



Sanna Uusivirta

Analyysialgoritmin kehittäminen molekyylidiagnostiselle testille

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Sähkö- ja automaatiotekniikka

Insinöörityö

03.05.2021

Tiivistelmä

Tekijä:	Sanna Uusivirta
Otsikko:	Analyysialgoritmin kehittäminen molekyyliagnostiselle testille
Sivumäärä:	22 sivua + 1 liite
Aika:	03.05.2021
Tutkinto:	Insinööri (AMK)
Tutkinto-ohjelma:	Sähkö- ja automaatiotekniikka
Ammatillinen pääaine:	Automaatiotekniikka
Ohjaajat:	Lehtori Timo Kasurinen Senior R&D Manager Juha Saharinen

Mobidiag Oy:n kehittämä Novodiag[®]-testialusta on täysin automatisoitu *In vitro* -diagnostinen (IVD) lääkinnällinen laite tartuntatauteja aiheuttavien patogeeneiden osoittamiseen erilaisista näytemateriaaleista. Testialusta koostuu itse laitteesta, kertakäyttöisestä testikasetista sekä laitteen käyttöliittymäohjelmistosta ja testikohtaisesta ohjelmistokomponentista. Testikohtainen ohjelmistokomponentti (plugin) sisältää parametrit, minkä perusteella tulos tulkitaan joko positiiviseksi tai negatiiviseksi kunkin patogeenin osalta.

Insinööriyön tarkoituksena oli määritellä analyysiparametrit Novodiag Bacterial GE+ -testille, joka tunnistaa joukon gastroenteriittiä aiheuttavia bakteereita ulostenäytteestä. Kyseisessä testissä lopullinen tunnistus tapahtuu näytteen esikäsittelyn, bakteeri-DNA:n eristämisen ja ennalta määritellyn DNA-sekvenssin monistuksen (PCR) jälkeen mikrosirulla käyttäen kohdepatogeenispesifisiä PCR-alukkeita ja koettimia sekä DNA-amplikoniin PCR-monistuksessa liitettyjä fluoresoivia leimamolekyylejä.

Mikrosirulta saatuun fluoresenssidataan vaikuttavat monet laitteesta ja sen kuvantamisominaisuuksista, testikasettien tuotannosta sekä molekyylibiologisista menetelmistä (kuten DNA-eristys ja PCR) ja biologisesta näytemateriaalista johtuvat tekijät. Tästä johtuen lopullista analyysialgoritmia varten tuotettiin dataa, jossa kohdebakteereita lisättiin näytteisiin tunnettuja pitoisuuksia. Jotta dataa voitiin käyttää algoritmin rakentamiseen, siihen linkitettiin näytteen sisällöstä kertovaa metadataa tätä tarkoitusta varten luotuun SQL-tietokantaan.

Varsinainen analyysialgoritmi kehitettiin käyttäen Mobidiagin sisäistä tarkoitukseen kehitettyä tuotekehitystyökalua, ARRY-ohjelmistoa, joka hyödyntää ohjattua koneoppimista ja jolla dataa voidaan tarvittaessa myös visualisoida. Lopuksi analyysialgoritmin säännöt käytiin manuaalisesti läpi, koska lopullisen analyysialgoritmin antamien tulosten on täytettävä testille asetetut suorituskykyvaatimukset herkkyuden, spesifisyyden sekä positiivisen ja negatiivisen ennustearvon osalta.

Insinööriyön lopputuloksena saatiin rakennettua esiversio analyysialgortimista kyseisen Novodiag-testin kaikille kohdebakteereille, joka toimii odotetusti simuloitulla ja tunnetuilla näytteillä tuotetulla datalla ja jota voidaan käyttää pohjana lopullisen, tuoteseen päätyvän analyysialgoritmin rakentamisessa.

Avainsanat: algoritmi, koneoppiminen, molekyyliagnostiikka

Abstract

Author: Sanna Uusivirta
Title: Analysis Algorithm Development for Molecular Diagnostics Assay
Number of Pages: 22 pages + 1 appendix
Date: 03 May 2021

Degree: Bachelor of Engineering
Degree Programme: Degree Program in Electrical and Automation Engineering
Professional Major: Automation technology
Instructors: Timo Kasurinen, Senior Lecturer
Juha Saharinen, Senior R&D Manager

Mobidiag Oy's Novodiag® System is a fully automated molecular diagnostics platform intended to be used for detecting pathogens causing infectious diseases from various sample materials. The system is an IVD medical device, which consists of the Novodiag instrument, disposable test cartridge, user interface software and a test specific software component called plugin. The assay specific plugin contains an analysis algorithm for each target panel which determines whether the result is interpreted as positive or negative for the pathogen target in question.

Purpose of the thesis study was to build an analysis algorithm for Novodiag Bacterial GE+ assay which identifies several gastroenteritis-causing bacteria from stool samples. The final pathogen detection is done on a microarray utilizing pathogen specific DNA probes. Before the detection, the pathogen DNA is extracted and amplified in a PCR reaction in which the target sequence is labeled with a fluorophore. This fluorophore is then detected in the microarray.

There are several sources of variation affecting the fluorescence data obtained from the microarray, such as the biological sample material, DNA extraction and PCR reactions, instrumentation, imaging properties of the Novodiag instrument and raw material variation in the cartridge production. To cover all the variation in the algorithm development process, true data was generated by running samples with spiked target pathogens. The generated data was annotated with metadata containing information of the used sample. This metadata was stored in an SQL database, in order to use it to build the algorithm.

The analysis algorithm was built using Mobidiag's internal R&D software, ARRY, which utilizes supervised machine learning and can be used to visualize the rules and datasets as well. All rules were also manually inspected after building to confirm that the assay fulfills its criteria for sensitivity, specificity, PPV and NPV. As a result of the thesis work, a primary version of the analysis algorithm was generated which worked as expected with the training dataset. This algorithm can be further utilized when developing the final algorithm for the market approved product.

Keywords: algorithm, machine learning, molecular diagnostics

Sisällys

Lyhenteet ja termit

1	Johdanto	1
2	Novodiag	2
2.1	Laitteisto	3
2.2	Testikasetti	5
2.3	Ohjelmistot	6
2.4	Novodiag Bacterial GE+	6
2.5	IVD-lääkinnällinen laite	8
3	Mikrosirutunnistus	9
3.1	Hybridisaatio	9
3.2	Fluoresenssimittaus	11
3.3	Variaatiolähteet	12
4	Koneoppiminen	14
4.1	Ohjattu oppiminen	14
4.2	Ohjaamaton oppiminen	15
4.3	Vahvistusoppiminen	15
5	Menetelmät	16
6	Yhteenveto	18
	Lähteet	20

Liite 1. Algoritmin rakentamisen yksityiskohtaiset menetelmät (SALAINEN)

Lyhenteet ja termit

ARRY	Mobidiag Oy:n sisäinen tuotekehitysohjelmisto, jota käytetään analyysialgoritmien rakentamiseen ja visualisointiin.
COU	Lyhenne sanoista <i>Cartridge Operating Unit</i> . Yksi Novodiag-laitteen neljästä toiminnallisesta yksiköstä.
DNA	Lyhenne sanoista <i>Deoxyribonucleic acid</i> eli deoksiribonukleiinihappo. Biomolekyyli, joka sisältää eliöiden geneettisen tiedon.
IVD	Lyhenne sanoista <i>In Vitro</i> Diagnostiikka. Tarkoittaa elävän yksilön ulkopuolella tehtävää diagnostiikkaa, eli käytännössä elävästä yksilöstä otetuille näytteille tehtäviä kokeita.
Metadata	Liitännäis- tai kuvailutieto, eli lisätieto jostain toisesta datasta tai datasetistä.
NPV	Lyhenne sanoista <i>Negative Predictive Value</i> eli negatiivinen ennustearvo. Kuvaa millä todennäköisyydellä negatiivisen testituloksen saaneella potilaalla todellisuudessa ei ole testattua tautia tai ominaisuutta.
PCR	Lyhenne sanoista <i>Polymerase Chain Reaction</i> eli polymeraasiketjureaktio. Molekyylibiologian menetelmä, jossa ennalta määrätystä DNA-sekvenssistä tuotetaan kopioita eksponentiaalisesti käyttäen hyväksi DNA-polymeraasientsyymejä sekä DNA-sekvenssin määrittäviä alukkeita.
Plugin	Ohjelmistokomponentti, liitännäinen, joka toimii yhdessä toisen ohjelmiston kanssa täydentäen tämän pääohjelmiston ominaisuuksia.

PPV	Lyhenne sanoista <i>Positive Predictive Value</i> eli positiivinen ennustearvo. Kuvaa millä todennäköisyydellä positiivisen testituloksen saaneella potilaalla todellisuudessa on testattu tauti tai ominaisuus.
Probe	Koetin. Yksijuosteinen DNA-sekvenssijakso, joka vastaa tiettyä osaa kyseisen koettimen kohteen geenisekvenssistä.
Probeset	Usean rinnakkaisen koettimen eli proben tuottama yhdistelmä.
RNA	Lyhenne sanoista <i>Ribonucleic acid</i> eli ribonukleiinihappo. Biomolekyyli, joka muistuttaa DNA:ta. Toimii muun muassa geneettisen tiedon välittäjänä eliöissä.
Spot	Yhden mikrosirukoettimen tuottama signaali.
SQL	Lyhenne sanoista <i>Structured Query Language</i> . Standardoitu kyselykieli, jota käytetään relaatiotietokantojen hallintaan ja tietokantahakuihin.
Target	Kohde eli testin tunnistama patogeeni.

1 Johdanto

Infektiodiagnostiikka on perinteisesti perustunut tautia aiheuttavien bakteerien viljelyyn näytteestä, patogeenien osoittamiseen antigeenitesteillä tai näytteen mikroskopointiin. Nämä menetelmät vievät aikaa, ja niillä pystytään tunnistamaan vain rajallinen määrä taudinaiheuttajia, mikä johtaa pitkittyneisiin diagnooseihin sekä tartuntatautien nopeaan leviämiseen. Viimeisen vuosikymmenen aikana on kehitetty useita molekyylibiologisia menetelmiä, joiden etuna perinteisiin menetelmiin on nopeus, tarkkuus ja laajempi taudinaiheuttajien kattavuus. [1; 2.]

Mobidiag Oy on vuonna 2000 perustettu yritys, joka kehittää ja myy kliiniseen *In Vitro* -diagnostiikkaan (IVD) tarkoitettuja laitteita ja reagensseja. Yrityksen kaksi suurinta tuoteperhettä Amplidiag® ja Novodiag® hyödyntävät molekyylibiologian menetelmiä, ja patogeenien tunnistus perustuu spesifisten DNA- tai RNA-sekvenssien eristämiseen, monistamiseen ja näiden sekvenssien tunnistamiseen. Tämä mahdollistaa taudinaiheuttajan osoittamisen näytteestä jopa alle tunnissa. [3; 4.]

Novodiag-testialusta koostuu itse Novodiag-laitteesta, kullekin patogeenipaneelille erikseen osoitetusta kertakäyttöisestä testikasetista sekä laitteen käyttöliittymäohjelmistosta ja testikohtaisesta erillisestä ohjelmistokomponentista (plugin). [4.] Testikohtainen plugin pitää sisällään muun muassa laitteen toimintaa ja testikasetissa tapahtuvaa fluidiikkaa ja kohdebakteerigeenien kohdealueiden PCR-monistamista säätelevän ohjaussekvenssin sekä lopulliseen patogeenitunnistukseen tarvittavan analyysialgoritmin.

Tämän työn tarkoituksena oli kehittää alustava versio analyysialgoritmista Novodiag Bacterial GE+ -testin kohdebakteereille niin, että tunnistus täyttää testille asetetut vaatimukset herkkyyden ja spesifisyyden suhteen. Kyseisessä testissä bakteri-DNA:n tunnistus tapahtuu mikrosirulla käyttäen spesifisiä koettimia ja fluoresenssiin perustuvaa kuvantamista [5]. Kunkin bakteerin monistettavalle

kohdegeenisekvenssille on mikrosirulla oma koettimensa ja, jokainen koetin on mikrosirulla useana kopiona, joten lopullinen sääntö kunkin bakteerin tunnistamiseen voi pitää sisällään eri yhdistelmiä useiden koettimien tuottamista signaaleista. Koska kyseessä on useita reaktiovaiheita yhdistelevä molekyylibiologiaan perustuva testi, jossa primäärinäytteen laatu ja bakteerien pitoisuudet näytteessä vaihtelevat suuresti, myös mikrosirun kuvantamisesta saatavan raakadatan laatu vaihtelee paljon. Tästä syystä algoritmin kehitystä varten tuotettiin dataa ajamalla näytteitä, jotka sisälsivät tunnettuja pitoisuuksia testin kohdebakteereja. Dataa tuottaessa pyrittiin myös käyttämään mahdollisimman montaa eri laitetta sekä useampaa kasettituotantoerää, jotta mukaan saataisiin useita mahdollisia variaatiolähteitä.

Jokaiseen näyteajoon linkitettiin näytteen sisällöstä, käytetystä laitteesta ja kasettierästä kertovaa metadataa, joka tallennettiin SQL-tietokantaan. Tämä metadata myös mahdollisti analyysialgoritmin rakentamisen käyttäen Mobidiagin sisäistä tuotekehitystyökalua, ARRY-ohjelmistoa. ARRY hyödyntää ohjattua koneoppimista ja pystyy käyttämään edellä mainittuun tietokantaan lisättyä metadataa laskiessaan parhaita mahdollisia sääntöjä kullekin kohdebakteerille käyttäen tuotettua raakadataa oppimismateriaalina. ARRY:ssa algoritmin rakennus on jaettu useaan eri tasoon, joissa seuraava rakennetaan aina edellisen päälle. Lopputuloksena ARRY:sta saadaan tunnistussäännöt kullekin testin kohdebakteerille ja -geenille, joita voidaan edelleen muokata ja visualisoida käyttäjän toimesta tarpeen mukaan. Koska kyseessä on biologinen reaktio, lopulliset säännöt tarkastetaan tarvittavan kokemuksen ja tiedon hallitsevan tutkijan toimesta.

2 Novodiag

Novodiag-testialusta on täysin automatisoitu, ”näyte-sisään-tulos-ulos”-periaatteella toimiva kokonaisuus, jossa kaikki kohdepatogeenin tunnistamiseen tarvittava tapahtuu saman, kertakäyttöisen testikasetin sisällä. Tällä hetkellä markkinoilla on viisi erilaista Novodiag-testiä eri kohdepatogeeneille, jotka kaikki hyödyntävät samaa Novodiag-laitetta ja toimintaperiaatetta. [3; 4.]

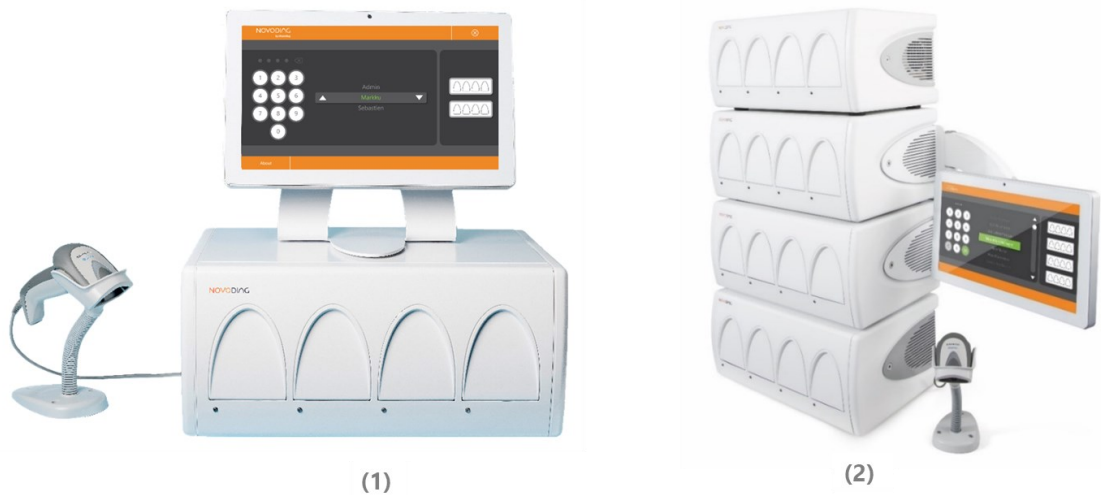
Testausprosessissa biologinen primäärinäyte (esimerkiksi uloste- tai nenä-nielunäyte) siirrostetaan ensin testille sopivaan näyteliuokseen. Tämän jälkeen näyte pipetoidaan testikasettiin, valitaan testille sopiva ohjelma ja ajo käynnistetään. Testi on valmis ja tulos nähtävissä testistä riippuen noin tunnin tai kahden kuluttua. Koko testausprosessi on esitetty kuvassa 1.



Kuva 1. Novodiag-testausprosessi. [6]

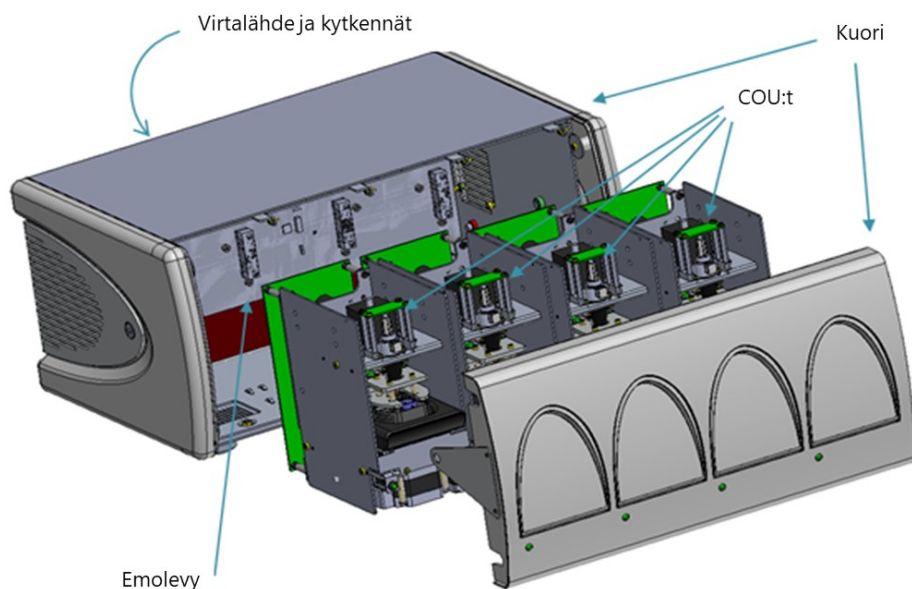
2.1 Laitteisto

Novodiag-laitteisto koostuu itse Novodiag-laitteesta, kosketusnäytöllä toimivasta tietokoneesta sekä viivakoodinlukijasta. Koko laitteisto sekä sen kaksi mahdollista asennustapaa on esitetty kuvassa 2. Yhdessä laitteessa on neljä yksikköä, joten sillä voidaan ajaa neljä testikasettia samaan aikaan. Laitteita voidaan koota korkeintaan neljä päällekkäin, jolloin samalla tietokoneella voidaan ohjata yhteensä 16 yksikköä eli ajaa 16 testikasettia samanaikaisesti. Laitetta ohjataan kosketusnäyttötietokoneelta, ja viivakoodinlukijaa käytetään valitsemaan kullekin testikasetille oikea protokolla lukemalla testikasetin viivakoodi, ja tarvittaessa myös näytteen viivakoodi. [4.]



Kuva 2. Novodiag-laitteisto ja sen kaksi mahdollista asennustapaa. Laitetta voidaan käyttää yksittäin (1) tai laitteita voidaan asentaa korkeintaan neljä päällekkäin (2). [4.]

Jokainen Novodiag-laite koostuu neljästä itsenäisesti toimivasta yksiköstä eli COU:sta (COU = Cartridge Operating Unit). Koska yksiköt toimivat itsenäisesti toisistaan riippumatta, voidaan laitteella tarvittaessa ajaa samanaikaisesti myös eri kohdepatogeeneille tai paneeleille tarkoitettuja testikasetteja. [4; 6.] Yksi COU sisältää useita moottoreita ja lämmittimiä, joita käytetään ohjaamaan testikasetissa tapahtuvaa mikrofluidiikkaa sekä biologisia reaktioita. Lisäksi jokaisessa COU:ssa on kamera mikrosirun kuvaamista varten. Räjähäytysmalli Novodiag-laitteesta on esitetty kuvassa 3.

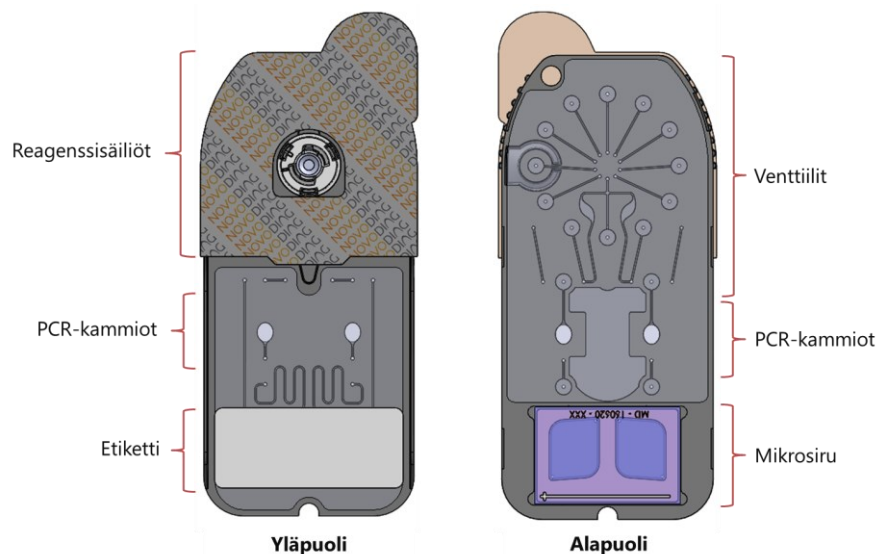


Kuva 3. Novodiag-laitteen rakenne.[6]

Jokaisella COU:lla on oma tytärlevynsä (daughter board), joka ohjaa kunkin COU:n toimintaa ja kiinnittyy Novodiag-laitteen emolevyyn (mother board). Tämä mahdollistaa jokaisen yksikön itsenäisen toiminnan, vaikka yhteys laitteiston tietokoneeseen katkeaisi ajon ollessa käynnissä. Mikäli laite vikaantuu, voidaan siihen vaihtaa yksittäinen COU verrattain helposti esimerkiksi paikan päällä asiakkaan luona ilman, että koko laitetta tarvitsee huoltaa. [6.]

2.2 Testikasetti

Novodiag-testikasetit sisältävät kaiken tarvittavan kohdepatogeeni-DNA:n tai RNA:n eristämiseen, monistamiseen PCR-tekniikalla sekä monistetun patogeeni-DNA:n visualisointiin. Testikasetit ovat kertakäyttöisiä, joten jokaiselle näytteelle täytyy olla oma testikasettinsa, joka tulee valita primäärinäytteen ja halutun testipaneelin mukaan. Kuvassa 4 on esitetty Novodiag-testikasetin rakenne ja sisältö. [4; 5; 6.]



Kuva 4. Novodiag-testikasetin rakenne ja osat. [5]

Kasetin yläosassa sijaitsevat reagenssisäiliöt ja niiden pohjalla venttiilien läpimenoaukot sekä nestekanavat, joita pitkin nesteitä liikutellaan kasetin yläosan keskellä olevaan ruiskusäiliöön. Kasetin keskiosassa sijaitsevat PCR-kammiot,

joissa tapahtuu kohdepatogeenien nukleiinihappojen (DNA tai RNA) monistaminen hybridisaatiota ja mikrosirutunnistusta varten. Nesteet siirretään kanavia pitkin PCR-kammioihin käyttäen kasetin ruiskusäiliötä ja venttiileitä, kun näyte on ensin puhdistettu ja näytteen nukleiinihapot eristetty muusta biologisesta materiaalista kasetin yläosan reagenssisäiliöitä käyttäen. Kasetin alaosassa sijaitsevat kasetin etiketti, jolle on printattu tuotekohtainen viivakoodi, mikä sisältää tiedot muun muassa kasettierästä, viimeisestä käyttöpäivästä sekä tarvittavasta ohjausekvenssistä ja analyysiparametreista. Kasetin alapuolella alaosassa sijaitsevat hybridisaatiokammiot ja niiden päällä mikrosiru. Kaikki reaktiot ja fluidiikka tapahtuvat kertakäyttöisen kasetin sisällä suljetussa tilassa. Tämä ehkäisee mahdollisia kontaminaatioita tehokkaasti ja parantaa laitteen käytettävyyttä, kun eri testikasetteja voidaan ajaa peräjälkeen samoilla laitteilla ilman välipuhdistusta tai muuta vastaavaa toimenpidettä. [5; 6.]

2.3 Ohjelmistot

Novodiag-laitteistoa ohjataan systeemiin kuuluvalla kosketusnäyttötietokoneella, johon on asennettu laitteen käyttöliittymäohjelmisto. Käyttöliittymäohjelmiston kautta hallitaan muun muassa laitteiden konfiguraatiota, käyttäjiä, ajojen käynnistämistä ja päättämistä sekä ajojen tuloksia, joita voidaan tarvittaessa myös tulostaa tai viedä eteenpäin muihin järjestelmiin. Lisäksi jokainen testi tarvitsee toimiakseen testikohtaisen ohjelmistokomponentin (plugin), joka sisältää kullekin testille spesifisen ohjausekvenssin ja analyysiparametrit. Tämä plugin asennetaan laitteelle käyttöliittymäohjelmiston kautta ja oikea plugin kullekin testille valitaan lukemalla viivakoodi testikasetista ennen ajon aloittamista. [4.]

2.4 Novodiag Bacterial GE+

Novodiag Bacterial GE+ on yksi viidestä markkinoilla olevista Novodiag-tuotteesta, joka on tarkoitettu käytettäväksi bakteerien aiheuttaman gastroenteriitin eli mahasuolitulehduksen diagnostiikkaan ulostenäytteestä. Se tarjoaa vaihtoehdon aikaa ja paljon laboratoriokapasiteettiä vieville bakteeriviljelyille, kun koko prosessi on automatisoitu ja tulokset ovat saatavilla alle kahdessa tunnissa. [5.]

Ripuli on yksi yleisimmistä sairauksista maailmanlaajuisesti, ja se on toiseksi yleisin kuolinsyy lapsilla kehittyvissä maissa [7; 8]. Oikean hoidon saamiseksi, taudin pitkittymisen ehkäisemiseksi ja turhien antibioottien määräämisen välttämiseksi on tärkeää saada selville ripulin ja siihen liittyvän ruuansulatuskanavan tulehduksen aiheuttaja mahdollisimman nopeasti ja kustannustehokkaasti. [5; 7.]

Novodiag Bacterial GE+ tunnistaa 13 yleisintä gastroenteriittiä aiheuttavaa bakteeria. Nämä bakteerit sekä näiden bakteerien tunnistamiseen käytetyt geenit on esitetty taulukossa 1.

Taulukko 1. Novodiag Bacterial GE+ testin targetit. [5]

Bakteeri	Kohdegeeni(t)
<i>Campylobacter coli</i>	gyrB
<i>Campylobacter jejuni</i>	rimM
<i>Clostridium difficile</i> toksiini B	tcdB
EAEC	aggR
EHEC	eae, stx1, stx2
EPEC	eae
ETEC	elt, est
<i>Salmonella</i> spp.	invA
<i>Shigella</i> spp. / EIEC	ipaH, invE
<i>Vibrio cholerae</i>	ctx
<i>Vibrio parahaemolyticus</i>	toxR
<i>Yersinia enterocolitica</i>	virF
<i>Yersinia pseudotuberculosis/pestis</i>	rumB

Testikasetissa kohdebakteerien DNA-monistamiseen ja hybridisaation tarvittavat koettimet ja alukkeet on jaettu kahteen eri kammioon ja jokaiselle kohdebakteerille on mikrosirulla useita eri koettimia, joiden perusteella lopullinen tunnistus tapahtuu.

2.5 IVD-lääkinnällinen laite

In vitro -diagnostiikka (IVD) tarkoittaa potilaasta otetusta näytteestä tehtävää lääketieteellistä tutkimusta tai koetta sairauden määrittelemiseksi tai muun tiedon saamiseksi. Terminä *in vitro* tarkoittaa elävän organismin ulkopuolella tehtävää koetta. [9.] Lääkinnällisiä laitteita koskevan EU-direktiivin (98/79/EY) mukaan *In vitro* -diagnostiikkaan tarkoitettu lääkinnällinen laite on laite, joka on

reagenssi, reagenssituote, kalibraattori, vertailumateriaali, testipakkaus, instrumentti, laite, laitteisto tai järjestelmä ... jonka valmistaja on tarkoittanut käytettäväksi *in vitro* ihmiskehosta otettujen näytteiden, mukaan lukien veren ja kudosten luovutukset, tutkimuksissa yksinomaisena tai pääasiallisena tarkoituksena saada tietoa:

- fysiologisesta tai patologisesta tilasta tai
- synnynnäisestä epämuodostumasta tai
- turvallisuuden ja yhteensopivuuden määrittämiseksi mahdollisten saajien kannalta tai
- hoitotoimenpiteiden tarkkailemiseksi. [10.]

In vitro -diagnostiikkaan käytettävien lääkinällisten laitteiden tuotekehitystä, testaamista, markkinoille saattamista ja valmistusta säädellään Suomessa tarkasti EU-tasolla ja direktiiveistä johdetulla kansallisella lainsäädännöllä, jotta voidaan varmistaa laitteen käyttäjäturvallisuus sekä laitteen tuottaman tiedon oikeellisuus. Lisäksi myös lääkinällisillä laitteilla tulee olla CE-merkintä ennen kuin tuotetta saa myydä. Laitteeseen liitettävällä CE-merkinnällä valmistaja vakuuttaa, että se täyttää kyseiselle tuotteelle asetetut vaatimukset EU-tasolla [11]. Mikäli tuotetta halutaan myydä EU:n ulkopuolella, tulee täyttää vielä kohdemarkkina-alueen kansalliset vaatimukset. Tällä hetkellä on meneillään siirtymäkausi, jossa vanha EU-direktiivi (98/79/EY) korvataan vuonna 2017 voimaan tulleella EU-asetuksella (2017/746/EU) tarkoituksena yhtenäistää EU-alueen sääntelyä IVD-lääkinällisten laitteiden osalta ja poistaa tarve kansalliselle lainsäädännölle. [10;12.]

IVD-lääkinnällisiä laitteita koskeva lainsäädäntö, asetukset ja direktiivit velvoittavat siis valmistajan vakuuttamaan, että markkinoille tuotu tuote on käynyt läpi muodollisen verifikaatio- ja validaatioprosessin, ja myös julkistamaan näiden prosessien perusteella saadut tuotteen suorituskykyä ja turvallisuutta kuvaavat arvot ja tulokset. Käytännössä tämä tarkoittaa tuotteen valmistajan kannalta merkittävää työmäärää, koska tuotteen testaaminen erilaisissa olosuhteissa on valmistajan vastuulla. Usein tämän sääntelyn piirissä olevien tuotteiden kehitys onkin jo varsin aikaisesta vaiheesta asti tiukasti säänneltyä, ja verifikaatio- ja validaatioprosessit vievät merkittävän osan tuotekehitysprojektin resursseista.

3 Mikrosirutunnistus

Novodiag Bacterial GE+ -testissä kohdepatogeenien tunnistus perustuu DNA-mikrosirulta luettaviin fluoresenssisignaaleihin. Ennen mikrosirutunnistusta haluttu kohde-DNA tulee monistaa ja hybridisoida mikrosirun pinnalla oleviin koetimiin signaalin tuottamiseksi. Lopuksi mikrosiru kuvataan, ja lopullinen bakteerin tai geenin tunnistus tehdään fluoresenssisignaalien ja niiden mahdollisesti tuottamien yhdistelmien perusteella, kun tiedetään, mihin kohtaan mikrosirua minkäkin bakteeri-DNA-jakson eli DNA-sekvenssin emäsjärjestystä vastaava koetin on liitetty.

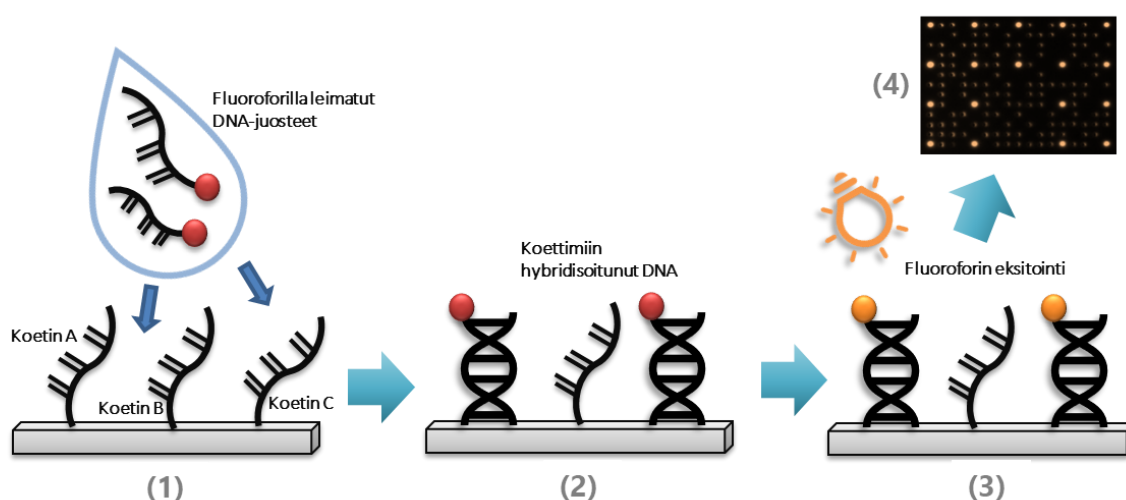
3.1 Hybridisaatio

DNA-hybridisaatio on menetelmä, jota on hyödynnetty molekyylibiologiassa jo useita vuosikymmeniä. Se perustuu nukleiinihappojen, eli DNA:n ja RNA:n, emäsjärjestykseen ja pariutumissääntöihin sekä siihen, miten tämä nukleiinihappojärjestys eroaa eri eliöiden ja jopa yksilöiden välillä ja kuinka eripituiset nukleiinihappojuosteet pariutuvat eri lämpötiloissa. [13.] Hybridisaatioon perustuvia sovelluksia on monia, mutta tässä esitellään perusperiaate, jota myös Novodiag-laitteisto ja teknologia hyödyntävät

Varsinaista hybridisaatioreaktiota varten tutkittava kohde-DNA monistetaan ensin PCR-reaktiossa ja siihen lisätään fluoresenssitunnistuksen mahdollistavia

fluorofori-molekyylejä. Mikäli kyseistä kohde-DNA:ta ei näytteestä löydy, ei tätä monistumista myöskään tapahdu ja tulos hybridisaatioreaktiosta on tälle kohteelle negatiivinen eli tunnistettavaa fluoresenssia ei nähdä.

Hybridisaatiossa käytettävän mikrosirun pinnalle on kiinnitetty kohde-DNA:lle spesifisiä nukleinihappokoettimia, jotka on suunniteltu ja syntetisoitu niin, että niihin sitoutuvat vain kohde-DNA:lla komplementaariset DNA-juosteet. Kaavakuva hybridisaatioreaktiosta on esitetty kuvassa 5.



Kuva 5. DNA-hybridisaation periaate.

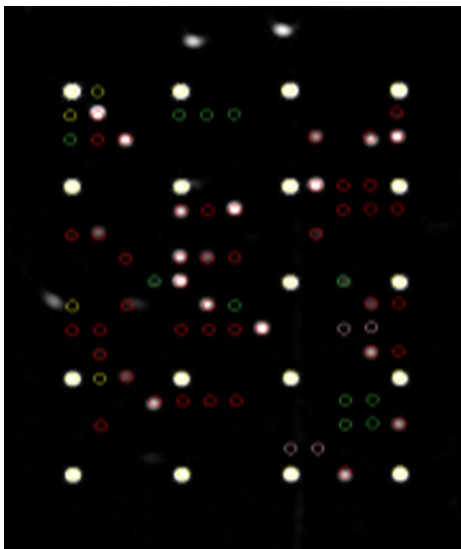
PCR-monistuksen jälkeen nestemäinen reaktioseos viedään mikrosirun pinnalle, siihen lisätään reaktio-olosuhteet mahdollistavaa puskuriliuosta ja lämpötila nostetaan reaktiolle optimaaliseksi. Tämä lämpötila riippuu muun muassa kohde-DNA:n ja nukleinihappokoettimen emäsjärjestyksestä ja pituudesta [13]. Hybridisaatioreaktiossa fluorofori-molekyylin sisältävät DNA-juosteet sitoutuvat niille sopiviin nukleinihappokoettimiin mikrosirun pinnalla. Reaktion lopuksi ylimääräinen, ei-hybridisoitunut DNA pestään sirun pinnalta pois, mikrosirulle johdetaan valoa tietyllä aallonpituudella, jolla fluorofori-molekyylit viritetään halutulle taajuudelle (eli molekyylin atomi absorboi fotonin). Tämän virityksen purkauksessa molekyylit vapauttavat eli emittoivat fotonin, mikä detektoidaan fluoresenssina tietyllä, fluorofori-molekyylin ominaisuuksista riippuvalla aallonpituudella. [14.]

3.2 Fluoresenssimittaus

Fluoresenssia käytetään tunnistuksessa useissa molekyylibiologissa menetelmissä. Myös Novodiag-laitteisto hyödyntää fluoresenssitunnistusta DNA-hybridisaation osoituksessa.

Mittaukseen varten tarvitaan fluoresoiva molekyyli eli fluorofori, jonka ominaisuuksiin kuuluvat tietty eksitaatio- ja emissioaallonpituus. Näiden perusteella laitteelta vaaditaan valonlähde, joka sopii käytetyn fluoroforin eksitaatioaallonpituudelle. Lisäksi laitteelta vaaditaan detektori ja siihen liitetty suodatin, jotka pystyvät erottelemaan käytetyn fluoroforin emissiosignaalin muusta valosta. [14.] Novodiagin käyttämässä menetelmässä fluorofori liitetään osaksi kohde-DNA-juostetta PCR-reaktiossa.

Novodiagissa mikrosirusta otetaan hybridisaation jälkeen tarkoitukseen kehitetyllä detektorikameralla useita kuvia eri valotusajalla, joista saadaan lopullinen kuva, jonka perusteella varsinainen analyysi tehdään. Useiden eri valotusajalla otettujen kuvien käyttäminen mahdollistaa sen, että analyysiin saadaan mukaan sekä heikon että vahvan signaalin antavat koettimet. Kuvassa 6 on esimerkki Novodiag-analyysiin käytettävästä kuvasta. Kuvassa kirkkaimmat, tasaisin välein näkyvät 18 spottia ovat grid- eli kohdistuskontrolleja, joiden avulla kuva kohdistetaan ja jokaisen spotin oletettu sijainti määritellään.



Kuva 6. Esimerkki Novodiagin hybridisaatiokuvasta.

3.3 Variaatiolähteet

Novodiag-testistä lopulta saatavaan fluoresenssi-raakadataan vaikuttavat useat eri lähteistä tulevat tekijät. Kyseessä on useita monimutkaisia molekyylibiologisia reaktioita ketjuttava prosessi, jossa jokainen vaihe ja olosuhde vaikuttavat lopputulokseen.

3.3.1 Primäärinäyte

Primääri- eli suoraan potilaasta tullut, käsittelemätön näyte on aina laadultaan ja biologisilta ominaisuuksiltaan ainutkertainen. Testiprotokollassa primäärinäyte siirrostetaan ja sekoitetaan näyteliuokseen, ja lopullinen testikasettiin päätyvä osuus on noin 10 % itse primäärinäytettä näyteliuoksessa [5]. Näytteen ominaisuuksista johtuvaa variaatiota pyritään hallitsemaan ja ennakoimaan testin kaikissa kehitysvaiheissa käyttämällä mahdollisimman paljon erilaisia primäärinäytetaustoja testejä tehdessä.

3.3.2 DNA:n eristäminen

Novodiag-testikasetissa näytteen kaikki nukleiinihapot eristetään muusta näytemateriaalista ja puhdistetut nukleiinihapot viedään edelleen monistettavaksi ja siitä edelleen hybridisaatioon. Se, kuinka paljon nukleiinihappoja mistäkin näytteestä saadaan eristettyä, riippuu suuresti näytteen laadusta ja koostumuksesta. Lisäksi seuraavien vaiheiden onnistumiseen vaikuttaa paljon myös se, kuinka paljon muita nukleiinihappoja näytteestä eristyy halutun kohdebakteeri-DNA:n lisäksi.

3.3.3 DNA-monistus ja hybridisaatio

DNA-monistaminen eli PCR sekä DNA-hybridisaatio ovat molekyylibiologisia menetelmiä, jossa reaktio perustuu tiettyjen, ennalta määriteltyjen DNA-jaksojen ominaisuuksiin. Jokaisen DNA-jakson monistumiselle ja hybridisaatiolle voidaan laskea ja testata optimaaliset olosuhteet. Suurimmassa osassa nykyään käytössä olevissa molekyyli diagnostiikan testeissä halutaan kuitenkin käsitellä useita eri DNA-jaksoja samassa reaktiossa, joten valitut olosuhteet ovat aina kompromissi. Näin ollen toiset DNA-jaksot monistuvat PCR-reaktiossa tehokkaammin ja sitoutuvat koettimiinsa paremmin hybridisaatiossa ja tuottavat lopulta vahvemman tunnistussignaalin.

3.3.4 Laitteisto

Novodiagin fluorenssimittauksessa lopulliseen signaaliin vaikuttavat myös testikasetin sisällä tapahtuvien reaktioiden ulkopuoliset tekijät. Näitä ovat muun muassa kasetin tuotantoon käytettävien komponenttien mahdollinen laatu vaihtelu sekä yksilölliset erot eri laitteiden välillä, jotka liittyvät esimerkiksi fluoresenss-detektoriiin tai fluoroforin eksitointiin tarvittavaan valonlähteeseen. Näitä variaatiolähteitä pyritään kontrolloimaan useissa eri laite- ja kasettituotannon vaiheissa tekemällä laadunvalvontatestejä jo puolivalmisteille ja sisääntulotarkastuksia tuotannossa käytettäville materiaaleille.

4 Koneoppiminen

Koneoppiminen on tekoälyn osa-alue, jossa pyritään rakentamaan algoritmeja, malleja ja ohjelmistoja, millä ennustetaan tulevia tapahtumia, kuvaillaan nykytilannetta tai pyritään ratkaisemaan jokin ongelma perustuen olemassa olevaan dataan. Koneoppimisessa hyödynnetään matematiikkaa, tilastollisia menetelmiä sekä tietojenkäsittelytiedettä, jotta sovellus voi oppia ja kehittyä ilman, että sitä on erikseen ohjelmoitu tekemään kyseinen tehtävä, ja sitä hyödynnetään erityisesti, kun kyseessä on monimutkainen ongelma tai suuri määrä dataa. Perinteisesti ohjelmoitaessa pyritään tekemään ohjelma, joka esimerkiksi ratkaisee jonkin yhtälön ja antaa tuloksen, kun sille syötetään jotkin tietyt arvot, tai vaihtoehtoisesti tehdään ohjelma, joka jaottelee sille syötetyn datan tietyllä tavalla. Koneoppimisessa sen sijaan pyritään siihen, että kone oppii itse esimerkiksi rakentamaan yhtälön tai jaottelemaan datan, kun sille annetaan käytettävät arvot ja haluttu lopputulos tai suuri määrä dataa jaoteltavaksi. [15, s. 2–4, 16, s. 2.]

Koneoppiminen voidaan jakaa niin sanottuun ohjattuun oppimiseen, ohjaamattomaan oppimiseen ja vahvistusoppimiseen, jotka eroavat toisistaan käytettyjen menetelmien, sekä opetusdatan luonteen osalta. [15, s. 19–20.]

4.1 Ohjattu oppiminen

Ohjatussa oppimisessa koneelle annetaan oppimista varten datasetti (*input*), joka sisältää myös sille sopivan vasteen (*output*) ja sitä pyydetään tämän datasetin perusteella rakentamaan ohjelma, joka tuottaa tietyn vasteen tai tuloksen myös silloin, kun sille syötetään vastaava data, jonka vastetta ei vielä tiedetä. Esimerkki tällaisesta opetusdatasta voi olla esimerkiksi joukko kuvia erilaisista maatilan eläimistä sekä niihin yhdistetty tieto siitä, mikä eläin kuvassa on. Koneoppimisprosessin tuloksena syntyy tällöin ohjelma, joka tunnistaa, mikä maatilan eläin on kuvassa, kun se on ensin opetellut kunkin eläimen ominaispiirteet sille annetusta opetusdatasta. [15, s. 20.]

Ohjatun koneoppimisen menetelmät voidaan jakaa edelleen regressio- ja luokittelualgoritmeihin. Edellinen maatalan eläimet -esimerkki oli juuri luokittelualgoritmi, missä kuvat jaettiin eri kategorioihin niiden ominaispiirteiden avulla. Regressioalgoritmeilla pyritään sen sijaan ennustamaan jonkin muuttujan arvoa. Tällainen voisi olla esimerkiksi osakkeen hinnan ennustaminen tietyin aikavälein perustuen pörssin historiatietoon. [15, s. 25.] Muutamia käytetyimpiä ohjatun oppimisen algoritmialleja ovat muun muassa lineaarinen tai logistinen regressio, K :n lähimmän naapurin luokitin (*K-nearest neighbors*, *KNN*), päätöspuut (*decision trees*) ja satunnaismetsät (*random forests*). [16.]

4.2 Ohjaamaton oppiminen

Ohjaamaton koneoppiminen eroaa ohjatusta oppimisesta siinä, että opetusdata-setti (*input*) ei sisällä tietoa vasteesta (*output*) ja sen tarkoituksena on löytää tästä datasta kaavoja, malleja tai korrelaatioita, minkä perusteella datasta voidaan tehdä päätelmiä tai ennustaa tulevaa. [16, s. 11.] Edellisen kappaleen kuvaesimerkkiin viitaten ohjaamattomassa oppimisessä koneelle annettaisiin suuri määrä kuvia maatalan eläimistä, mutta sille ei kerrotaisi, mikä eläin on kuvassa. Näin kone voi esimerkiksi luokitella kuvat eri ryhmiin sen perusteella, mitkä muistuttavat toisiaan ja oppia yhdistämään kunkin eläimen oikeaan ryhmään. Tätä menetelmää kutsutaan klusteroinniksi. Muutamia käytetyimpiä klusterointimenetelmiä ovat muun muassa K :n keskiarvon klusterointi sekä hierarkkinen klusterointi. [15, s.21, 16, s. 162.]

4.3 Vahvistusoppiminen

Vahvistusoppimisesta koneoppimisen osa-alueena puhutaan, kun kone oppii toimimaan tietyllä tavalla suhteessa ympäristöön muuttuvassa tilanteessa. Se on yksi eniten tutkituista koneoppimisen osa-alueista tänä päivänä [16, s. 13]. Käytännön esimerkkinä vahvistusoppimisesta voidaan esittää esimerkiksi shakkipeliä, jossa koneelle voidaan antaa tiedot pelin säännöistä ja sallituista siirroista, mutta sen tulee muuttaa toimintaansa vastustajan siirtojen mukaan toivottu lopputuloksen saamiseksi. [15, s. 23.]

5 Menetelmät

Työ analyysialgoritmin rakentamiseksi tehtiin kolmessa osassa. Ensin tuotettiin tarvittava määrä dataa sekä positiivisilla että negatiivisilla kliiniseen taustaan tehdyillä näytteillä käyttäen mahdollisimman monipuolisesti saatavilla olevia laitteita ja kasettieriä. Jotta dataa voitiin käyttää algoritmin rakentamiseen, siihen linkitettiin näytteen sisällöstä kertovaa metadataa tätä tarkoitusta varten luotuun SQL-tietokantaan. Seuraavaksi jokaiselle testin targetille rakennettiin tunnustus-säännöt ARRY-ohjelmistolla (sisäinen tuotekehitysohjelmisto, Mobidiag Oy) käyttäen ensimmäisessä vaiheessa luotua dataa. Viimeisessä vaiheessa säännöt tarkistettiin.

Yksityiskohtaiset menetelmät algoritmin rakentamiseksi on esitetty liitteessä 1, joka on julistettu salaiseksi Mobidiag Oy:n pyynnöstä liikesalaisuuteen vedoten.

Lopullinen analyysialgoritmi piti sisällään säännöt kaikille Novodiag Bacterial GE+ -testin kohdebakteereille. Lopulliset säännöt on esitetty liitteessä 1. ARRY:n koneoppimistyökalulla määritetyt sääntöjä käytettiin fluoresenssin mediaani-intensiteettiarvojen määrittämisessä. Koneoppimista kokeiltiin hyödyntää myös muiden parametrien määrittelyssä, mutta näiden kohdalla manuaalinen sääntöjen kirjoittaminen tuotti paremman lopputuloksen.

Opetusdatasta saadut testin suorituskykyä kuvaavat arvot kullekin targetille on esitetty taulukossa 2. Testille asetetut vaatimukset suorituskyvyn suhteen ovat seuraavat: herkkyuden tulee olla suurempi kuin 80 % ja tarkkuuden suurempi kuin 95 %. Käytännössä tämä tarkoittaa, että tarkkuutta painotetaan herkkyyttä enemmän. Taulukossa esitetyt arvot PPV ja NPV ovat kliinisessä diagnostiikassa käytettyjä parametreja, jotka kuvaavat positiivisen ja negatiivisen tuloksen ennustearvoa eli sitä, millä todennäköisyydellä positiivisen tai negatiivisen testituloksen saanut potilas todella on tai ei ole kyseisen bakteerin kantaja. Näille arvoille ei ole asetettu erikseen kriteeriä, mutta pyrkimys on saada nekin mahdollisimman lähelle 100 % kun ne lasketaan käytetyn opetusdatan perusteella.

Taulukko 2. Analyysialgoritmillä saadut suorituskykyarvot Novodiag Bacterial GE+ -testin targeteille.

Bakteeri	Kohdegeeni	Herkkyys %	Tarkkuus %	PPV %	NPV%
<i>C. coli</i>	gyrB	83,0 %	100 %	100 %	98,6 %
<i>C. jejuni</i>	rimM	100 %	100 %	100 %	100 %
<i>C. difficile</i>	tcdB	81,6 %	99,8 %	97,6 %	98,6 %
EAEC	aggR	98,1 %	100 %	100 %	99,8 %
EHEC	eae	89,5 %	100 %	100 %	98,3 %
	stx1	92,0 %	100 %	100 %	99,4 %
	stx2	96,0 %	100 %	100 %	99,7 %
EPEC	eae	89,5 %	100 %	100 %	98,3 %
ETEC	elt, est	98,5 %	100 %	100 %	99,8 %
<i>Salmonella</i> spp.	invA	100 %	100 %	100 %	100 %
<i>Shigella</i> spp. / EIEC	ipaH	80,4 %	100 %	100 %	97,0 %
	invE	89,8 %	100 %	100 %	99,2 %
<i>V. cholerae</i>	ctx	100 %	100 %	100 %	100 %
<i>V. parahaemolyticus</i>	toxR	94,0 %	100 %	100 %	99,5 %
<i>Y. enterocolitica</i>	virF	68,6 %	100 %	100 %	97,5 %
<i>Y. pseudotuberculosis</i> / <i>pestis</i>	rumB	100 %	100 %	100 %	100 %

Taulukosta 2 nähdään, että lähes kaikki testin targetit täyttivät testille asetetut kriteerit herkkyiden ja tarkkuuden suhteen. Erityisesti tarkkuus on kaikilla targeteilla erinomainen. Ainoastaan yhdellä targetilla, *Y. enterocolitica*, herkkyys jäi alle 80 %:n. Tämän targetin osalta tehtiin tarkempi selvitys, jossa tutkittiin samoista näytteistä vaihtoehtoisella menetelmällä saadut tulokset. Tällä mene-

telmä antoi herkkyudeksi 80 %, joten selitys ei löydy pelkästään käytetystä näytteestä. Myös raakadataa tutkittiin tarkemmin, ja huomattiin että osassa vääriksi negatiivisiksi jääneistä ajoista ei todellakaan nähty detektoitavaa signaalia, mikä eroaisi selvästi negatiivisilla näytteillä ajetuista ajoista. Tämän yhden targetin tutkimiseksi tarvitaan vielä lisää dataa ja mahdollisesti uusi opetusdatasetti, jotta säännöstä saadaan rakennettua parempi.

Vaikka kaikki muut targetit täyttivät testille asetetut kriteerit, osalla herkkyys näyttäisi olevan lähellä testille asetettua 80 %:n rajaa. Nämä targetit olivat *C. coli*, *C. difficile* ja *Shigella* (ipaH-geeni). Myös näiden targetien osalta tehtiin tarkempi tutkimus, jossa todettiin, että myös verrokkimenetelmällä samoista näytteistä saadut herkkyysarvot olivat noin 80 % tai jopa sen alle. Näin ollen voitiin todeta, että hieman alhaisempi, vaikkakin kriteerit täyttävä, herkkyys johtui näiden targetien kohdalla käytetystä näytteestä.

6 Yhteenveto

Tämän työn tarkoituksena oli rakentaa Novodiag Bacterial GE+ -testille analyysialgoritmi, jolla analysoidut tulokset täyttävät testille asetetut suorituskykykriteerit ja joka voidaan ottaa käyttöön sellaisenaan kyseisen testin jatkokehitykseen. Analyysialgoritmin kehityksessä oli tarkoitus käyttää hyväksi koneoppimista mahdollisuuksien mukaan, ja tätä varten luotiin merkittävä määrä dataa ajamalla näytteitä, joihin oli lisätty tunnettuja määriä testin kohdebakteeritargetteja. Tähän dataan liitettiin SQL-tietokantaan tallennettua metadataa, joka kuvasi käytettyjen näytteiden sisältöä, jolloin sitä voitiin käyttää myös opetusdatana Mobidiag Oy:n sisäisen tuotekehitysohjelmisto ARRY:n koneoppimisominaisuudelle.

Algoritmia rakentaessa oli selvää jo hyvin aikaisessa vaiheessa, että pelkästään koneoppimista ei voitu käyttää algoritmin sääntöjen rakentamiseen vaan lisäksi tarvittiin asiantuntija-arviota. Tämä johtui muun muassa siitä, että osa käytetyistä parametreista haluttiin mukaan syistä, jotka eivät käyneet ilmi käytetystä

opetusdatasta. Näitä olivat erityisesti Novodiag-kasetin mikrosirutuotantoon liittyvät laadunvalvontakriteerit ja tuotantolaitteiston asettamat rajoitukset sekä puhtaasti biologiaan, eli kohdepatogeenien geneettisiin ominaisuuksiin liittyvät syyt. Kasettituotantoon liittyvät parametrit eivät ota sellaisenaan kantaa siihen, onko lopputulos jollekin patogeenille positiivinen vai negatiivinen, joten näitä parametreja ei olisi voitu tällä lähestymistavalla määrittää joka tapauksessa. Käytännössä ei ollut myöskään mahdollista saada opetusdataan mukaan kaikilla mahdollisilla eri bakteerivarianteilla ajettuja näyteajoja.

ARRY:n koneoppimistyökalu osoittautui kuitenkin hyödylliseksi algoritmin rakentamisessa, kun sille oli ensin määritelty selkeät rajat ja kohde. Tässä työssä käytetyn testin osalta olisi todennäköisesti selvitty myös jollain yksinkertaisemminkin lähestymistavalla, kun lopulta koneoppimistyökalulle jäi määritettäväksi vain muutama fluoresenssin mediaani-intensiteetin raja-arvo kullekin patogeenille. Koneoppimisominaisuudesta oltaisiin mahdollisesti saatu myös enemmän hyötyä irti, mikäli opetusdataa olisi tuotettu merkittävästi enemmän ja käyttäen myös todellisia kliinisiä näytteitä. ARRY:n koneoppimistyökalu antoi kuitenkin merkittävää tietoa siitä, millä parametreilla oli merkitystä sille, onko lopputulos negatiivinen vai positiivinen käytetyn opetusdatan perusteella. Työstä saatiin myös paljon arvokasta tietoa ja käyttökokemusta ARRY-ohjelmiston toiminnasta yleensä, mikä palvelee tulevaisuudessa muiden, mahdollisesti monimutkaisempien analyysialgoritmien kehityksessä.

Työn tuloksena saatua lopullista analyysialgoritmia ei tulla implementoimaan lopulliseen, markkinoilla olevaan tuotteeseen sellaisenaan, mutta sitä voidaan käyttää tulosten analysointiin tuotteen jatkokehityksessä. Algoritmia täytyy testata vielä todellisilla kliinisillä näytteillä, ja yhden targetin (*Y. enterocolitica*) osalta huonomman herkkyuden syy täytyy selvittää. IVD-lääkinnällisiin laitteisiin liittyvät direktiivit, standardit ja lainsäädäntö, sekä niihin perustuva Mobidiag Oy:n laatujärjestelmä myös vaativat, että kaikki markkinoille päätyvät tuotteet käyvät läpi muodollisen verifikaation ja validaation, mikä koskee kaikkia Novodiag-systeemin osia, myös ohjelmistoja ja tätä analyysialgoritmia.

Lähteet

- 1 L.E.S. Bruijnesteijn van Coppenraet, M. Dullaert-de Boer, G.J.H.M. Ruijs, W.A. van der Reijden, A.G.M. van der Zanden, J.F. L. Weel, T.A. Schuurs, *Case-control comparison of bacterial and protozoan microorganisms associated with gastroenteritis: application of molecular detection*. *Clinical Microbiology and Infection*, 2015; 21: 592.e9-592.e19. <https://www.sciencedirect.com/science/article/pii/S1198743X15002955> .
- 2 H. Zhang, S. Morrison, Y. Tang. *Multiplex Polymerase Chain Reaction Tests for Detection of Pathogens Associated with Gastroenteritis*. *Clinics in Laboratory Medicine*, 2015; 35: 461-486. <https://www.sciencedirect.com/science/article/abs/pii/S0272271215000256> .
- 3 Mobidiag Oy. *Novodiag-system*. Verkkoaineisto. <https://mobidiag.com/products/novodiag/> Luettu 31.03.2021.
- 4 Mobidiag Oy. *Novodiag® System Instructions for Use*, Version 5-0. 2021. Tuotenumero: NVD-SYST-IFU.
- 5 Mobidiag Oy. *Novodiag® Bacterial GE+ Instructions for Use*, Version 6-0. 2021. Tuotenumero: NVD-BGE-IFU.
- 6 Mobidiag Oy:n sisäinen koulutusmateriaali, ei julkaistu. Luettu 31.03.2021.
- 7 WHO. *Diarrhoeal disease fact sheet*. Verkkoaineisto, 2017. <https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease> Luettu 31.03.2021.
- 8 P. Kelly. *Infectious diarrhoea*. *Medicine*, 2011; 39: 201-206.
- 9 Euroopan kemikaalivirasto. *In vitro -menetelmät*. Verkkoaineisto. <https://echa.europa.eu/fi/support/registration/how-to-avoid-unnecessary-testing-on-animals/in-vitro-methods> Luettu 31.03.2021.
- 10 Euroopan Komissio. *Tiedote in vitro -diagnostiikkaan tarkoitettujen lääkinällisten laitteiden valmistajille*, Euroopan Komission asiakirja 20.11.2018, ISBN: 978-92-79-96574-6.
- 11 Tukes. *CE-merkintä*. Verkkoaineisto. <https://tukes.fi/tuotteet-ja-palvelut/ce-merkinta#5f0eb58b> Luettu 31.03.2021.
- 12 Euroopan komissio. *Medical Devices sector – lääkinälliset laitteet*. Verkkoaineisto. https://ec.europa.eu/health/md_sector/overview_fi Luettu 31.03.2021.
- 13 H. Koltai, C. Weingarten-Baror, *Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction*. *Nucleic Acids Research*, 2008; 36: 2395–2405. <https://doi.org/10.1093/nar/gkn087>

- 14 Integrated DNA Technologies. *Fluorescence and Fluorescence Applications*. Verkkoaineisto. https://www.researchgate.net/profile/Eric-Devor/publication/251811656_Fluorescence_and_Fluorescence_Applications/links/53dfa57c0cf2a768e49bc48d/Fluorescence-and-Fluorescence-Applications.pdf Luettu 31.03.2021.
- 15 G. Rebala, A. Ravi, S. Churiwala. *An Introduction to Machine Learning*. Switzerland, Springer Nature AG; 2019. ISBN: 978-3-030-15729-6 (e-kirja).
- 16 E. Alpaydin. *Introduction to Machine Learning*, Third Edition. Cambridge, The MIT Press; 2014. ISBN: 978-0-262-02818-9 (e-kirja).