including decision tree, random forest, KNN, support vector machine, ensemble voting) are compared for their classification performance. The most effective classification method identified in this study will be implemented in the standard machine learning system in the software for H2B-Spectral.

Keywords: Fluorometer, fluorescence, bacteria, machine learning.

CERCS: P176 Artificial Intelligence, P180 Metrology, instrumentation, P200 Electromagnetism, optics, acoustics



Sisukord

Sissejuhatus

Mikrobioloogilise saaste (käesoleval juhul täpsemalt "pinna reostatuse") tuvastamine ja haldamine on väljakutse paljudes valdkondades, ulatudes toiduainetööstusest ühistranspordi, haiglate ja laboriteni. Käesoleva kümnendi alguses ühiselt läbitud ülemaailmse COVID pandeemia valguses on üha suurenev tähelepanu suunatud mikroobse saastumise riskidele, mis tekivad siis, kui tahked pinnad puutuvad kokku juhuslikult, looduslikult või tahtlikult ladestunud submikro- ja mikromeetriliste bioloogiliste osakestega. Need osakesed koosnevad valdavalt mikroobidest, mis võivad kiiresti põhjustada muu hulgas biokile teket. Raskesti eemaldatava biokile kiire ja efektiivne tuvastamine on elutähtis, eriti olukordades, kus on kahtlus patogeense mikroorganismi talletumisele ja levikule läbi biokile [1].

Traditsioonilised mikrobioloogilised meetodid, nagu mikrobioloogilise objekti kultuuri kasvatamisel põhinevad testid, nõuavad sageli mitme päeva pikkust inkubeerimist, et tuvastada ja identifitseerida kasvanud mikroorganisme. See ajaaken võimaldab biokile ja saasteainete levikul jätkuda, mis võib raskendada olukorda ja suurendada saastumise ulatust. Viimaste aastate jooksul on arendatud kiiremaid diagnostilisi meetodeid, nagu reaalajas polümeraasi ahelreaktsioon (PCR) meetod ja massispektromeetria [2], mis võimaldavad kiiremini tuvastada patogeene ilma pikaajalise kultiveerimiseta.

Kuigi tänapäevased meetodid nagu PCR ja massispektromeetria võimaldavad tuvastada mikroorganismid mõne tunni jooksul, eeldavad need siiski proovide võtmist. Erinevalt traditsioonilistest meetoditest võimaldab LDI Innovation OÜ poolt arendatud H2B-Spectral, mis töötab mitmekanalilise fluoromeetrina, anda tulemuse mõne sekundi jooksul, ja seda proovi võtmata, kontaktivabalt. See uuenduslik lähenemine võimaldab pindade puhtuse kiiret kontrolli ja vähendab muret mittevajalike proovide võtmise pärast. Käesolevas töös uuritakse, milliseid masinõppe meetodeid saab kõige tõhusamalt rakendada antud seadme andmete analüüsimiseks, et tuvastada ja klassifitseerida erinevaid baktereid, pakkudes seeläbi uudseid lahendusi mikroobse saastumise kiireks tuvastamiseks.

Viimaste kümnendite tehnoloogilised edasiminekud ultravioletsete valgusdioodide tehnoloogias, ülimalt tundlike optiliste detektorite (räni-fotokordistid) ning isegi akutehnoloogias on küll võimaldanud uudsete sensorite arendamist, mis objektispetsiifilist "spektraalset fluorestsents sõrmejälge" [3] kasutades mikrobioloogilisi objekte suudavad käeskantava väikeseadmega tuvastada, aga ilma objektispetsiifilise referentsandmebaasita ning töökindla analüüsimetoodikata on selline lähenemine kasutu. Eelmainitud

tehnoloogiatele tuginedes arendatud H2B-Spectral seade nimelt mõõdab ülinõrkasid optilisi fluorestsents-signaale kõigest kaheksal kanalil ning vastab üldjoontes küsimusele, kas antud pind on mikrobioloogilistest objektidest puhas. Et mõõtetulemused sõltuvad taustpinnast, kasvukeskkonnast, mikrobioloogilise objekti tüübist, eluetapist jne, siis mõõtetulemuste interpreteerimine ilma analüüsietapita pole otstarbekas. Loomulikult, on erinevate mõõtepindade spektrid omavahel väga erinevad, aga erinevad mikrobioloogiliste objektide tüübid (nt pärmid vs bakterid) on omavahel sarnasemad ja erinevad bakterite tüübid veel sarnasemad. Käesolevas töös on põhiliselt uuritud H2B-Spectral seadme mõõtmisvõimekuse piirjuhtu, mis käsitleb sarnaselt kasvatatud ja sarnase kontsentratsiooniga bakteritüvede eristamise võimet mitmetel erinevatel taustpindadel. Võrreldud on klassifitseerimisalgoritme nagu otsustuspuu, juhuslikud metsad, K-lähimad naabrid ja tugivektormasin.

1. Mõisted, terminid ja kasutatavad lühendid

Ergastus - emissiooni maatriks (*ingl. excitation-emission matrix, EEM***)** – efektiivne viis andmete esitamiseks, kus iga rida vastab konkreetsele ergastuslainepikkusele ja iga veerg konkreetsele emissoonilainepikkusele.

Geobacillus stearothermophilus **(GS)** – termofiilne bakter, mis on võimeline kasvama ja paljunema kõrgetel temperatuuridel

Interkvartiilide vahe (*ingl. interquartile range, IQR***)** – statistiline mõiste, mida kasutatakse selleks, et tuvastada kõrvalekaldeid andmestikus. IQR on kolmanda ja esimese kvartiili vahe (Q3 - Q1), mis määratleb andmestiku keskse vahemiku.

Juhuslik mets (*ingl. Random Forest, RF***)** – masinõppe meetod, mis koosneb paljudest otsustuspuudest, mis töötavad kui ansambel. Igale puule antakse juhuslik alamhulk algsetest andmetest ja tunnustest, ja tulemuseks on keskmine või enim hääli saanud klass.

K-lähimate naabrite meetod (*ingl. K-nearest neighbours, KNN***)** – masinõppe meetod, mis klassifitseerib uued objektid vastavalt nende lähimate naabrite enamusklassile treeningandmestikus.

Ristvalideerimine (*ingl. cross-validation***)** – meetod mudeli üldistamisvõime hindamiseks masinõppes. See protseduur hõlmab andmestiku jaotamist mitmeks alamhulgaks. Tüüpiline meetod on korduv *N*-kordne valideerimine (*ingl*. *k-fold cross-validation*), kus andmestik jaotatakse *N* võrdsesse ossa. Seejärel treenitakse mudel *N*-1 osaga ja valideeritakse järelejäänud ühe osaga. Protsessi korratakse *N* korda, iga kord kasutades erinevat osa valideerimiseks ja ülejäänud osi treenimiseks.

Segadusmaatriks (*ingl. confusion matrix***)** – viis mudeli tulemuste hindamiseks. Maatriksis kuvatakse mudeli ennustuste tulemusi võrreldes tegelike tulemustega.

Spektraalne Fluorestsents Sõrmejälg (*ingl. spectral fluorescence signatures, SFS***)** – meetod bioloogiliste ainete, nagu mikroorganismide, tuvastamiseks nende loomuliku "fluorestsents-sõrmejälje" ehk kindlakujulise ergastus-emissioon-maatriksi põhjal. SFS meetodit kasutatakse meditsiinis, keskkonnauuringutes ja bioturvalisuses. [3]

Otsustuspuu (*ingl. Decision Tree, DT***)** – masinõppe meetod, mis moodustab mudeli otsustusreeglitest hierarhilises struktuuri, mis sarnaneb puule. Seda kasutatakse nii klassifikatsioonis kui ka regressioonis.

Polümeraasi ahelreaktsioon (*ingl. polymerase chain reaction, PCR***)** – Biokeemiline tehnoloogia, mida kasutatakse DNA järjestuste paljundamiseks ja võimendamiseks, et võimaldada detailsemat analüüsi ja uuringuid.

Tugivektorklassifikaator (*ingl. Support Vector Classification, SVC***)** – SVM-i variant, mis on spetsiaalselt kavandatud andmeklassifikatsiooniks. See leiab andmekategooriate vahelise optimaalse hüpertasandi vt. ka tugivektormasinad.

Tugivektormasin (*ingl. Support Vector Machine, SVM***)** – masinõppe meetod, mida kasutatakse klassifitseerimis- ja regressiooniprobleemide lahendamiseks. SVM töötab andmete esitamise kaudu kõrgemates mõõtmetes, et leida parim eraldaja.

Ultraviolettkiirgus (*ingl. ultraviolet, UV***)** – Elektromagnetilise spektri osa, mis jääb nähtava valguse ja röntgenkiirguse vahele.

2. Teoreetiline taust

2.1 Andmeteaduse ja Masinõppe Ülevaade

Andmeteaduse (ingl. *data science*) **mõiste** ümbritseb erinevaid protsesse, mis on suunatud andmete analüüsimisele ja nende põhjal strateegiliste otsuste langetamisele, et täita konkreetseid eesmärke. Eesti andmeteaduse kommuun määratleb andmeteadust kui valdkonda, mis koondab endas erinevaid tegevusi, mille eesmärk on andmete põhjal kasulike otsuste tegemine[1].

Masinõpe (ingl. ML, *machine learning*) keskendub algoritmide ja statistiliste mudelite arendamisele. Masinõpet võib defineerida kui arvuti võimet õppida ilma otseste juhisteta [4]. Selle asemel analüüsivad arvutid andmestikke, et tuvastada mustreid ja seoseid, mis aitavad ennustada või klassifitseerida uusi andmeid.

Masinõpe töötab andmekogumiga, püüdes ise leida korrelatsioone ja mustreid, mis toetavad uute situatsioonide mõistmist ja ennustamist. Masinõpet on võimalik liigitada mitmesse kategooriasse: juhendatud, juhendamata ja stiimulõpe[2]. Nendes kategooriates optimeerivad arvutid mudeli parameetreid andmete põhjal eesmärgiga parandada mudeli tulemuslikkust. Mudeli efektiivsust hinnatakse testandmete abil.

Juhendatud õpe (*ingl. supervised learning*) on kõige levinum masinõppe meetod. Seda kasutatakse näiteks igapäevaste e-kirjade seast rämpskirjade filtreerimiseks või meditsiinilistes diagnoosimise süsteemides. Selles õppe vormis moodustavad andmed x ja y paare, eesmärgiga ennustada y silte x tunnuste põhjal. Sisendid x võivad olla klassikalised vektorid kui ka keerukamad objektid nagu dokumendid, pildid, DNA järjestused või graafid. Sarnaselt võib ka siltidel y olla erinevaid struktuure. Lihtsam on binaarne kuju (näiteks "rämpskiri" või "ei ole rämpskiri"), mitmeklassiline klassifikatsioon (kus y võtab ühe K sildist) või lausa mitmesildiline klassifikatsioon (kus y on samaaegselt märgistatud mitme K sildiga)[5].

Juhendamata õpe (ingl. *Unsupervised learning*) on masinõppe meetod, kus mudel töötab andmetega, mis ei sisalda eelnevalt määratud sildistusi. Selle asemel püüab mudel ise leida andmetest struktuuri, kasutades näiteks klastrianalüüsi või mõõtmete vähendamise tehnikaid [5]. Üheks selliseks näiteks on peakomponentanalüüs (*ingl. Principal Component Analysis,*

[1] Mis on andmeteadus? | Data Science Estonia. http://datasci.ee/sissejuhatus/mis-on-andmeteadus (14.04.2024).
[2] Närvivõrkude ja masinõppe sõnastik. | Data Science Estonia. http://datasci.ee/masinoppe-sonastik/ (14.04.2024).

PCA), mis on lineaarne dimensioonide vähendamistehnika, püüdes olemasolevate andmete variatsiooni maksimeerida.

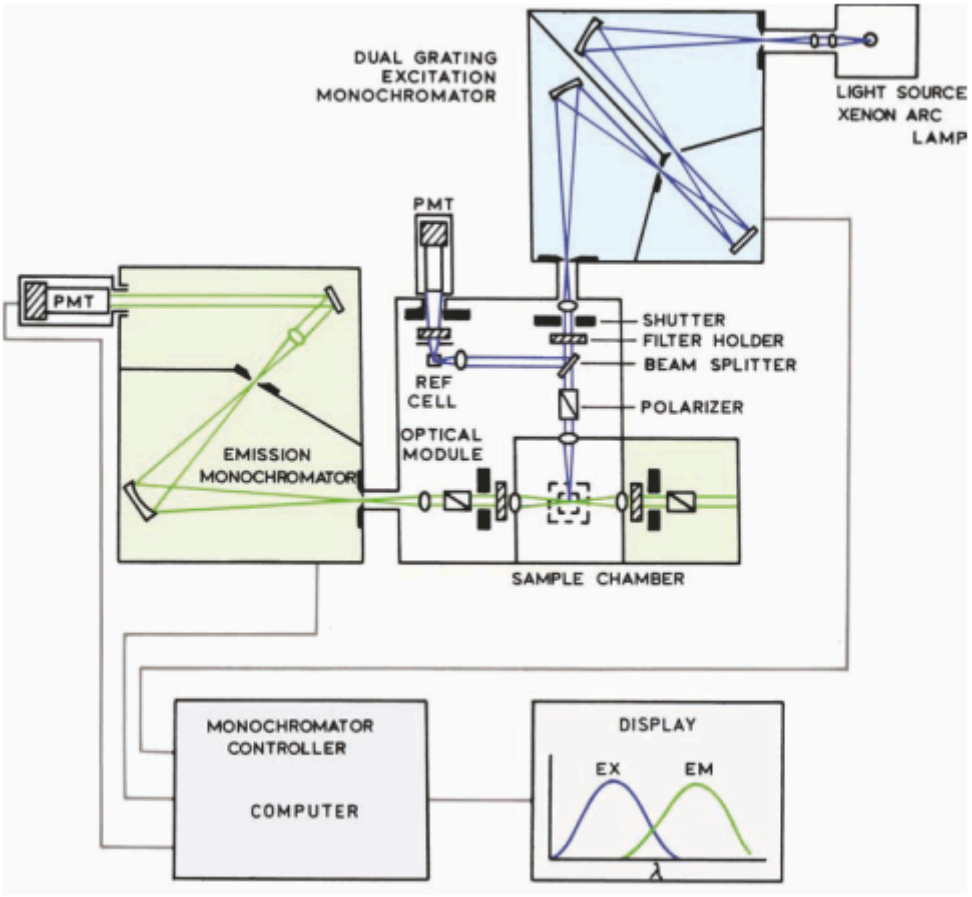
Peter Flach ütleb oma 2012 a. raamatus: "*Machine learning is all about using the right features to build the right models that achieve the right task.*" [6]. Selline määratlus on eriti asjakohane juhendatud õppe kontekstis, nagu ka käesolevas magistritöös, kus andmed on sildistatud eesmärgiga ennustada tundmatuid andmeid.

2.2 Fluorestsents

Fluorestsents on füüsikaline nähtus, kus aine neelab valgust (s.h. ultravioletset) ühel lainepikkusel ja kiirgab valgust teisel, tavaliselt pikemal lainepikkusel. See nähtus on **lühiajaline**, kestes tavaliselt nanosekundeid ("fluoresestsents") pärast ergastava valguse eemaldamist [7]. Nanosekond-ajaskaalas fluorestsents koos mikrosekund-ajaskaala fosforestsentsiga on mõlemad fotoluminestsentsi nähtused ("foto" - valguse poolt põhjustatud), mida käesolevas töös käsitletav fluoromeeter mõõdab, keskendudes fluorestsents-signaalile.

2.2.1 Fluorestsents-spektroskoopia

Erinevate ainete ja materjalide fluorestsentsi uuritakse **fluorestsents-spektroskoobiga** (ka mõnikord nimetatud spektrofluoromeetriks või spektrofluorimeetriks), mille põhimõtteline skeem on näha joonisel 2.2.1. Taoline instrument mõõdab muudetava lainepikkusega valgusega ergastatud aine (proovi) poolt emiteeritud valguse intensiivsust erinevatel lainepikkustel ja mõõtmise tulemuseks on fluorestsentsi spekter (maatriks), mis näitab fluorestsentsi intensiivsust erinevatel lainepikkuste paaridel, nii et iga mõõtepunkt on määratud tema ergastuslainepikkuse, emissioonilainepikkuse ning selle paari korral mõõdetud intensiivsusega.



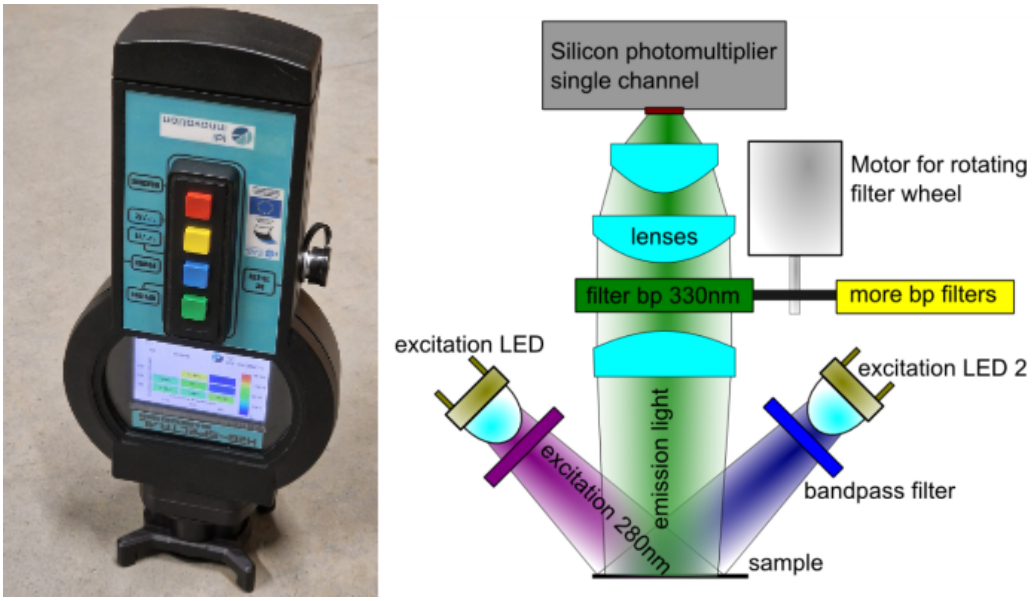
Joonis 2.2.1. Fluorestsents-spektroskoobi põhimõtteskeem [7].

2.2.2 Fluoromeeter

Fluoromeeter on sageli lihtsam seade, mida kasutatakse fluorestsentsi intensiivsuse mõõtmiseks ette määratud lainepikkustel, sageli on need lainepikkused fikseeritud ja seadmespetsiifilised. Nii töötab näiteks hapnikusisalduse andur veekeskkonnas [8] või õlidetektorid veepinnal [9]. Need seadmed on lihtsamad ja neid kasutatakse juhtudel, kui detailne spektraalanalüüs ei ole vajalik. Käesolevas töös kasutatavat sensorit võiks nimetada **mitmekanaliliseks fluoromeetriks**, kuna see jääb lihtsa fluoromeetri ja fluorestsents-spektroskoobi vahepeale.

H2B-Spectral on LDI Innovation OÜ poolt välja töötatud optiline seade (joonis 2.2.2.1), mis põhineb mittekontaktsel mõõtmisel ja on peamiselt mõeldud mikroorganismide ning bioloogilist päritolu reostuse tuvastamiseks erinevatel tahketel pindadel. Seade toimib mitmekanalilise fluoromeetrina, mille lainepikkuste valik põhineb fluorestsents-spektroskoobiga tehtud relevantsete mõõteobjektide spektraalsel analüüsil

(joonis 2.2.2.1). Mõõtepinda ergastatakse järgemööda kolme erineva ultravioletse lainepikkusega (280 nm, 310 nm ja 340 nm), mõõtes emiteeruvat optilist fluorestsents-signaali kolmel erineval kanalil (340 nm, 405 nm ja 460 nm). Ergastamine toimub **kolme erineva ultravioletse valgusdioodiga** ja sobivate filtrite abil, kuna valgusdioodide spektrid sisaldavad laia spektrit (pikemaid lainepikkusi), mis mõõteobjektilt peegeldudes põhjustaks väärlugemeid sama lainepikkusega loetava fluorestsents-signaali juures. Samal ajal loetakse emissiooni läbi kõrgekvaliteediliste filtrite, mis tagavad, et optiline tihedus teistele lainepikkustele on suurem kui 6 (s.t. 10^6 korda maha surutud) ja ergastuslainepikkuse peegeldus ei mõjuta mõõtetulemust. Footoneid loetakse üksikfooton-lugemise režiimis, kasutades kaasaegset räni-fotokordisti skeemi, mis on klassikalisest fotoelektronkordistist suurusjärgu võrra odavam ning väiksem.

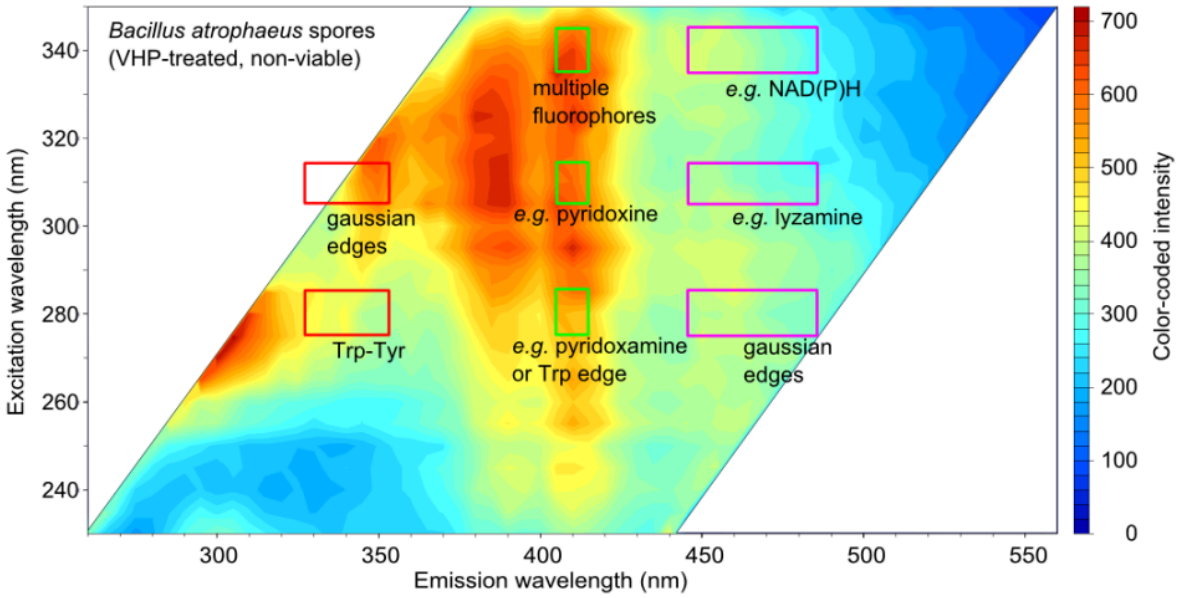


Joonis 2.2.2.1. H2B-Spectral fluoromeeter (vasakul), seadme diagramm [10] (paremal).

2.2.3 Spektraalse Fluorestsentsi Sõrmejälg ehk ergastus - emissiooni maatriks

Fluorestsents-spektroskoopia andmeid on efektiivne esitada ergastus-emissiooni maatriksite (EEM) kujul, mis võimaldavad visualiseerida ja analüüsida valguse ergastamise ja emissiooni interaktsioone. Joonisel 2.2.3.1 on kujutatud SFS seadme tüüpiline ergastus-emissiooni maatriks, mis näitab, kuidas erineva lainepikkusega valgus ergastab proovi, tuues esile iseloomulikud fluorestsentsi piirkonnad. H2B-Spectral seadme jaoks sobilikud lainepikkused on valitud just selle EEMi analüüsi põhjal, kus mikroobide fluorestsents on kõige intensiivsem ning indikatiivne mikrobioloogiliste objektide olemasolule. Selline andmete

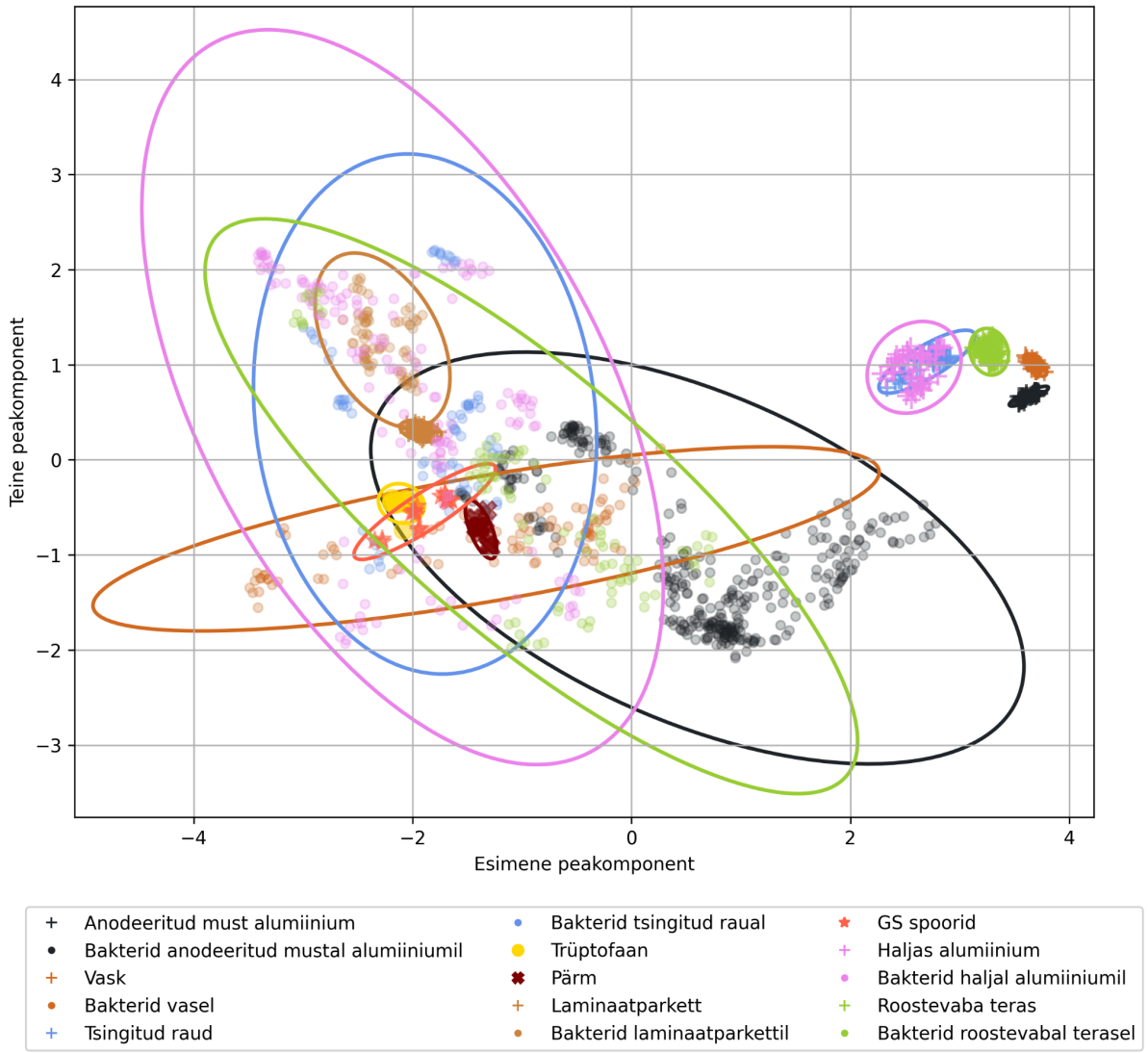
esitlus aitab määrata optimaalsed ergastus- ja emissioonilainepikkused, mis on kriitilised mikroorganismide efektiivseks tuvastamiseks.



Joonis 2.2.3.1. H2B-Spectral seadme jaoks valitud lainepikkused, SFS spektrist. Gaussi äärteks nimetatud spektraal alad ei ole valitud ühelgi kindlal põhjusel, vaid esinevad, muude valikute tõttu kuid sisaldavad siiski kasulikku teavet [10].

2.3 Varasemad mõõtmised H2B-Spectral seadmega

H2B-Spectral seadmega on varasemalt mõõdetud *Geobacillus stearothermophilus* (GS) spoore, trüptofaani ja pärmi. Joonisel 2.3.1 on kujutatud peakomponentanalüüsi tulemust, mis sisaldab antud töö raames mõõdetud bakterite proove erinevatel pindadel, koos pindadega ja GS spooride, trüptofaani ning pärmi mõõtmisi.



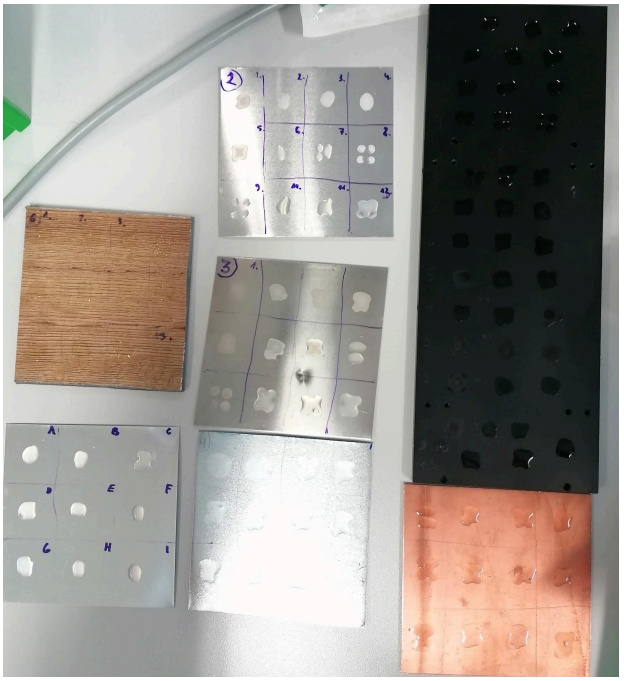
Joonis 2.3.1. Peakomponentanalüüs kuue erineva pinnaga, bakterid erinevatel pindadel ning spoor, seen ja trüptofaan haljal alumiiniumil.

3. Andmed ja metoodika

3.1 Andmete kogumine ja ettevalmistamine

3.1.1 Mikroorganismide ettevalmistus

Töös kasutatud mikroorganismid olid bio-ohutuse taseme 1 ehk inimesele ohutud bakterid, mis olid kasvatatud ja ettevalmistatud Tartu Ülikooli Tehnoloogia Instituudi antimikroobsete ainete laboris. Kõik bakterid on kasvatatud lüsogeense puljongi (ingl. *Lysogeny Broth, LB*) kasvulahuses, mis on levinud vedel toitekeskkond bakterite kasvatamiseks laboritingimustes. Seejärel on bakterid lahjendatud fosfaatpuhvris ühesuguse optilise tiheduse juurde, mis võimaldab eeldada väga sarnaste kontsentratsioonide saavutamist pindadel. Igast lahusest tilgutati taustmaterjalidele täppispipetiga 80 mikroliitrit lahust, millel lasti enne mõõtmiste alustamist täielikult pindadele ära kuivada (joonis 3.1.1.1).

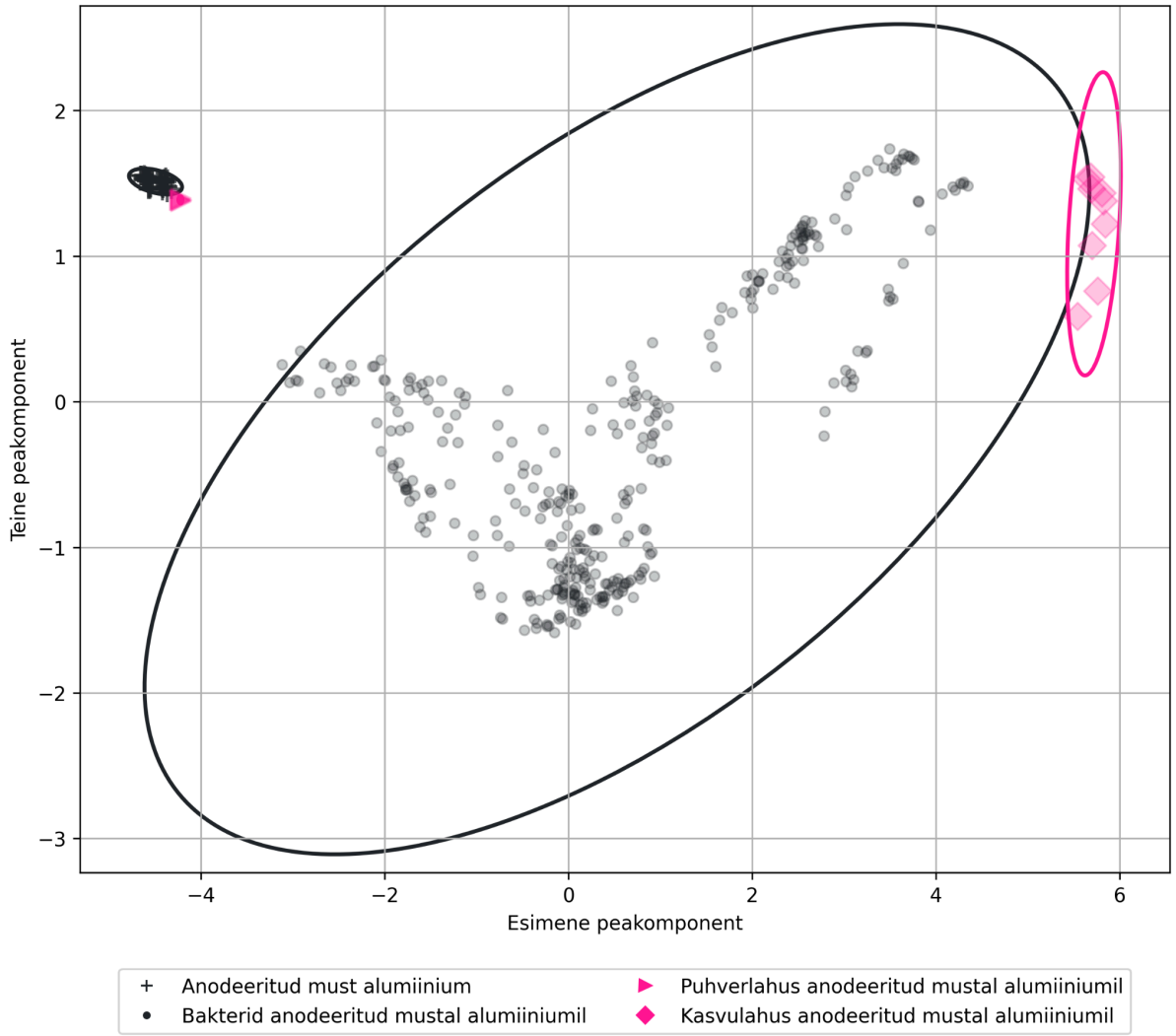


Joonis 3.1.1.1. Bakteriproovid kuivamas taustmaterjalidel. Laminaatparkett (vasak üleval), haljas alumiinium (vasakul all ja keskmises tulbas üleval), roostevaba teras (keskmises tulbas keskel), tsingitud raud (keskmises tulbas all), mustaks anodeeritud alumiinium (paremal üleval), vask (vasakul all).

Kasutatud bakterid hõlmasid mitmeid erinevaid *Escherichia coli* tüvesid (BL21(DE3), DH5, Nissle V1, Nissle V2), *Pseudomonas putida* ja *Pseudomonas stutzeri* bakteriliike, samuti

Staphylococcus epidermidis ja *Shewanella oneidiensis*-i. Lisaks olid kasutusel ka *Bacillus subtilis* ja tema tüvede isolaadid numbritega: 240, 241, 244 ja 245. Kasutatud bakterid olid valitud kättesaadavuse ja eksperimentaalse ohutuse tõttu.

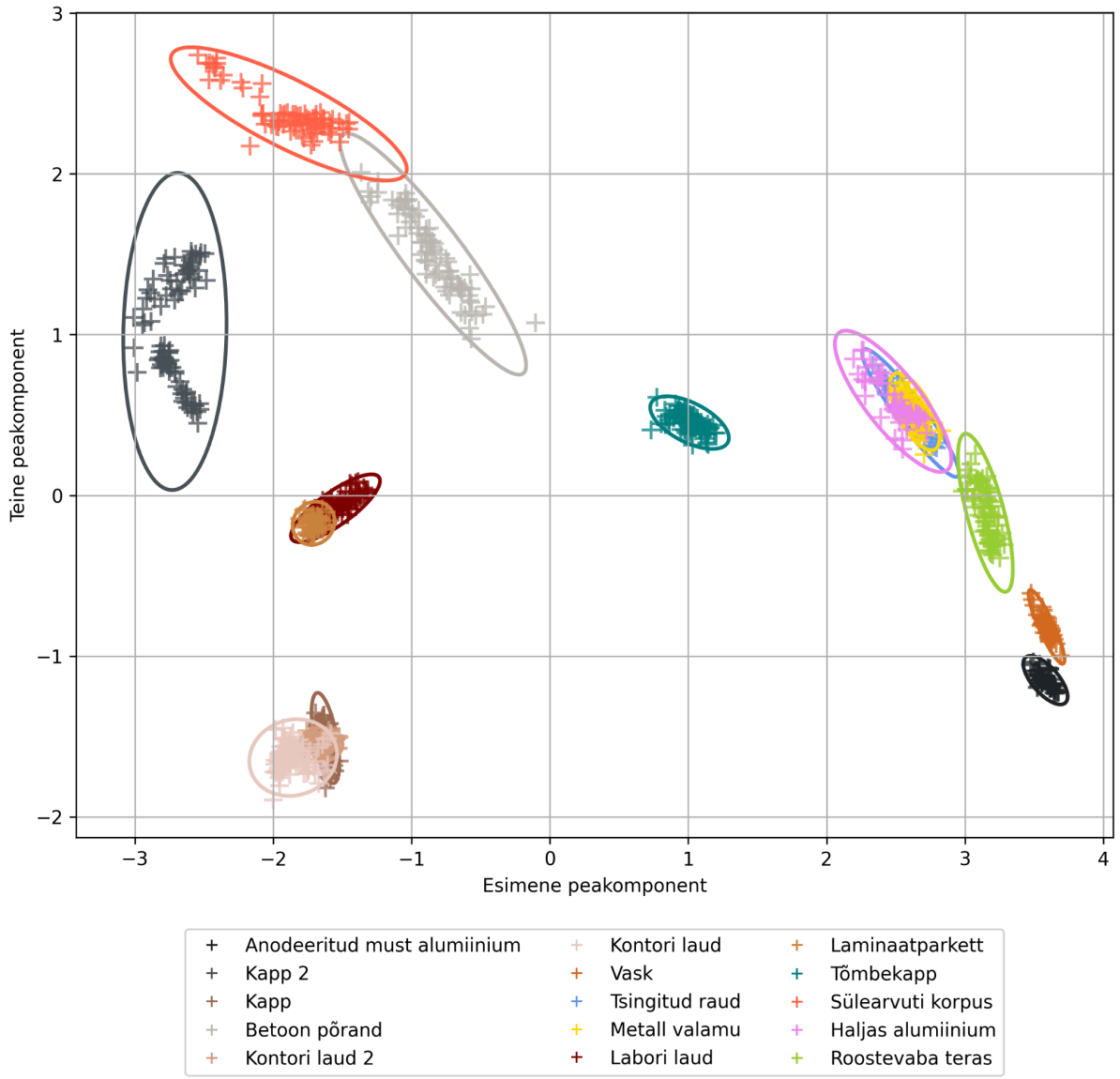
Enne bakterite eristamisele keskendumist mõõdeti töö käigus üle ka kasvu- ning puhverlahuste spektrid mustaks anodeeritud ja haljal alumiinium, tsingitud raud ja laminaatparkett pindadel. Peakomponentanalüüsi tulemuste visuaalse kuvamisega veenduti, et kasvulahus ja puhverlahus tõesti erinevad bakteriproovidest olulisel määral. Joonisel 3.1.1.2 on näha peakomponentanalüüsi kahemõõtmelisi tulemusi mustaks anodeeritud alumiiniumi kohta koos sellel materjalil mõõdetud bakteriproovide ning kasvu- ja puhverlahuse proovide vahel.



Joonis 3.1.1.2. Peakomponentanalüüs mustaks anodeeritud alumiinium pinnaproovide, sellel mõõdetud bakteriproovide ning kasvu- ja puhverlahuse proovide vahel.

3.1.2 Taustmaterjalid

Mõõtmisi tehti kuuel erineval taustmaterjalil: mustaks anodeeritud alumiinium, haljas alumiinium, roostevaba teras, tsingitud raud, vask ning laminaatparkett. Enamusele taustmaterjalidele mahtus 12 mõõtmiseks sobilikku pindala, millel katsealuseid bakteriproove probleemivabalt mõõta. Musta anodeeritud alumiiniumi pindala oli suurem ja sellele mahtusid ära kõigi bakterite proovid kolme paralleelina (joonis 3.1.1.1). Taustmaterjalidele lisaks on seadmega mõõdetud muud tavalised kontoris või laboris esinevad taustpinnad, demonstreerimaks kuivõrd hästi on eristuvad nende spektrid võrreldes sarnaste bakterite eristamisega (joonis 3.1.2.1).



Joonis 3.1.2.1 Erinevate kontori, labori ja bakterite mõõtmise pindade eristus peakomponentanalüüsiga.

3.1.3 Kogutud andmed ja töötlemine

Kokku sai tehtud 2678 mõõtmist H2B-Spectral seadmega. Nende mõõtmiste sisse on arvestatud bakterite, sööda- ja puhverlahuste mõõtmised ning erinevate bakterite, kontori ja labori pindade mõõtmised. Konkreetse töö jaoks olulised mõõtmised on bakterid ja pinnad, millel bakterite proovid olid, kokku 1571 mõõtmist. Tabelis 3.1.3.1 ja 3.1.3.2 on väljatoodud bakterite ja pindade tunnus arv, mida kasutatakse ka tulemuste esitamisel ning mõõtmiste arv enne ja peale eeltöötlust. Ühe proovi kohta tehti vähemalt kümme mõõtmist. Iga mõõtmise järel pöörati seadet, et imiteerida uue proovi mõõtmist.

Tabel 3.1.3.1. Bakteri mõõtmiste jaotus üle kõigi pindade enne ja peale andmete puhastamist

Bakteri nimetus	**Bakteri tunnus**	**Mõõtmiste arv**	**Mudelites kasutatav mõõtmiste arv**
Escherichia coli BL21, DE3	1	97	71
Escherichia coli DH5	2	98	73
Escherichia coli Nissle V1	3	105	75
Escherichia coli Nissle V2	4	98	73
Pseudomonas putida	5	95	64
Pseudomonas stutzeri	6	98	75
Staphylococcus epidermidis	7	98	72
Bacillus subtilis	8	97	62
Bacillus subtilis isolaat 240	9	87	65
Bacillus subtilis isolaat 241	10	88	62
Bacillus subtilis isolaat 244	11	87	63
Bacillus subtilis isolaat 245	12	34	0
Shewanella oneidiensis	13	98	73
	KOKKU	**1180**	**828**

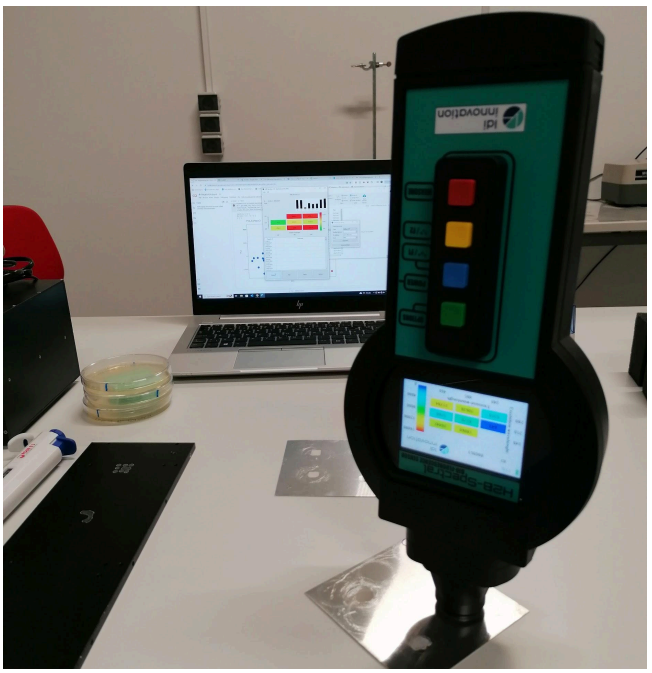
Tabel 3.1.3.2. Pinna mõõtmiste jaotus enne ja pärast andmete puhastamist.

Pinna nimetus	**Pinna tunnus**	**Mõõtmiste arv**	**Mõõtmised, mida kasutati keskmistamiseks**
Anodeeritud must alumiinium	1	52	48
Haljas alumiinium	2	66	52
Roostevaba teras	3	90	78
Tsingitud raud	4	62	47
Vask	5	60	44
Laminaatparkett	6	61	43
	KOKKU	**391**	**312**

H2B-Spectral seadme jaoks on arvutisse paigaldatav tarkvara ja seadme saab läbi WiFi ühendada andmete kogumiseks arvutiga. Tarkvara võimaldab mõõtmisteks luua uue andmebaasi või kasutada juba olemasolevat andmebaasi, kuhu mõõtmisi salvestada. Andmebaasi salvestatakse mõõtmise aeg, kaheksa EEM spektraalkanali intensiivsuse väärtust ja kasutajal on võimalik igale mõõtmisele juurde lisada kommentaar. Käesolevas töös sai kommentaari lahtrisse salvestatud mõõdetava objekti tunnus. Mõõtmiste lõpetamisel on võimalik tarkvarast välja eksportida andmed nii andmebaasi endana kui CSV failina. Töös kasutati CSV faili, mida on lihtne erinevate programmidega lugeda ja töödelda.

Proovide mõõtmise ajal üritati tähele panna ebaõnnestunud mõõtmisi, kus mõni kanal andis nullväärtuseid, ja need kohe kogutavast andmestikust eemaldada. Andmete eeltöötlemisel tuleb siiski veenduda, et andmetesse ei jääks ühtegi proovi, kus oleks nullväärtuseid.

Seadme ja selle andmete kasutamise kogemusest on täheldatud, et mõni EEM väärtus võib olla küllastuses või erineda tunduvalt, kui näiteks väline valgus on sattunud mõõtmise alale. See võis tähendada, et seade ei ole olnud korralikult vastu pinda, et tõkestada välise valguse mõju mõõtmisele. Joonisel 3.1.3.3 on näha seadet ja kuidas sellega mõõtmisi teostati bakteri proovide mõõtmisel. Joonisel 2.2.2.1 on näha seadme mehaanilist edasiarengut, mis tagas stabiilsema püsti seismise ja parema valgustõkestuse pindade mõõtmisel.



Joonis 3.1.3.3. H2B-Spectral seadme kasutus proovide mõõtmisel. Lisaproovide imiteerimiseks pöörati seadet pinna peal ja sooritati uus mõõtmine.

Andmete puhastamisel eemaldati kõik proovid, millel mõni EEM elemendi väärtus oli null. Lõplikust andmehulgast jäeti välja ka lahjendatud kontsentratsiooniga proovid ja alles jäänud andmetele rakendati interkvartiil vahemikku igale pinnale ja bakteri kombinatsioonile eraldi. Interkvartiili vahemik on läbi korrutatud 1.3-ga. Sellel viisil eemaldatakse andmestikust anomaalsed mõõtmistulemused, näiteks sellised, kus mõni EEM elemendi väärtus on tõenäoliselt sattunud küllastusse. Mudelite treenimiseks kasutatavatest andmetest jäeti välja ka bakteri nr 12 mõõtmised, kuna need olid mõõdetud vaid ühel pinnal - mustaks anodeeritud alumiiniumil.

Andmeanalüüs viidi läbi kasutades programmeerimiskeelt Python[3] keskkonnas Google Colaboratory. Pandas[4] ja NumPy[5] teegid olid peamiselt kasutusel andmestruktuuride haldamiseks ja andmete eeltöötluseks, mis võimaldasid efektiivselt käsitleda suuri andmekogumeid ning teostada andmete ühendamise ja transformeerimise toiminguid. Scikit-Learn[6] teek oli kasutusel mudelite treenimiseks ja Matplotlib[7] andmete ja tulemuste visualiseerimiseks.

---

[3] Welcome to Python.org. https://www.python.org/ (14.05.2024).
[4] API reference - pandas 2.2.2 documentation. https://pandas.pydata.org/docs/reference/index.html (14.05.2024).
[5] NumPy reference - NumPy v1.26 Manual. https://numpy.org/doc/stable/reference/ (14.05.2024).
[6] API Reference - scikit-learn 1.4.2 documentation. https://scikit-learn.org/stable/modules/classes.html (14.05.2024).
[7] API Reference - Matplotlib 3.8.4 documentation. https://matplotlib.org/stable/api/index.html (14.05.2024).

3.2 Tunnused

Suurendamaks mudelite treenimise efektiivsust ja täpsust on antud töös bakteri EEM elementidele juurde lisatud tunnuseid, mis jagunes kolmeks etapiks. Esimeses etapis kasutatakse ära pindade mõõtmisi ja igale bakteri mõõtmisele lisatakse juurde vastav keskmistatud pinna EEM elementide väärtused.

Keskmiste arvutamine iga pinna *s* ja iga pinna EEM elemendi *i* kohta:

$$\mu_{s,i} = \frac{1}{n_s} \sum_{j=1}^{n_s} P_{s,i,j} \tag{1}$$

kus *n* on mõõtmiste arv pinnal *s*, $P_{s,ij}$ on *j*-s mõõtmine *i*-nda EEM elemendi väärtuse puhul pinnal *s*.

Teises etapis on rakendatud *one-hot-encoding*-ut pindade kategoriseerimiseks mitmeks binaarseks tunnuseks, et hõlbustada masinõppe algoritmide kasutamist. See tähendab, et algsed pinna kategooriad on jagatud uuteks tunnusteks ja vastav tunnus on 1, kui proov on antud kategoorias ja 0 vastasel juhul. See lähenemine võimaldab masinõppe mudelitel efektiivsemalt õppida ja tuvastada mustreid pindade kategooriate vahel, parandades seeläbi ennustuste täpsust.

Viimases etapis lisati andmestikele juurde EEM väärtused mis kirjeldasid proovi ja pinna erinevust ehk proovi EEM väärtustest lahutati maha pinna EEM väärtused.

Vahe arvutamine iga bakteri *B* EEM elemendi *i* ja keskmistatud pinna EEM elemendi *i* kohta:

$$\Delta_{s,i} = B_{s,i} - \mu_{s,i} \tag{2}$$

kus $\Delta_{s,i}$ on vahe bakteri *i*-nda EEM elemendi väärtuse ja pinna *i*-nda keskmistatud EEM elemendi väärtuse vahel pinnal *s*.

3.3 Mudelid

Masinõppe mudelite treenimise protsess antud töös on keskendunud mitmete klassifitseerimise algoritmide kasutamisele ja nende optimeerimisele, et leida H2B-Spectral seadmele parim klassifitseerimise algoritm.

3.3.1 Kasutatud klassifitseerimise algoritmid

Otsustuspuu (*ingl. Decision Tree, DT*) algoritm on valitud oma lihtsa implementatsiooni poolest H2B-Spectral seadet silmas pidades. Otsustuspuud on lihtsad, kuid võimsad klassifitseerimise algoritmid, mis õpivad andmetest otsustusreegleid, et jõuda järeldusteni. Need on intuitiivselt mõistetavad ja lihtsad implementeerida. Iga sõlm puus esindab otsustavat küsimust või omadust ning iga haru tulemust. Leitud tingimusi saaks otse seadme lähtekoodis kasutada.

Juhuslikud metsad (*ingl. Random Forests, RF*) on ansambel-õppe meetod, mis koosneb paljudest otsustuspuudest. Iga puu treenitakse veidi erineva andmekogumi ja tunnuste alamhulgaga, mis aitab vähendada ületreenimist ja parandada mudeli üldist jõudlust. Juhuslikud metsad on eriti tõhusad, kui on vaja töödelda suuri andmekogumeid ja keerukaid omadusi.

K-lähimate naabrite (*ingl. K-Nearest Neighbors, KNN*) algoritm klassifitseerib proovid, leides treeningandmetest k lähimat naabrit ja hääletades nende klasside põhjal. See on lihtne ja efektiivne meetod, mis sobib hästi väikestele andmekogumitele. KNN algoritmi efektiivsus sõltub õigesti valitud kaugusmeetrikast ja *k* väärtusest.

Tugivektormasinad (*ingl. Support Vector Machines, SVM*) on võimas klassifikaator, mis otsib hüperplaani, mis optimaalselt eraldab erinevad klassid. SVM on tuntud oma robustsuse poolest, kui on vaja eraldada keerulisi ja segaseid mustreid.

Töös proovime ka ansambel hääletust (*ingl. Ensemble Voting*) uurimaks, kas mitme mudeli peale on võimalik paremat tulemust klassifitseerimisel saavutada või mitte. Ansambel hääletus on pehmes režiimis (*ingl. Soft Voting*), et arvesse võtta kõiki sisend mudeleid. Lisaks võtab mudel kaaludena arvesse iga mudeli täpsust ehk parema õigsusega mudelil on rohkem kaalu klassi hääletamisel.

3.3.2 Hüperparameetrite optimeerimine ja ristvalideerimine

Hüperparameetrite optimeerimiseks kasutati *GridSearchCV*-d, mis on sklearni raamistikus pakutav meetod. See meetod võimaldab automaatselt katsetada mitmesuguseid hüperparameetrite kombinatsioone, et leida parim mudel. *GridSearchCV* kasutab ristvalideerimist, mis jagab andmed korduvateks treening- ja testkomplektideks, et hinnata mudeli jõudlust erinevatel andmekogumitel.

Iga mudeli jaoks määrati konkreetne hüperparameetrite komplekt, mida optimeeriti, sealhulgas puude arv juhuslikes metsades, naabrite arv KNN-is, ning *C* ja gamma väärtused SVM-is. Mudelite treenimise käigus jälgiti õigsus skoori, et hinnata ja valida parimate hüperparameetritega mudelid.

Mudelite üldise jõudluse hindamiseks kasutati kümnekordset ristvalideerimist. See aitab tagada, et mudeli hinnangud on usaldusväärsed ja et mudel suudab uusi andmeid efektiivselt üldistada. Ristvalideerimise jaotused valiti selliselt, et igas jaotuses oleks esindatud kõik bakterite ja pindade kombinatsioonid, tagades seeläbi iga bakteritüübi ja pinna esindatuse igas jaotuses.

4. Tulemused

Antud peatükis esitame ja analüüsime viie erineva masinõppe mudeli tulemusi, mida rakendati bakterite tuvastamiseks erinevatelt pindadelt. Iga mudeli puhul arvutati mitmeid jõudlusnäitajaid nagu mudeli üldine õigsus ning iga klassi kohta eraldi täpsus, saagis ja f1 skoor. Kõik väärtused on keskmistatud üle ristvalidatsiooni tulemuste, et hinnata mudelite robustsust ja jõudlust erinevates tingimustes.

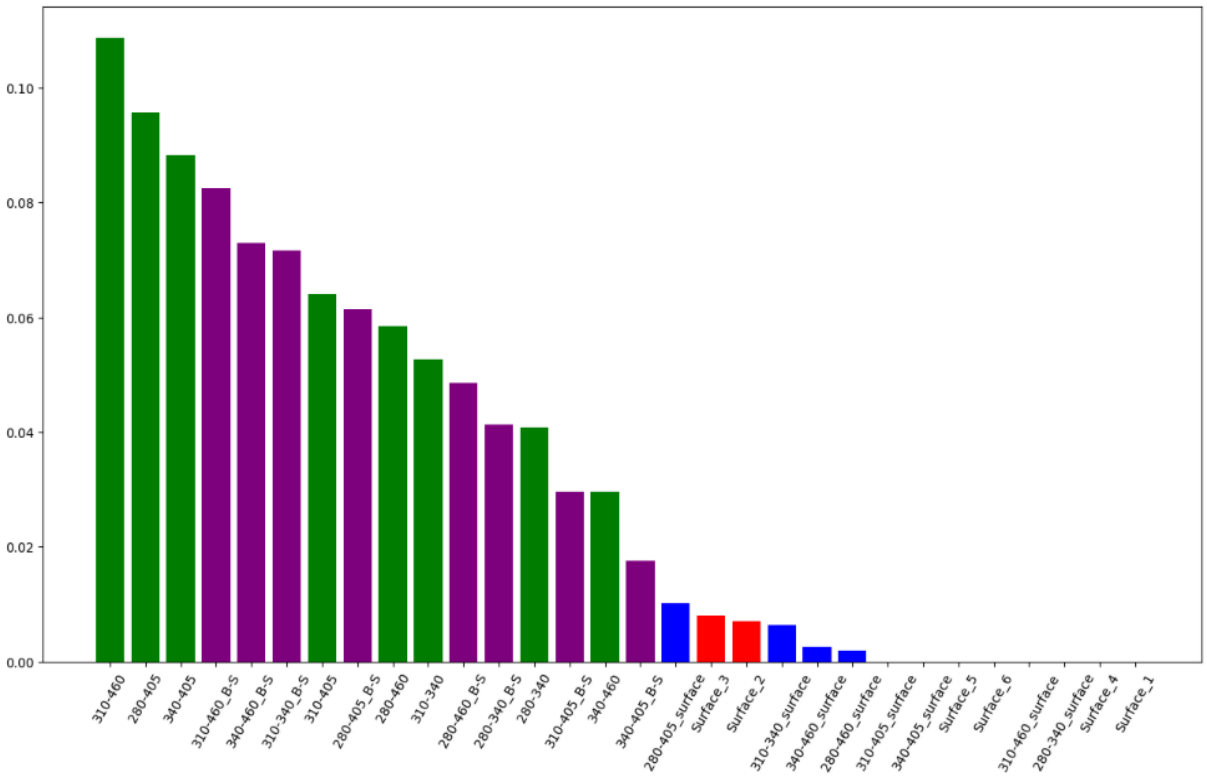
4.1 Otsustuspuu tulemused

Otsustuspuu mudeli keskmine õigsus oli 0.744, mis viitab mõõdukale jõudlusele. Mudeli parimaks tulemuseks oli bakter 8, kus täpsus ulatus 0.924-ni ja saagis 0.835-ni, näidates suurepärast suutlikkust selle klassi ennustamisel. Vastupidiselt, bakter tunnusega 1 näitas kõige madalamat täpsust (0.570) ja saagist (0.516), mis osutab sellele, et mudelil oli raskusi selle klassi õigesti klassifitseerimisega. Klasside põhised tulemused leiab tabelist 4.1.1.

Tabel 4.1.1. Otsustuspuude keskmistatud meetrika skoorid iga bakteri kohta.

Bakteri tunnus	**Täpsus**	**Saagis**	**F1-skoor**
1	0.57	0.516	0.528
2	0.631	0.627	0.613
3	0.709	0.732	0.707
4	0.615	0.624	0.614
5	0.851	0.858	0.85
6	0.782	0.733	0.752
7	0.692	0.734	0.7
8	0.924	0.835	0.87
9	0.843	0.787	0.794
10	0.812	0.85	0.822
11	0.831	0.776	0.787
13	0.879	0.893	0.88

Joonis 4.1.2 esitab otsustuspuu tunnuste tähtsuse, kust ilmneb, et otsustuspuu on oodatult pidanud oluliseks bakterite spektrit. Lisaks on märkimisväärset tähtsust omistatud tunnustele, mis kajastavad informatsiooni bakterite EEM väärtuste ja pinna EEM väärtuste erinevuste kohta.



Joonis 4.1.2. Otsustuspuu tunnuste olulisus. Roheliselt bakterite EEM väärtused, lillalt bakterite EEM ja pinna EEM vahe, siniselt pinna EEM väärtused ja punaselt pinna *one-hot-encoding* tunnused.

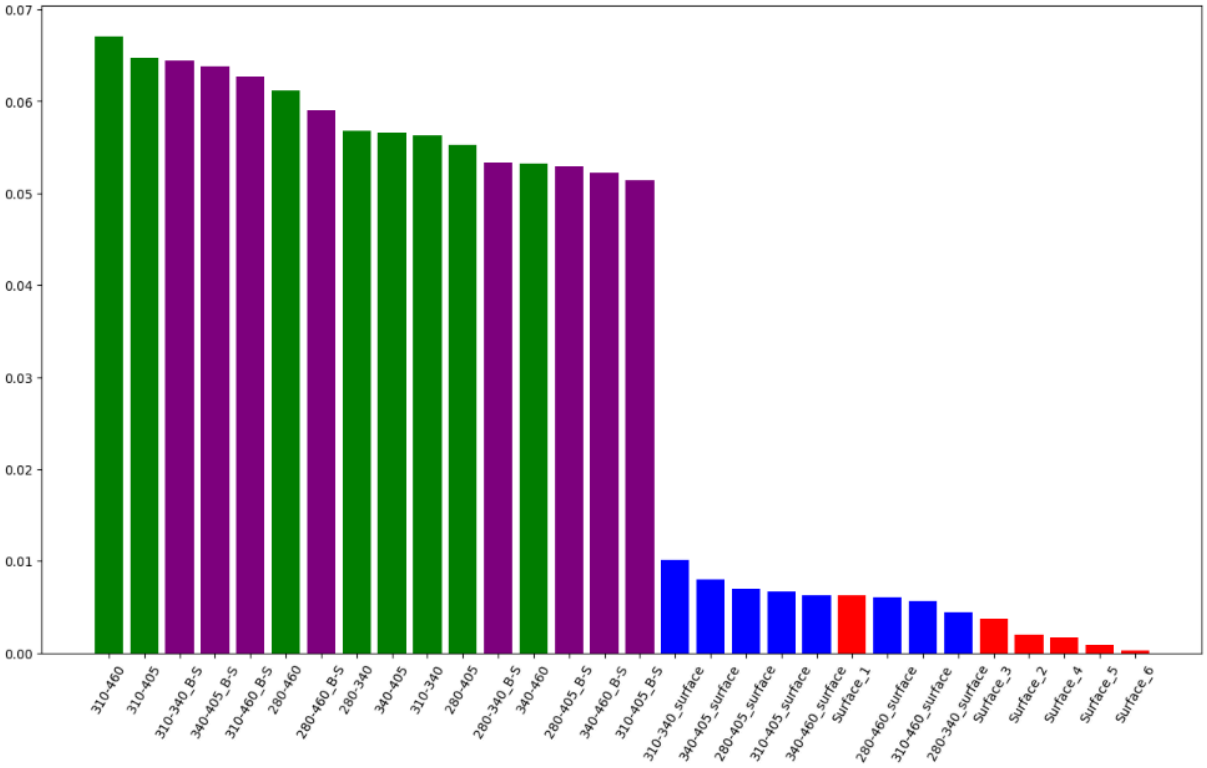
4.2 Juhuslike metsade tulemused

Juhusliku metsa mudeli õigsus oli märkimisväärselt kõrgem otsustuspuu omast, saavutades keskmise õigsuse skoori 0.888. Märkimisväärseks tuleb ka pidada bakteri 13 saagise tulemust 1.000, mis on maksimaalne tulemus ja näitab, et üle terve ristvalideerimise suutis mudel tuvastas kõik selle klassi tegelikud juhud. Kehvemad meetrikad olid aga bakteril 4, mille saagis oli 0.648 ja täpsus 0.853. Klasside põhised tulemused leiab tabelist 4.2.1.

Tabel 4.2.1. Juhuslike metsade keskmistatud meetrika skoorid iga bakteri kohta.

Bakteri tunnus	**Täpsus**	**Saagis**	**F1-skoor**
1	0.817	0.749	0.771
2	0.856	0.828	0.829
3	0.781	0.914	0.837
4	0.853	0.648	0.719
5	0.955	0.957	0.955
6	0.923	0.917	0.917
7	0.893	0.946	0.914
8	0.971	0.968	0.967
9	0.924	0.959	0.937
10	0.904	0.939	0.914
11	0.905	0.85	0.873
13	0.961	1	0.979

Ka juhuslikel metsadel sai vaadatud millised on olulisemad tunnused. Joonisel 4.2.2 on näha, et ka juhusliku metsa korral olulised tunnused on bakterite spektri väärtused ja bakterite EEM väärtuste ja pinna EEM väärtuste vahe. Võrreldes otsustus puuga on need tunnused aga võrdsemalt olulised.



Joonis 4.2.2. Juhusliku metsa tunnuste olulisus. Roheliselt bakterite EEM väärtused, lillalt bakterite EEM ja pinna EEM vahe, siniselt pinna EEM väärtused ja punaselt pinna *one-hot-encoding* tunnused.

4.3 K-lähimate naabrite tulemused

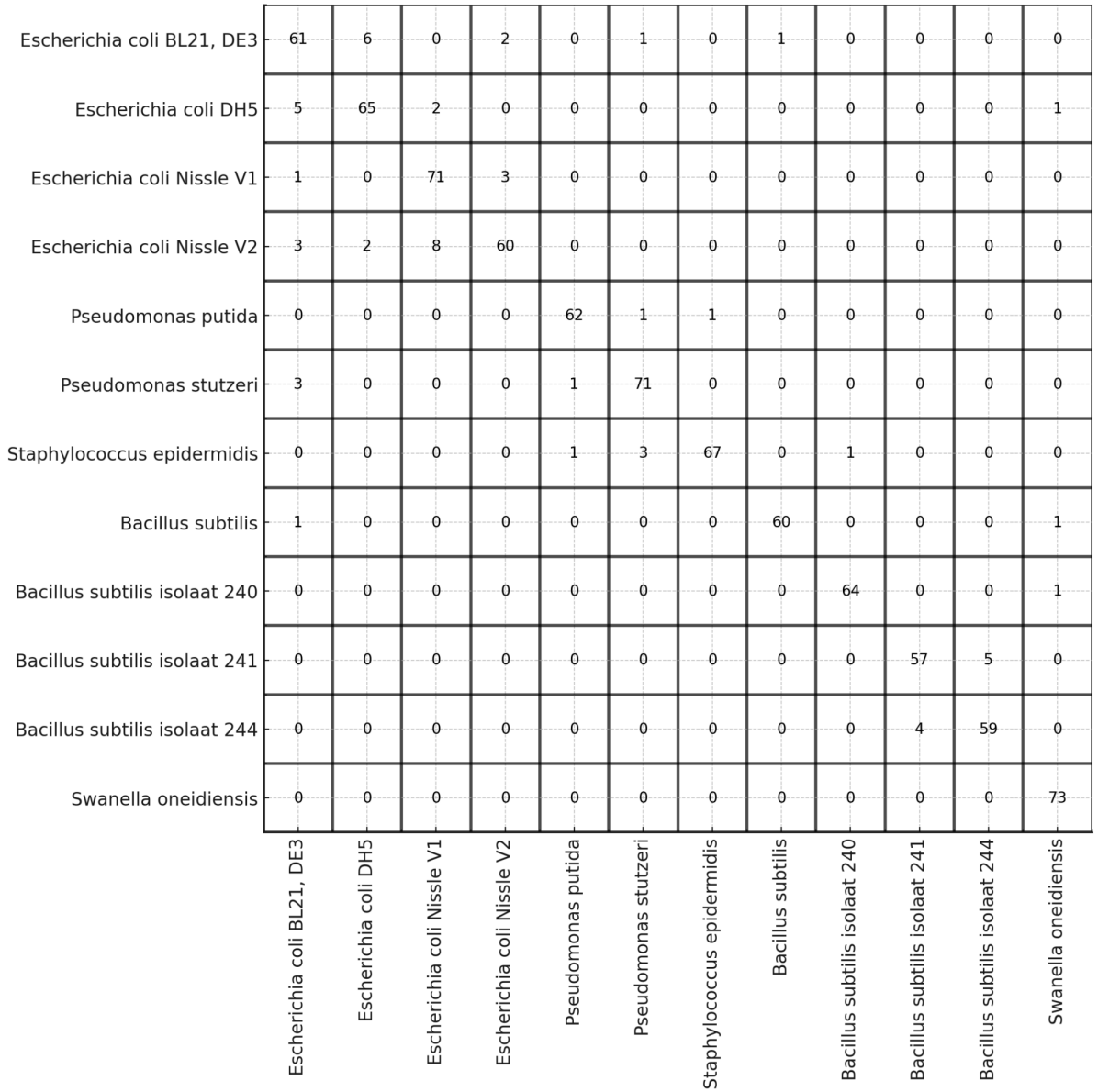
K-lähimate naabrite meetodi keskmine õigsus oli 0.894. Nagu juhusliku metsa puhul, saavutas ka KNN mudel bakteri 13 täiuslikud tulemused (saagis 1.000). Tasub mainimist, et kõigi mudelite seast oli KNN mudeli parameetrid stabiilseimad, kasutades kõigil juhtudel kaalutud kaugusi ja väikest naabrite arvu (3). Samuti varieerus kaugusmõõt "*euclidean*" ja "*manhattan*" vahel. KNN mudeli klassi põhilised meetrikad on esitatud tabelis 4.3.1.

Tabel 4.3.1. KNN mudelite keskmistatud meetrika skoorid iga bakteri kohta.

Bakteri tunnus	**Täpsus**	**Saagis**	**F1-skoor**
1	0.862	0.777	0.801
2	0.854	0.838	0.831
3	0.804	0.85	0.819
4	0.83	0.715	0.762
5	0.988	0.971	0.978
6	0.937	0.94	0.936
7	0.914	0.961	0.934
8	0.958	0.938	0.944
9	0.94	0.961	0.945
10	0.921	0.904	0.906
11	0.905	0.87	0.882
13	0.926	1	0.96

4.4 Tugivektormasina tulemused

Tugivektormasinate kummuleeritud segadusmaatriks (joonis 4.4.1) toob hästi esile mudeli eksimused *Escherichia coli* tüvede eristamisel ja vähesed eksimused erinevate bakterite ja bakteri isolaatide eristamisel.



Joonis 4.4.1 Tugivektormasinate kummuleeritud segadusmaatriks üle ristvalideerimise. *x*-teljel on ennustatud klassid ja y-teljel tegelikud klassid.

SVC mudel näitas kõigist mudelitest kõige kõrgemat keskmist õigsust - 0.930. Nagu juhuslike metsade ja KNN mudelite puhul, siis ka tugivektormasinad said bakteri 13 saagiseks maksimaalse skoori. Tabelist 4.4.2 võib veel tähele panna, et bakteri 4 saagis on oluliselt parem kui teistes mudelites, kuigi ka SVC mudelis jääb ta kõige kehvema saagisega bakteri klassiks.

Tabel 4.4.2. Tugivektormasin mudelite keskmistatud meetrika skoorid iga bakteri kohta.

Bakteri tunnus	**Täpsus**	**Saagis**	**F1-skoor**
1	0.851	0.87	0.848
2	0.902	0.884	0.884
3	0.884	0.938	0.906
4	0.924	0.823	0.867
5	0.973	0.971	0.97
6	0.944	0.942	0.939
7	0.989	0.934	0.956
8	0.989	0.968	0.976
9	0.989	0.988	0.987
10	0.935	0.924	0.926
11	0.924	0.94	0.928
13	0.962	1	0.98

4.5 Ansambel hääletuse tulemused

Ansambel hääletuse tulemused põhinevad kõigil teistel mudelitel, mis sageli osutub täpsemaks kui ükski individuaalne mudel eraldi. Antud töös on näha, et ansambli keskmine õigsus on 0.908, mis on samuti kõrge aga jääb siiski alla SVC mudelile. Tabelis 4.5.1 on välja toodud bakteri klasside meetrikad.

Tabel 4.5.1. Ansambel hääletuse keskmistatud meetrika skoorid iga bakteri kohta.

Bakteri tunnus	**Täpsus**	**Saagis**	**F1-skoor**
1	0.838	0.794	0.809
2	0.895	0.844	0.861
3	0.813	0.901	0.849
4	0.869	0.74	0.791
5	0.973	0.971	0.97
6	0.943	0.957	0.946
7	0.931	0.946	0.934
8	0.989	0.968	0.976
9	0.955	0.958	0.955
10	0.938	0.922	0.926
11	0.924	0.92	0.919
13	0.94	1	0.967

4.6 Tulemustest kokkuvõtvalt

Tabelis 4.6.1 on välja toodud kõigi mudelite keskmistatud meetrikad. On näha, et teostatud mõõtmistel eelnimetatud bioloogilise materjaliga (vt tabel 3.1.3.1) saadud tulemused saame hinnata heaks kõik mudelid peale otsustuspuu (DT) mudeli. Parimaks igas meetrikas osutus tugivektormasin.

Tabel 4.6.1. Mudelite keskmistatud tulemused

Mudel	**Õigsus**	**Täpsus**	**Saagis**	**F1-skoor**
DT	0.744	0.762	0.747	0.743
RF	0.888	0.895	0.890	0.884
KNN	0.894	0.903	0.894	0.892
SVC	**0.930**	**0.939**	**0.932**	**0.931**
Ansambel hääletus	0.908	0.917	0.910	0.909

Kokkuvõte

Antud töö käigus uuriti LDI Innovation OÜ poolt arendatud seadmele H2B-Spectral mõõtmisvõimekuse piirjuhtu, mis käsitleb sarnaselt kasvatatud, sarnase kontsentratsiooniga bakteritüvede eristamise võimet mitmetel erinevatel taustpindadel. Katsetes kasutati taustmaterjalidena kuut erinevat pinnakattematerjali ja mikroobidena bioohutust arvestades kasutati viie erineva bakteri tüvesid ja isolaate. Töös otsiti ka parimat masinõppe meetodit bakteri eristamiseks, mida oleks võimalik kas otse seadmesse või arvutisse käivasse tarkvarasse integreerida.

Töös tuli välja, et kasutades erinevaid masinõppe mudeleid taustpinna mõju tunnusena, osutus oluliseks otsustuspuu ja juhusliku metsa mudelites.

Antud töös bakterite eristamisel kõige ebatäpsem kasutatud masinõppe mudelitest oli otsustuspuu. Kõigi meetrikate osas parimaks mudeliks osutus tugivektormasin.

Mõõtmistulemuste analüüsil selgus, et sama bakteri tüvede eristamisel esines mõningaid eksimusi. Erinevate bakterite (ja nende isolaatide) eristamisel eksimusi oli selgelt vähem.

Ehkki töös kasutatud erinevate bakterite ja proovide hulk oli väike, viitab antud töö seadme võimekusele eristada ka piirjuhte, kasutades masinõppe mudelit andmete töötlemisel.

Kogu töö kood on avalikult kättesaadav GitHubi repositooriumis
https://github.com/LonLoF/magistritoo

Tänuavaldused

Soovin tänada firmat LDI Innovation OÜ võimaluse eest magistritööks teostada uuringuid firmas ja kasutada firmas loodud innovaatilist seadet H2B-Spectral.

Suur tänu antimikroobsete ainete tehnoloogia professor Tanel Tensonile võimaluse eest teostada mõõtmiseid Tartu Ülikooli Tehnoloogia Instituudi antimikroobsete ainete laboris nende poolt ettekasvatatud bakterkultuuridega.

Tänan oma juhendajaid doktor Ott Rebast ja doktor Anna Aljanakit innustuse, kaasamõtlemise ja kannatlikkuse eest.

Viidatud kirjandus

[1] Piette A.S., Vybornova O., Bentahir M., Gala J.L. CBRN: Detection and identification innovations. *Crisis Response Journal*, 2014, nr 10, lk 36-38.

[2] Zhao X., Lin C.-W., Wang J., Oh D. H. Advances in Rapid Detection Methods for Foodborne Pathogens. J. Microbiol. Biotechnol., 2014, 24(3), lk 297–312.

[3] Babichenko, S. Spectral Fluorescent Signatures in Diagnostics of Water Environment. Tallinn: Tallinna Pedagoogiline Ülikool, 2001.

[4] Mitchell T. M. Machine Learning. New York: McGraw-Hill. 1997.

[5] Jordan M. I., Mitchell T. M. Machine learning: Trends, perspectives, and prospects. *Science*, 2015, vol. 349, no. 6245, pp. 255-260.

[6] Flach P. Machine Learning: The Art and Science of Algorithms That Make Sense of Data. Cambridge: Cambridge University Press. 2012.

[7] Lakowicz J.R. Principles of Fluorescence Spectroscopy. 3rd ed. New York: Springer. 2006.

[8] Castellano, F. N., Lakowicz, J. R. A Water-Soluble Luminescence Oxygen Sensor. Photochemistry and Photobiology, 1998, 67(2), pp 179–183.

[9] Chase, C. R., Van Bibber, S., Muniz, T. P. Development of a Non Contact Oil Spill Detection System. Proceedings of OCEANS 2005 MTS/IEEE, Washington, DC, USA, 17-23 September 2005.

[10] Rebane O. *In situ* non-contact sensing of microbiological contamination by fluorescence spectroscopy. Tartu Ülikooli füüsikainstituudi doktoritöö. Tartu: Tartu Ülikool, 2022.

Lisad

I. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Rimmo Rõõm,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose
   Bakterite eristamine fluoromeetri spektrist masinõppe abil
   mille juhendaja(d) on Ott Rebane ja Anna Aljanaki,
   reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Rimmo Rõõm

15.05.2024