

**Karri Kalliala**

**Koneoppimisen soveltaminen sydän- ja verisuonitautien  
hoidossa**

Tietotekniikan kandidaatintutkielma

9. toukokuuta 2022

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

**Tekijä:** Karri Kalliala

**Yhteystiedot:** karri.h.kalliala@student.jyu.fi

**Ohjaaja:** Timo Tiihonen

**Työn nimi:** Koneoppimisen soveltaminen sydän- ja verisuonitautien hoidossa

**Title in English:** Application of Machine Learning in the treatment of cardiovascular diseases

**Työ:** Kandidaatintutkielma

**Opintosuunta:** Kaikki opintosuunnat

**Sivumäärä:** 27+0

**Tiivistelmä:** Sydän- ja verisuonisairaudet ovat johtava kuolinsyy maailmanlaajuisesti, ja niiden tuottama kuormitus vaivaa etenkin maita, joiden resurssit eivät riitä perusterveydenhuoltoon. Tässä tutkielmassa perehdytään, miten koneoppimista on sovellettu sydän- ja verisuonitautien hoidossa ja ennaltaehkäisyssä, sekä pohditaan kirjallisuuden perusteella optimaalista ajankohtaa koneoppimisen käytölle sairauden kulun ajalta. Tutkielmassa huomattiin, että koneoppimista käytetään sairauden eri vaiheissa hyvin kattavasti ja monipuolisesti, mutta yksikään vaiheista ei noussut toista merkittävämmäksi.

**Avainsanat:** Koneoppiminen, Tekoäly, Hoito, Sydän- ja verisuonitaudit, SVT,

**Abstract:** Cardiovascular disease is the leading cause of death worldwide, and its burden is particularly severe in countries with insufficient resources for primary health care. This dissertation examines how machine learning has been applied in the treatment and prevention of cardiovascular disease, and considers the optimal time for the use of machine learning during the course of the disease. The dissertation found that machine learning is used very comprehensively and in various ways in different stages of the disease, but none of the stages became more significant than the other.

**Keywords:** Machine Learning, ML, Artificial Intelligence, AI, Treatment, Cardiovascular disease, CVD

## Termiluettelo

CA	Coronary angiography, koronaariangiografia eli sepelvaltimoiden varjoainekuvaus
CDSS	Clinical Decision Support System, eli kliinisen päätöksenteon tukijärjestelmä
EKG	Elektrokardiogrammi, sydänsähkökäyrä
Eteisvärinä	Sydämen rytmihäiriö, jolle on ominaista useiden ylimääräisten tahdistinalueiden aiheuttama eteisten värisevä liike
HFrEF	Heart Failure with reduced Ejection Fraction, eli systolinen sydämen vajaatoiminta
Holtertutkimus	Holtertutkimuksessa tutkittavalta rekisteröidään sydänsähkökäyrän muutoksia yli vuorokauden ajalta.
Sinusrytmi	Sydämen normaali rytmi, jonka tahdistajana toimii sinussolmuke
SPECT	Single Photon Emission Computed Tomography, eli sydänlihaksen perfuusion gammakuvausmenetelmä
SVT	Sydän- ja verisuonitauti
TT-kuvaus	Tietokonetomografia (Computer Tomography)

## **Kuviot**

Kuvio 1. Ylisovittaminen (IBM 2021) .....	9
Kuvio 2. Päättöspuu (Sitar-Taut ym. 2009) .....	9

## **Taulukot**

Taulukko 1. Koneoppimisalgoritmien tarkkuus Framingham Risk Score-datalla, EKG-datalla sekä näiden datojen yhdisteellä (Jayasudha ym. 2021) .....	11
Taulukko 2. Tiainen ym. (2006) Suomennos Glasgow-Pittsburgh Cerebral Performance Category - luokituksesta .....	14

# Sisällys

1	JOHDANTO .....	1
2	SYDÄN- JA VERISUONITAUTIEN DIAGNOSTIIKKA .....	2
2.1	Sydän- ja verisuonitaudit.....	2
2.2	Lääketieteellinen diagnostiikka ja ennaltaehkäisy menetelmät .....	3
2.3	Informaatioteknologian käyttö kliinisessä päätöksenteossa .....	4
3	KONEOPPIMINEN KLIINISESSÄ KÄYTÖSSÄ .....	5
3.1	Koneoppiminen.....	5
3.2	Data ja sen muodot .....	7
3.3	Koneoppimisen käytön haasteet .....	7
4	KONEOPPIMISEN SOVELTAMINEN SVT:N ERI VAIHEISSA .....	10
4.1	Koneoppiminen varhaisessa tunnistamisessa .....	10
4.2	Koneoppiminen diagnoosissa .....	12
4.3	Koneoppiminen hoidon aikana .....	13
5	YHTEENVETO .....	16
	LÄHTEET .....	18

# 1 Johdanto

Sydän- ja verisuonitaudit eli SVT:t ovat tällä hetkellä maailman yleisin kuolinsyy (THL 2020). Deaton ym. (2011) mukaan sekä kehittyvissä että kehittyneissä maissa sydän- ja verisuonitautien riskitekijät ovat esiintymismääriltään nousussa. SVT:t voivat aiheutua geneettisistä tekijöistä tai elintapoihin liittyvistä tekijöistä, kuten liikalihavuudesta, tupakan poltosta, alkoholin käytöstä ja stressistä (Ahn ym. 2021).

Koneoppiminen voidaan määritellä tekoälyn ja tietotekniikan haaraksi, jossa pyritään imitoimaan ihmisten oppimistapoja datan ja algoritmien avulla. Viime vuosien aikana Big Data -teknologioiden kehittyminen on Duan, Edwards ja Dwivedi (2019) mukaan vienyt tekoälykehitystä eteenpäin, mikä on osaltaan edistänyt lääketieteen alalla muun muassa kliinisen päätöksenteon tukemista ja tautien ennaltaehkäisyä.

Tämän tutkielman tavoitteena on selvittää koneoppimisen käyttöä sydän ja verisuonitautien hoidon aikana sekä pohtia koneoppimisen hyödyntämisen optimaalista ajankohtaa sairauden elinkaaren aikana. Tutkimuksessa pyritään vastaamaan seuraaviin kysymyksiin:

1. Miten koneoppimista on sovellettu sydän- ja verisuonitautien eri vaiheissa?
2. Missä vaiheessa sydän- ja verisuonitautien elinkaarta koneoppimisesta on eniten todettua hyötyä?

Tutkielma on toteutettu kirjallisuuskatsauksena. Katsauksessa haetaan vastauksia edellä mainittuihin kysymyksiin kirjallisuushaun kautta löydettyistä tutkimusartikkeleista. Luvussa 2 käsitellään sydän- ja verisuonitautien teoriaa sekä lääketieteellisen diagnostiikan perusteita ja menettelytapoja. Luvussa 3 taas käsitellään koneoppimisen teoriataustaa, dataa ja sen muotoja sekä koneoppimisen käyttöön liittyviä haasteita. Luvussa 4 käydään läpi miten koneoppimista on käytetty sydän- ja verisuonitautien elinkaaren eri vaiheissa. Luvussa 5 pyritään muodostamaan yhteenveto kirjallisuuskatsauksesta ja pohditaan koneoppimisen käytölle optimaalista ajankohtaa SVT:n elinkaaren aikana.

## 2 Sydän- ja verisuonitautien diagnostiikka

Sydän ja verenkiertoelimistö mahdollistavat muiden elinten toiminnan, mutta ovat myös monin tavoin haavoittuvainen osa ihmisen elimistöä laaja-alaisuutensa vuoksi. Niihin kohdistuvat sairaudet, eli sydän- ja verisuonitaudit, ovat sekä potilaalle että hoitavalle taholle usein vaikeita tunnistaa ja diagnosoida lievien tai vähäisten oireiden takia (Ahn ym. 2021). Tässä luvussa esitellään tietoa sydän- ja verisuonitaukeista, niiden diagnostiikkaa sekä informaatioteknologian käyttöä kliinisen päätöksenteon tukena.

### 2.1 Sydän- ja verisuonitaudit

Sydän- ja verisuonitaudit ovat sydämen tai verenkierron heikkenemiseen tai häiriöihin liittyviä sairauksia. Yleisimpiä sydän- ja verisuonitaukeja ovat sepelvaltimotauti, sydämen vajaatoiminta ja aivoverenkiertohäiriöt (THL 2020). Enemmistö tavanomaisista sydän- ja verisuonitaukeista on niin sanottuja monitekijäisiä tauteja eli taudin syntyyn vaikuttavat perinnöllisten tekijöiden lisäksi elintavat ja muut ympäristötekijät (Airaksinen 2016, s.76). Esimerkiksi sydäninfarktin merkittävimpiä riskitekijöitä ovat INTERHEART-tutkimuksen mukaan ovat tupakointi, dyslipidemia eli veren rasva-arvojen poikkeavuus, kohonnut verenpaine, diabetes, keskivartalolihavuus, stressi, riittämätön kasvisten syöminen, säännöllinen alkoholin kulutus ja fyysinen inaktiivisuus (Yusuf ym. 2004).

WHO (2021) mukaan sydän- ja verisuonitaudit aiheuttavat noin kolmasosan kaikista maailman kuolemista, joista yli kolme neljäsosaa tapahtuu matalan ja keskitulotason maissa, joissa usein ihmisillä ei ole mahdollisuutta perusterveydenhuoltoon. Tämä estää SVT:n riskitekijöiden ja taudin varhaisen tunnistamisen potilaalta, mikä taas voisi ehkäistä tulevia suurempia ja kalliimpia operaatioita (WHO 2021). Myös kehittyneissä maissa SVT:t aiheuttavat edelleen paljon kuolemia; Suomessa sydän- ja verisuonitaudit aiheuttavat noin puolet työikäisten kuolemista (THL 2020).

Suuren kuolleisuuden lisäksi SVT:t aiheuttavat myös paljon hoito- ja lääkekuluja. Suomalaisen sydän- ja verisuonitaukeihin liittyvät lääkekulut vuonna 2019 ylsivät 191 miljoonaan euroon (THL 2020). Kulujen ja kuolleisuuden pienentämiseen voisi toimivalla ennaltaehkäi-

systrategialla olla merkittävä vaikutus, kun tauti tunnistettaisiin ajoissa.

## **2.2 Lääketieteellinen diagnostiikka ja ennaltaehkäisy menetelmät**

Airaksinen (2016, s.296) mukaan valtimosairauksien ennaltaehkäisy terveissä ihmisissä perustuu kahteen lähestymistapaan: väestöstrategiaan ja riskialttiiden yksilöiden tunnistamiseen ja hoitamiseen. Väestöstrategiassa pyritään vaikuttamaan koko populaatioon vähentämällä vaaratekijöitä ja muokkaamalla elintapoja ja -ympäristöä terveellisemmiksi. Tästä esimerkkinä Airaksinen (2016, s.310) kertoo työikäisten sepelvaltimotaudin sairastavuuden ja kuolleisuuden vähentyneen merkittävästi viimeisten vuosikymmenien aikana. Kokonaisuudessaan sairastavien määrä ja hoidon tarve ei ole kuitenkaan laskenut eli taudin luonne on muuttunut tappavasta kroonisen ja vakaan tautimuodon suuntaan (Airaksinen 2016, s.310).

Airaksinen (2016, s.92) mukaan sydänsairauksien diagnostiikalle ja hoidon seurannalle luodaan vankka pohja kliinisellä tutkimuksella yhdessä anamneesin, eli haastatteleamalla kerätyn potilashistorian, kanssa. Merkittävä osa informaatiosta saadaan yksinkertaisin keinoin: kyselemällä, katselemalla, kuuntelemalla ja tunnustelemalla. Kliiniset tutkimusmenetelmät ovat myös halpoja ja välittömästi saatavilla, minkä lisäksi niiden avulla koneelliset jatkok tutkimukset ja niiden kiireellisyys pystytään kohdentamaan potilaisiin, joilla diagnoosin tarkentaminen ja sairauden vaikeusasteen määrittäminen vaikuttavat jatkohoitoon (Airaksinen 2016, s.92).

Sydänsairauden tarkka diagnostiikka vaatii anamneesin ja kliinisen tutkimuksen lisäksi usein vähintään EKG:n, joitakin laboratoriokokeita, keuhkokuvan ja usein sydämen kaikututkimuksen (Airaksinen 2016, s.93). Ahn ym. (2021) mukaan yleisimmät sydän- ja verisuonitauteihin liittyvät kliiniset tutkimukset ovat sydämen kaikututkimus, SPECT-gammakuvaus, EKG, holteritutkimus ja sepelvaltimon TT-kuvaus. Näistä tutkimuksista EKG:ta kehdetaan tärkeäksi ja potilaalle harmittomaksi tavaksi ennustaa ja diagnosoida sydän- ja verisuonitauteja (Liu ym. 2021).



## 2.3 Informaatioteknologian käyttö kliinisessä päätöksenteossa

Kliinisessä ympäristössä liikkuu valtavasti tietoa, kuten potilaiden perustietoja, tutkimuskuvia ja -signaaleja sekä laboratoriotuloksia. Johnson ym. (2018) mukaan lääkäreiden ja muiden kliinikoiden on jo pitkään täytynyt osata tunnistaa, mitata ja tulkita erilaisten muuttujien välisiä suhteita. Lääkäreiden käyttöön tulvii koko ajan lisää ammattitaitoa vaativaa tietoa ja samalla heiltä odotetaan tehokkaampaa toimintaa (Johnson ym. 2018).

Tietotulvaan ratkaisuna ovat toimineet tietotekniset ja automatisoidut menetelmät kliinikkojen apuna. Esimerkiksi Airaksinen (2016, s.131) mukaan EKG-laitteissa on jo vuosia ollut automaattisia diagnostisia algoritmeja, jotka auttavat lääkäreitä diagnosoinnissa. Nämä algoritmit ovat ajan kuluessa parantuneet ja haastavat nykyisellään jo kokeneemmankin EKG:n tulkitsijan. Koneellisesti tulkittavan EKG-käyrän täytyy kuitenkin olla teknisesti laadukas, jotta algoritmi pystyy diagnosoimaan tarkasti ja tunnistamaan käyrän epänormaalisuudet oikein. EKG-algoritmeilla on edelleen selviä heikkouksia, kuten rytmihäiriöiden tulkinta. Lisäksi algoritmeja on tietoisesti yliviritetty havaitsemaan mahdolliset sairauksista johtuvat EKG-muutokset, mikä myös osaltaan lisää ylidiagnostiikkaa (Airaksinen 2016, s.131).

Ahn ym. (2021) mukaan tekoälyteknologioiden viimeaikainen kehitys on mahdollistanut useiden sairauksien varhaisen havaitsemisen ja osoittanut lääkäreitä parempaa kykyä diagnosoida sairauksia. Johnson ym. (2018) uskoo koneoppimisen olevan ratkaisu kliinisen informaation hyödyntämiselle sen voidessa tehostaa jokaista hoitovaihetta, tutkimuksesta diagnoosiin ja hoidon valintaan. Koneoppimisen sisällyttämisen tuloksena kliininen työskentely muuttuu tehokkaammaksi, kätevämmäksi ja yksilöllistetyimmäksi (Johnson ym. 2018).

### **3 Koneoppiminen kliinisessä käytössä**

Idea koneoppimisesta kehittyi jo varhain 1950-luvulla, ja IBM (2020) artikkelin mukaan termi "koneoppiminen" on Arthur Samuelin kehittämä. Samuel julkaisi vuonna 1959 oman tutkimuksensa koneoppimisen käytöstä Tammi-pelissä. Lääketieteellisessä analysoinnissa ja tiedonlouhinnassa tekoälyteknologiat alkoivat nousta esiin 1980-luvulla (Hamet ja Tremblay 2017). Terveystieteissä koneoppimista ja tiedonlouhintaa on Oswald, Jaya Sathwika ja Bhattacharya (2022) mukaan käytetty sopivien hoitomenetelmien määrittelyyn, sairauden riskitekijöiden ennakointiin ja kustannustehokkaiden hoitorakenteiden päättämiseen. Tässä luvussa esitellään kliinisestä näkökulmasta tarkasteltuna koneoppimisen perusteita ja algoritmeja, koneoppimismalleille syötettävää dataa sekä mallien analyttistä tulkintaa.

#### **3.1 Koneoppiminen**

Tekoäly on yleistermi, joka tarkoittaa älykkään käyttäytymisen mallintamista tietokoneella ilman ihmisen ohjausta (Hamet ja Tremblay 2017). Koneoppimista kuvaillaan sekä tekoälyn merkittäväksi osa-alueeksi (Nakamura ja Sasano 2022) että joukoksi menetelmiä, jotka mahdollistavat tekoälyn (Johnson ym. 2018). IBM (2020) artikkelissa koneoppiminen määritellään tekoälyn ja tietotekniikan haaraksi, joka perustuu ihmisten oppimistapojen imitointiin datan ja algoritmien avulla. Koneoppiminen on siis termi joukolle algoritmeja ja menetelmiä, joiden päätarkoituksena on aiempaa kokemusta ja uutta informaatiota yhdistelemällä omaksua uutta tietoa. Tätä tietoa voidaan hyödyntää esimerkiksi erilaisten ennusteiden ja toimintamallien kehittämisessä.

Koneoppimismallit ovat koneoppimismenetelmin koulutettuja matemaattisia malleja, joille on luonteenomaista muovautua itsenäisesti koulutusdatan mukaiseksi (Nakamura ja Sasano 2022). Koneoppimismallin asetukset ja piirteet siis määrittyvät opetuksen aikana vastaanottamaan dataa, jolla mallia opetetaan. Monet koneoppimisalgoritmit pohjautuvat enimmäkseen tilastotieteellisiin menetelmiin kuten regressioon. Alpaydin (2014, s.3) kertoo koneoppimisen käyttävän tilastotieteen teoriaa matemaattisten mallien rakentamisen pohjana, sillä kuten tilastotieteessä, myös koneoppimisessa on tarkoitus tehdä päätelmiä otoksesta. Johnson

ym. (2018) määritelmässä klassista tilastotiedettä ja klassista koneoppimista ei määritellä tarkasti toisistaan eroaviksi, vaan ne nähdään samankaltaisina, joskin usein eri kysymyksiin vastaavina. Tilastotieteessä keskitytään johtamaan päätelmiä otos- tai populaatioparametreista, kun taas koneoppimisessa pääpaino on havainnollistaa algoritmisesti datan rakennetta ja tehdä siitä ennusteita tai luokittelua (Johnson ym. 2018).

Koneoppiminen voidaan kategorisoida ohjattuun ja ohjaamattomaan oppimiseen sekä vahvistusoppimiseen. Nakamura ja Sasano (2022) mukaan useimmat lääketieteessä käytetyt tekoälymallit ovat ohjattua oppimista eli luokitteluun ja ennusteiden tekemiseen keskittyviä koneoppimismalleja. Johnson ym. (2018) mukaan ohjatussa oppimisessa yritetään mallintaa riippumattomien muuttujien, esimerkiksi iän, sukupuolen ja ammatin, suhdetta riippuvaan vastemuuttujaan, esimerkiksi sairauden riskiin. Lääketieteessä muuttujien suhteet saattavat olla hyvinkin monimutkaisia, sillä esimerkiksi jollain lääkkeellä saattaa olla merkityksellistä vuorovaikutusta toisen lääkkeen kanssa vain tietyissä olosuhteissa (Johnson ym. 2018).

Toisin kuin ohjatussa oppimisessa, ohjaamattomassa oppimisessa ei ole riippumatonta vastemuuttujaa, eikä muuttujien välisistä suhteista tiedetä mitään ja pyritään välttämään muuttujiin liittyviä oletuksia (Alashwal ym. 2019). Ohjaamattoman oppimisen menetelmät keskittyvät löytämään datan mallista hahmoja ja muuttujien välisiä suhteita (Johnson ym. 2018; Hamet ja Tremblay 2017). Ohjaamattomassa oppimisessa pyritään siis tiedonlouhinnan tavoin johtamaan algoritmisesti tietoa aineistosta, jota ei ole kerätty tilastollista analyysia varten. Yhtä ohjaamattoman oppimisen algoritmia kutsutaan klusteroinniksi, jossa algoritmi yrittää havaita datasta samankaltaisista havainnoista koostuvia ryhmittymiä eli klustereita. Alashwal ym. (2019) mukaan lääketieteen alalla klusterointi on osoittautunut tehokkaaksi työkaluksi kuvioiden ja datan rakenteen tunnistamisessa.

Hamet ja Tremblay (2017) mukaan vahvistusoppiminen tuottaa toimintamalleja tai toimintasekvenssejä. Oppimisprosessi perustuu toivotun käytöksen palkitsemiseen ja/tai ei-toivotun käytöksen rankaisemiseen (Hamet ja Tremblay 2017). Tällaisen toimintamallin solmukohta kertoo siis käyttäjälle tai sitä käyttävälle ohjelmalle, mikä olisi paras tapa jatkaa tästä vaiheesta eteenpäin. Johnson ym. (2018) mukaan vahvistusoppimisen käyttö lääketieteessä on ollut harvinaista. Potentiaalisia käyttökohteita vahvistusoppimisalgoritmeille ovat potilaan ominaisuuksien perusteella mukautuvat hoito-ohjelmat (Johnson ym. 2018).

## 3.2 Data ja sen muodot

Silverio ym. (2019) mukaan datan määrä kaksinkertaistuu joka toinen vuosi. Terveystieteiden kontekstissa tyypillinen suurikokoinen sairaala tuottaa noin 100 teratavua dataa vuodessa, ja Silverio ym. (2019) esimerkin mukaan yksi gigatavu vastaa noin 6 miljoonaa kirjaa (1Tb = 1000 Gb). Sähköiset terveystiedot sisältävät monenlaista informaatiota, minkä lisäksi ne mahdollistavat tarkkojen tietojen kirjaamisen ja säilyttämisen sairaalakäynnistä tai saadusta hoidosta (Silverio ym. 2019). Sähköisten terveystietojen lisäksi Miotto ym. (2017) mukaan modernissa lääketieteellisessä tutkimuksessa on noussut esiin muun muassa kuvantamista, edistyneitä analytiikkaa (esim. genomiikkaa), sensorien tuottamaa dataa sekä kompleksista tekstimuotoista dataa. Lääketieteellisen tekstimuotoisen datan keskiössä ovat potilaskertomukset, jotka saattavat olla edelleen joissain maissa käsin kirjoitetussa muodossa. Jotta käsin kirjoitetusta tekstistä on saatu dataa tietokoneen ymmärtämään muotoon, on Ahn ym. (2021) tutkimuksessa käytetty luonnollisen kielen käsittelyä (Natural Language Processing), joka on koneoppimista hyödyntävä tekoälyn muoto.

Data voidaan yleisesti kategorisoida strukturoituun ja strukturoimattomaan eli rakenteelliseen ja rakenteettomaan dataan. Silverio ym. (2019) kuvailee rakenteellista dataa erittäin järjestelmälliseksi sekä helposti käsiteltäväksi ja tulkittavaksi tiedoksi. Esimerkkejä rakenteellisesta datasta ovat muun muassa laboratoriotulokset ja sähköiset terveystiedot. Rakenteettomassa datassa ei rakenteellisesta poiketen ole ennaltamääritettyä mallia, minkä lisäksi se voi olla tekstimuodon sijaan myös jossakin muussa muodossa, kuten kuvana. Rakenteettomasta tiedosta voi kuitenkin olla hyötyä kokonaiskuvan muodostamisessa, sillä se saattaa tuoda ilmi muun muassa potilaan terveyteen vaikuttavia ympäristötekijöitä (Silverio ym. 2019). Lisäksi Ahn ym. (2021) mukaan rakenteettomien tietojen, kuten signaalidatan ja lääketieteellisten kuvien, analyysi syväoppimisalgoritmeilla on tuottanut parempia tuloksia kuin rakenteellisilla sähköisillä potilastiedoilla tehdyt mallit.

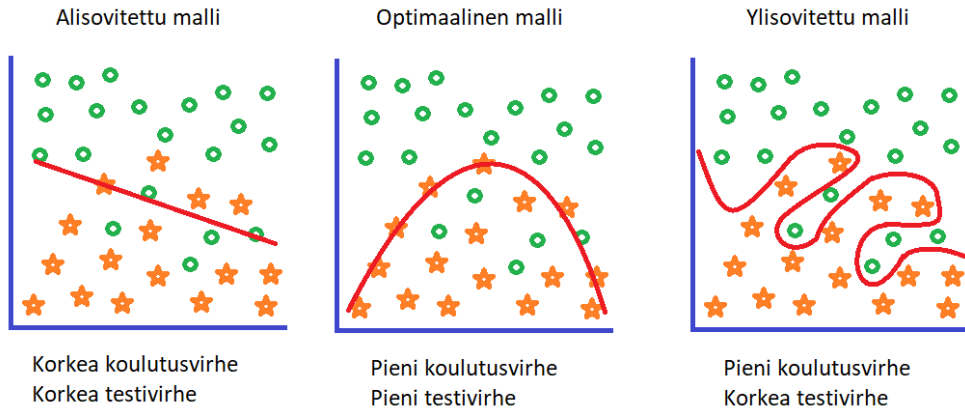
## 3.3 Koneoppimisen käytön haasteet

Deo (2015) väittää lääketieteellisen alan vastuullisen luonteen olevan osasy sille, ettei koneoppimisalgoritmeja vielä käytetä laajalti kliinisessä ympäristössä. Esimerkiksi hoitovir-

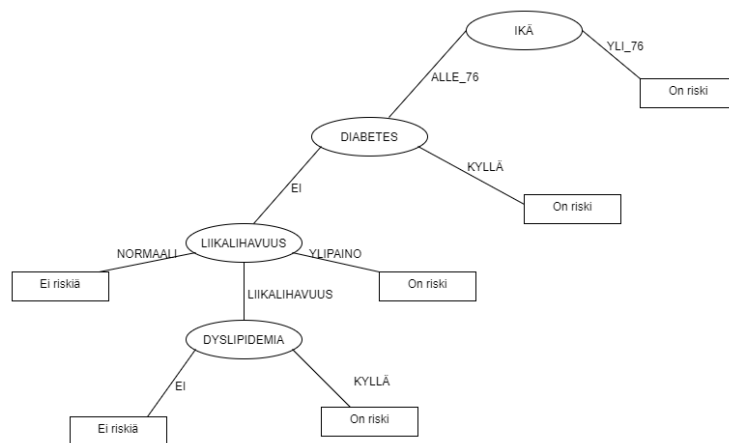
heiden korvausvelvollisuus ja vastuu hoidon kulusta sekä tietämättömyys koneoppimisesta ja siihen liittyvän ymmärryksen puute luovat tilanteen, jossa koneellinen apu tuntuu epämu-kavalta. Näiden syiden vuoksi siihen ei uskalleta luottaa, vaikka joissain tapauksissa olisi selvää koneoppimismenetelmän tuovan tarkempia tuloksia (Deo 2015).

Bi ja Ma (2021) mukaan lääketieteessä käytettävä koneoppimiselle relevantti data on tyypil-lisesti epätasapainoista eli terveiden ihmisten osuus datasta on suuri, kun taas sairaiden osuus on pieni. Mitä vakavampi ja harvinaisempi tauti on kyseessä, sitä enemmän datan epätasapai-noisuuden vaikutus korostuu. Tämä olisi tärkeää huomioida etenkin mallien validoinnissa ja testaamisessa. Muuten mallien yleistymiskyvystä voidaan tehdä vääriä johtopäätöksiä. Esi-merkiksi jos malli oppii luokittelemaan epätasapainoisesta datasta kaikki havainnot terveiksi, niin mallin tarkkuus voi olla lähes sata prosenttia, mutta taas sensitiivisyys, eli oikeiden po-sitiivisten tulosten osuus kaikista tutkittavista, on lähellä nollaa, joten mallista ei tule pätevää luokittelijaa.

Oswald, Jaya Sathwika ja Bhattacharya (2022) mukaan tämän hetkiset koneoppimismallit kärsivät datan ylisovittamisesta (Kuvio 1) ja heikosta tarkkuudesta. Esimerkiksi Johnson ym. (2018) mukaan päätöspuut (Kuvio 2) kärsivät ylisovittamisesta, sillä ne ovat korkean varianssin 'oppijoita' eivätkä yleisty helposti muuhun dataan. Tätä ongelmaa välttääkseen päätöspuissa käytetään bootstrap-menetelmiä, joilla simuloidaan yhden tietojoukon pilkko-mista osiin ja pyritään näin helpottamaan yleistymisongelmaa.



Kuvio 1. Kuviossa oikeanpuoleisin malli on ylisovittunut opetusdataan, eikä se yleisty enää helposti uudelle datalle. Kuva mukailtu (IBM 2021) teoksen mukaan.



Kuvio 2. Päätöspuussa luokitellaan potilaiden sepelvaltimotautiriskiä. Kuva mukailtu (Sitar-Taut ym. 2009) teoksen mukaan.

## 4 Koneoppimisen soveltaminen SVT:n eri vaiheissa

Johnson ym. (2018) mukaan sydän- ja verisuonitauteja tutkiva lääketiede voi hyötyä tekoälystä ja etenkin koneoppimisesta enemmän kuin aikaisemmista metodeista, kuten tilastotieteistä. Tilastollisissa malleissa joudutaan tekemään oletuksia ilmiöistä ja havainnoista. Kun nämä oletukset eivät ole voimassa, voidaan koneoppimisella saada tarkempia malleja ja tuloksia (Johnson ym. 2018).

Syväoppiminen on koneoppimisen yksi menestyneimpiä osa-alueita, jonka neuroverkkomallit sisältävät yhden piilokerroksen sijaan monia piilokerroksia. Erästä syväoppimisalgoritmia, konvoluutioneuroverkkoa, käytetään esimerkiksi kuvien analysointiin (Farah ym. 2021), joka vaatii grafiikkasuorittimelta paljon tehoa. Syväoppimisen merkitys on kasvanut viime vuosien aikana johtuen edistyksestä grafiikkasuorittimien rinnakkaislaskennassa (Johnson ym. 2018). Johnson ym. (2018) myös toteavat, että syväoppimismallit edustavat uusinta teknologiaa ja lääkäreiden tulisi ymmärtää niiden mahdollisuudet kliinisessä käytössä.

Sydän- ja verisuonitaudit vaativat aktiivista ja pitkäaikaista hoitoa. Tässä tutkielmassa sydän- ja verisuonitautien elinkaari jaetaan kolmeen tutkittavaan vaiheeseen. Ensimmäinen vaihe sisältää SVT:n varhaisen tunnistamisen. Toiseen vaiheeseen kuuluu sairauden diagnosointi, ja kolmanteen vaiheeseen hoito ja kuntouttaminen. Jokaiselle vaiheelle on tehty oma aineistohaku, joista kirjallisuuskatsaukseen on valittu sopivimmat tutkimukset.

### 4.1 Koneoppiminen varhaisessa tunnistamisessa

Tässä alaluvussa käydään läpi kaksi tutkimusta, joissa on pyritty parantamaan koneoppimisen avulla SVT-ennustemalleja. Tutkimuksissa perehdytään tarkkailemaan koneoppimismalleille syötettävän datan, kuten EKG-muuttujien tai potilastietokantojen merkitystä mallien suorituskykyyn.

Jayasudha ym. (2021) tutkimuksen päätavoitteena oli vertailla eri koneoppimismetodien tarkkuutta ja rakentaa moniulotteinen tietokanta, joka on tarkempi ja monin tavoin hyödynnettävissä. Tarkkailun kohteena vertailuissa olivat suorituskykymittarit, kuten mallin tark-

kuus (accuracy). Tutkimuksessa kerättiin dataa Framingham Risk Score-datan ja EKG-datan muuttujilla, jonka jälkeen kerätystä datasta luotiin yksi yhdistetty tietokanta. Yhdistetyn tietokannan tarkoitus on Jayasudha ym. (2021) mukaan mahdollistaa paremmat ennustemallit sydänsairauksille monipuolisella FRS-EKG-datalla. Tutkimuksen tuloksena kaikkien menetelmien virheiden määrä tippui alle kymmenesosaan siirryttäessä yhdistetyllä datalla koulutamiseen.

Data	Satunnaismetsä	Logistinen regressio	Naiivi Bayes	Neuroverkko
FRS	64.47%	69.73%	72.3%	67%
EKG	83.18%	79.64%	74.33%	71%
Yhdistetty	96.98%	96.23%	91.95%	97%

Taulukko 1. Koneoppimisalgoritmien tarkkuus Framingham Risk Score-datalla, EKG-datalla sekä näiden datojen yhdisteellä (Jayasudha ym. 2021)

Li, Hu ja Liu (2021) tutkimuksessa päätavoitteena oli luoda koneoppimismalli SVT:n ennustamiseen, perustuen EKG- ja FKG-signaaleihin (fonokardiogrammi). Tämänkin tutkimuksen, kuten Jayasudha ym. (2021), tavoitteena oli löytää tarkempi EKG-ennustemalli yhdistämällä kaksi erilaista tietojoukkoa. Sydän- ja verisuonitautien monimuotoisuuden ja lukuisten alalajien takia yhden työkalun käyttö on Li, Hu ja Liu (2021) sanoin riittämätöntä piilevien sydän- ja verisuonitautien hahmottamiseen, jonka takia tutkijat kehittivät montaa koneoppimisalgoritmia eri kerroksissa hyödyntävän sekä moniulotteista dataa hyödyntävän mallin. Tutkimuksen tuloksena 'multimodaalinen' koneoppimismalli suoriutui paremmin kuin yhdellä datalla koulutetut mallit. Varsinkin pelkillä fonokardiogramminauhotteilla koulutetut mallit suoriutuivat merkittävästi huonommin.

Yhteenvedona tutkimuksista voidaan todeta koneoppimismallien tuottavan parempia tuloksia, kun oppimisessa käytetty data on monipuolista ja sen käyttö on tarkkaan mietittyä. SVT:n monimuotoisuuden takia laajempi näkökulma ongelmien tarkastelussa on hyödyllistä.



## 4.2 Koneoppiminen diagnoosissa

Tässä alaluvussa seurataan kolmea tutkimusta ja niiden näkökulmia koneoppimisen hyödyntämiseen diagnosointivaiheessa. Tutkimuksissa perehdytään aikaleimattujen muuttujien vaikutukseen koneoppimismalleissa, sepelvaltimon kudosterrosten luokitteluun ja diagnosointia avustavaan AI-CDSS:ään.

E. Choi ym. (2016) tutkimuksen tavoitteena oli tutkia, vaikuttaako aikaleimattujen muuttujien sisällyttäminen mallin koulutukseen sen kykyyn ennustaa sydämen vajaatoiminnan ensidiagnoosia. Aikaleimattuja muuttujia tutkimuksessa olivat esimerkiksi reseptin määräyshetki ja toimenpiteen aloitusaika. Tutkimuksessa myös vertailtiin eri koneoppimisalgoritmien suorituskykyä aikaleimattujen muuttujien kanssa. Tutkimuksen yhteydessä kehitetty takaisinkytkentäneuroverkko (Recurrent Neural Network, RNN) menestyi vain vähän muita koneoppimismalleja paremmin, kun mallin tarkkailu- ja ennustusikkuna oli pituudeltaan 12 kuukautta. Kun aikaikkunaa pidennettiin 18 kuukauteen, RNN suoriutui huomattavasti vertailtavia algoritmeja paremmin. Vertailtavia koneoppimisalgoritmeja olivat regularisoitu logistinen regressio, monikerroksinen perseptroniverkko, tukivektorikone sekä k-lähin naapuri-menetelmä. Tuloksena huomattiin aikaleimattujen muuttujien sisällyttämisen koulutukseen parantavan mallien kykyä ennustaa sydämen vajaatoiminnan ensidiagnoosia.

Abdolmanafi ym. (2017) tutkimuksessa kehitettiin koneoppimismalli, joka oppi luokittelemaan sepelvaltimon sisäisiä kudosterroksia optisista koherenssitomografiakuvista (OCT). Koneoppimismallin tarkoituksena oli helpottaa erään lasten sepelvaltimotaudin, Kawasakin taudin, diagnosointi-, hoito- ja seurantaprosessia. Tutkijat huomasivat, että valmiiksi koulutetun konvoluutioneuroverkon (CNN) käyttö piirteiden erottelijana ja satunnaismetsän käyttö luokittelijana toi malliin tarkkuutta ja nopeutta, luokittelun tarkkuuden (accuracy) noustessa 96 prosenttia. Käyttäessä konvoluutioneuroverkkoa sekä piirteiden erotteluun että luokitteluun, saatiin tarkkuutta nostettua yhdellä prosenttiyksiköllä, mutta mallin koulutusaika oli huomattavasti pidempi.

D.-J. Choi ym. (2020) tutkimuksen tarkoituksena oli arvioida tekoälyllisen kliinisen päätöksenteon tukijärjestelmän (AI-CDSS) tarkkuutta sydämen vajaatoiminnan diagnoosissa. Koneoppimista käytettiin AI-CDSSn luomisessa vertaillen viittä eri algoritmia, joista lopulta

käyttöön päätyi korkeinta tarkkuutta näyttänyt CART, eli luokittelu- ja regressiopuu (Classification And Regression Tree). Lääkäreiden luoma CDSS yhdistettiin koneoppimismalliin, josta tehtiin hybridimalli AI-CDSS. Tutkimuksessa käytettiin suorituskykymittarina konkordanssiastetta, eli tässä tilanteessa yhdenmukaisuutta verrokkiparin diagnooseissa. Tutkimuksessa AI-CDSS ja erikoislääkäri diagnoosivat 98 prosenttia ajasta yhdenmukaisesti, kun taas erikoislääkärin ja yleislääkärin diagnoosit olivat yhdenmukaisia 76 prosenttia ajasta. Tästä johtuen tutkimustuloksena AI-CDSS:n todettiin olevan hyödyllinen silloin, kun paikalla ei ole erikoistuneita lääkäreitä tekemässä lopullista diagnoosia.

Yhteenvedona näistä tutkimuksista voidaan todeta niiden tarkastelevan SVT:n diagnosointia eri näkökulmista hyvin tuloksin. Sydämen vajaatoiminnan ensidiagnoosi voidaan ennustaa käyttäen mallin koulutuksessa aikaleimattuja muuttujia ja varsinaisen diagnoosin avuksi on kehitetty lähes erikoislääkärin tasoinen kliinisen päätöksenteon tukijärjestelmä. Myös lasten sepelvaltimoiden sisäisiä kudoskerroksia osataan luokitella, mikä auttaa Kawasakin taudin diagnosoinnissa ja hoidon seurannassa.

### **4.3 Koneoppiminen hoidon aikana**

Tässä alaluvussa seurataan neljää tutkimusta ja niiden näkökulmia koneoppimisen käytöstä hoidon aikana ja hoidon seurantavaiheessa. Tutkimuksissa tutkittiin sairaalassa tai sairaalan ulkopuolella tapahtunutta sydänpysähdystä (IHCA—OHCA) sekä koneoppimisen käyttöä näissä tilanteissa. Tutkimuksissa yhtenä potilaan toipumisen mittarina oli sydänpysähdysten jälkeinen aivojen toiminnallinen tila, joka on jaettu viiteen CPC-luokkaan (Cerebral Performance Category, Taulukko 2)

Mayampurath ym. (2022) tutkimuksessa vertailtiin eri koneoppimis- ja tilastomalleja, joilla ennustetaan IHCA-potilaiden todennäköisyyttä kotiutua sairaalasta suotuisien neurologisten oireiden kanssa (CPC-luokka yhtäsuuri tai pienempi kuin 2). Vertailussa oli mukana viisi koneoppimisalgoritmia; logistinen regressio, tukivektorikone, satunnaismetsä, Extreme Gradient Boosting sekä monikerroksinen perseptroniverkko, ja tilastollisin menetelmin luotu CASPRI-malli (Cardiac Arrest Survival Post-Resuscitation In-hospital). Extreme Gradient Boosting -malli oli kaikilla mittareilla tehokkaampi muihin malleihin verrattuna.

CPC-luokka	Tila
1	Tajuissaan, täysin omatoiminen, ei viitettä aivovauriosta
2	Tajuissaan, täysin omatoiminen, lievä kognitiivinen häiriö tai neurologinen poikkeavuus
3	Tajuissaan, tarvitsee runsaasti apua päivittäisissä toimissaan
4	Pysyvästi tajuton
5	Kuollut

Taulukko 2. Tiainen ym. (2006) Suomennos Glasgow-Pittsburgh Cerebral Performance Category - luokituksesta

Cheng ym. (2021) tutkimuksessa vertailtiin kolmea erilaista koneoppimismallia käytettäväksi OHCA-potilaiden hoitoon liittyvien ennusteiden kanssa. Vertailtavina algoritmeina olivat logistinen regressio, tukivektorikone ja Extreme Gradient Boosting. Tutkittavina tuloksina olivat suotuisa neurologinen tila kotiutuessa, kuten Mayampurath ym. (2022) tutkimuksessa, sekä 30 päivän selviytymisaste ja todennäköisyys selviytyä sairaalasta kotiuttamiseen asti. Tutkimuksessa todettiin Extreme Gradient Boosting -algoritmin tuottavan parhaita tuloksia kolmen vertailtavan algoritmin joukosta.

Chi ym. (2021) tutkimuksessa kehitettiin syväoppimismalli, jonka tarkoituksena oli ennustaa IHCA-potilaan kuolintodennäköisyyttä sekä sairaalaan takaisinottamisen todennäköisyyttä seuraavan 30 päivän ajalle. Datan epätasapainoisuus aiheutti tutkimukselle haasteita, sillä vain noin kymmenesosa koulutusdataan valituista havainnoista olivat IHCA-potilaita. Epätasapainoisuutta korjatakseen dataa painotettiin eli IHCA-potilaiden määrää korostettiin mallin kehityksessä. Tutkijat saivat mallin ennustamaan yli 80 prosentin todennäköisyydellä IHCA-kuolleisuuden ja yli 85 prosentin todennäköisyydellä IHCA-potilaan takaisinoton. Tutkijat uskovat ratkaisunsa datan epätasapainoisuuden kanssa olevan hyvä tapa kouluttaa parempia malleja epätasapainoisesta datasta.

Harford ym. (2022) tutkimuksen tarkoituksena oli opettaa koneoppimismalleille sairaalan käyttämä päätöksentekoprosessi määrätä CA-kuvaus OHCA-potilaille. Vertailtavia algoritmeja oli 9 kappaletta, mukaan lukien Extreme Gradient Boosting-algoritmi, joka on Cheng

ym. (2021) ja Mayampurath ym. (2022) tutkimuksissa osoittautunut menestyksekkääksi. Harford ym. (2022) saivat parhaita tuloksia kehittämällään EFCN-algoritmilla (Embedded Fully Convolutional Network). Algoritmi kehitettiin kahdessa vaiheessa; ensimmäisessä vaiheessa verkon neuronit koulutettiin CA-datalla ja toisessa vaiheessa neuronien aloitusarvoiksi annettiin ensimmäisen vaiheen koulutetut neuronit ja koulutusta jatkettiin CPC-datalla. Tämä monivaiheinen konvoluutioneuroverkko toi tutkijoille parhaat tulokset mallin luokittellessa 90 prosenttia potilaista oikein. Tutkijoiden kehittämällä mallilla tehdyn analyysin mukaan noin viidesosa potilaista olisi pitänyt kuvata CA:lla, minkä johdosta noin kolmasosalla heistä selviytymismahdollisuudet olisivat parantuneet, mikäli heidät oltaisiin kuvattu.

Tutkimukset rakentuivat samalle pohjalle, niiden tarkastellen IHCA- ja OHCA-potilaista koostuvaa dataa. Tutkimuksissa löydettiin Extreme Gradient Boosting-algoritmin sopivan potilaan neurologisen tilan ennustamiseen kotiutumishetkellä ja syväoppimisalgoritmien toimivan päätöksenteon tukemiseen kehitettävien mallien koulutuksessa.

## 5 Yhteenveto

Tutkielman päätavoitteena oli vastata tutkimuskysymykseen "Miten koneoppimista on sovellettu sydän- ja verisuonitautien eri vaiheissa?" sekä pohtia kirjallisuuskatsauksen perusteella, onko koneoppimisen käytöstä eniten hyötyä jossain sydän- ja verisuonitautien vaiheista.

Toisessa luvussa esiteltiin tietoa sydän- ja verisuonitaukeista sekä niiden tuottamista ongelmista terveydenhuollossa. Toisessa luvussa myös käytiin läpi lääketieteellistä diagnostiikkaa ja informaatioteknologian käyttöä kliinisen päätöksenteon tukena. Kolmannessa luvussa esiteltiin koneoppimista, koneoppimisalgoritmille syötettävää dataa ja sen muotoja sekä koneoppimisen käytön haasteita lääketieteessä. Aiheita tutkittiin sekä yleisestä että kliinisestä näkökulmasta. Neljännessä luvussa käytiin läpi aineistohaun tuottamaa kirjallisuutta jakaen sydän- ja verisuonitautien elinkaari kolmeen vaiheeseen ja tutkien koneoppimisen käyttöä näissä vaiheissa.

Kirjallisuudesta ilmeni, että koneoppimista on käytetty hyvin laajasti sydän- ja verisuonitautien hoidossa sekä yleisluontoisemmalla potilastietokantadatalla että yksityiskohtaisemmin tutkivalla kuvantamis- ja signaalidatalla. Esimerkiksi erilaisten tietojoukkojen, kuten elektrokardiogrammista ja fonokardiogrammista saadun datan, yhdistäminen kuolutusdataksi paransi merkittävästi koneoppimismallien suorituskykyä. Tutkimuksissa huomattiin datan algoritmien monipuolisuuden ja moniulotteisuuden tuovan parempia tuloksia. Monet tutkimuksissa kehitetyt koneoppimismallit luotiin myös jossain määrin vaiheittain, hyödyntäen algoritmien parhaita ominaisuuksia tietyissä vaiheissa. Tästä esimerkkinä konvoluutioverkon käyttö piirteiden erottelijana ja satunnaismetsän käyttö luokittelijana toi koulutukseltaan nopean, mutta pätevän mallin luokitteluun sepelvaltimon sisäisiä kudoskerroksia (Abdolmanafi ym. 2017).

Koneoppimisella on suuret mahdollisuudet viedä lääketiedettä eteenpäin, sillä maailma on muuttumassa koko ajan teknologiaa hyväksyvämmäksi ja dataa sekä tietotaitoa kehittää koneoppimismenetelmiä on paljon. Kirjallisuuskatsauksen perusteella on hankalaa sanoa, onko sydän- ja verisuonitautien elinkaareissa koneoppimisen käytölle eniten hyötyä tuottavaa vaihetta, sillä näkökulmia ja käyttötapauksia oli todella paljon. Kuitenkin eettisesti ja loogisesti

mietittynä taudin varhaisen tunnistamisen vaihe on osaltaan kriittinen hetki, sillä kun tauti tunnistetaan ajoissa, voidaan elintapamuutoksilla ja lääkityksillä vaikuttaa merkittävästi suuntaan, johon tauti on kehittymässä. Lisäksi varhaisen tunnistamisen vaiheessa voitaisiin hyötyä automatisoidummista menetelmistä, sillä kohdeyleisö on luonnollisesti suurempi silloin, kun etsitään onko henkilöllä tai populaatiolla riski sairastaa sydän- ja verisuonitauteja.

## Lähteet

Abdolmanafi, Atefeh, Luc Duong, Nagib Dahdah ja Farida Cheriet. 2017. “Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography”. *Biomed. Opt. Express* 8, numero 2 (helmikuu): 1203–1220. <https://doi.org/10.1364/BOE.8.001203>. <http://opg.optica.org/boe/abstract.cfm?URI=boe-8-2-1203>.

Ahn, Imjin, Wonjun Na, Osung Kwon, Dong Hyun Yang, Gyung-Min Park, Hansle Gwon, Hee Jun Kang ym. 2021. “CardioNet: a manually curated database for artificial intelligence-based research on cardiovascular diseases”. *BMC Medical Informatics and Decision Making* 21, numero 1 (tammikuu): 29. ISSN: 1472-6947. <https://doi.org/10.1186/s12911-021-01392-2>. <https://doi.org/10.1186/s12911-021-01392-2>.

Airaksinen, Juhani. 2016. *Kardiologia*. 3., uudistettu painos. Toimittanut Katriina Aalto-Setälä, Juha Hartikainen, Heikki52 Huikuri, Mika Laine, Jyri Lommi, Pekka Raatikainen ja Antti Saraste. Sisältää hakemiston. Helsinki: Duodecim.

Alashwal, Hany, Mohamed El Halaby, Jacob J. Crouse, Areeg Abdalla ja Ahmed A. Mustafa. 2019. “The Application of Unsupervised Clustering Methods to Alzheimer’s Disease”. *Frontiers in Computational Neuroscience* 13. ISSN: 1662-5188. <https://doi.org/10.3389/fncom.2019.00031>. <https://www.frontiersin.org/article/10.3389/fncom.2019.00031>.

Alpaydin, Ethem. 2014. *Introduction to machine learning*. Third edition. Adaptive computation and machine learning. Includes index. Cambridge, Massachusetts ; London, England: The MIT Press.

Bi, Wenbin, ja Rong Ma. 2021. “Unbalanced Data Set Processing Method for Colorectal Cancer Prediction in TCM Diagnosis”. Teoksessa *2020 IEEE International Conference on E-health Networking, Application Services (HEALTHCOM)*, 1–6. <https://doi.org/10.1109/HEALTHCOM49281.2021.9615914>.

Cheng, Chi-Yung, I-Min Chiu, Wun-Huei Zeng, Chih-Min Tsai ja Chun-Hung Richard Lin. 2021. “Machine Learning Models for Survival and Neurological Outcome Prediction of Out-of-Hospital Cardiac Arrest Patients”. *BioMed Research International* 2021 (syyskuu): 9590131. ISSN: 2314-6133. <https://doi.org/10.1155/2021/9590131>. <https://doi.org/10.1155/2021/9590131>.

Chi, Chien-Yu, Shuang Ao, Adrian Winkler, Kuan-Chun Fu, Jie Xu, Yi-Lwun Ho, Chien-Hua Huang ja Rohollah Soltani. 2021. “Predicting the Mortality and Readmission of In-Hospital Cardiac Arrest Patients With Electronic Health Records: A Machine Learning Approach”. *J Med Internet Res* 23, numero 9 (syyskuu). ISSN: 1438-8871. <https://doi.org/https://doi.org/10.2196/27798>. <http://www.ncbi.nlm.nih.gov/pubmed/34515639>.

Choi, Dong-Ju, Jin Joo Park, Taqdir Ali ja Sungyoung Lee. 2020. “Artificial intelligence for the diagnosis of heart failure”. *NPJ digital medicine* 3 (1): 1–6.

Choi, Edward, Andy Schuetz, Walter F Stewart ja Jimeng Sun. 2016. “Using recurrent neural network models for early detection of heart failure onset”. *Journal of the American Medical Informatics Association* 24, numero 2 (elokuu): 361–370. ISSN: 1067-5027. <https://doi.org/10.1093/jamia/ocw112>. eprint: <https://academic.oup.com/jamia/article-pdf/24/2/361/34148331/ocw112.pdf>. <https://doi.org/10.1093/jamia/ocw112>.

Deaton, Christi, Erika Sivaraja Froelicher, Lai Ha Wu, Camille Ho, Kawkab Shishani ja Tiny Jaarsma. 2011. “The Global Burden of Cardiovascular Disease”. *European Journal of Cardiovascular Nursing* 10, numero 2\_suppl (kesäkuu): S5–S13. ISSN: 1474-5151. [https://doi.org/10.1016/S1474-5151\(11\)00111-3](https://doi.org/10.1016/S1474-5151(11)00111-3). [https://doi.org/10.1016/S1474-5151\(11\)00111-3](https://doi.org/10.1016/S1474-5151(11)00111-3).

Deo, Rahul C. 2015. “Machine Learning in Medicine”. *Circulation* 132 (20): 1920–1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>. eprint: <https://www.ahajournals.org/doi/pdf/10.1161/CIRCULATIONAHA.115.001593>. <https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.115.001593>.



Duan, Yanqing, John S. Edwards ja Yogesh K Dwivedi. 2019. “Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda”. *International Journal of Information Management* 48:63–71. ISSN: 0268-4012. <https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2019.01.021>. <https://www.sciencedirect.com/science/article/pii/S0268401219300581>.

Farah, Yasmin, Syed Muhammad Ismail Shah, Aisha Naeem, Syed Muhammad Shujaiddin, Adina Jabeen, Sana Kazmi, Sarush Ahmed Siddiqui ym. 2021. “Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future”. *RCM* 22 (4): 1095–1113. <https://doi.org/10.31083/j.rcm2204121>.

Hamet, Pavel, ja Johanne Tremblay. 2017. “Artificial intelligence in medicine”. Insights Into the Future of Medicine: Technologies, Concepts, and Integration, *Metabolism* 69:S36–S40. ISSN: 0026-0495. <https://doi.org/https://doi.org/10.1016/j.metabol.2017.01.011>. <https://www.sciencedirect.com/science/article/pii/S002604951730015X>.

Harford, Samuel, Marina Del Rios, Sara Heinert, Joseph Weber, Eddie Markul, Katie Tataris, Teri Campbell, Terry Vanden Hoek ja Houshang Darabi. 2022. “A machine learning approach for modeling decisions in the out of hospital cardiac arrest care workflow”. *BMC Medical Informatics and Decision Making* 22, numero 1 (tammikuu): 21. ISSN: 1472-6947. <https://doi.org/10.1186/s12911-021-01730-4>. <https://doi.org/10.1186/s12911-021-01730-4>.

IBM. 2020. “What is machine learning?” Viitattu 15. kesäkuuta 2020. <https://www.ibm.com/cloud/learn/machine-learning>.

———. 2021. “Overfitting”. Viitattu 3. maaliskuuta 2021. <https://www.ibm.com/cloud/learn/overfitting>.

Jayasudha, B S K, P N Sudha, Rachana S, Sahana. Anagha A Kashyap ja Anusha L. 2021. “Machine Learning Techniques for Cardiovascular Risk Score -Prediction”. Teoksessa *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, 505–509. <https://doi.org/10.1109/MysuruCon52639.2021.9641581>.

Johnson, Kipp W., Jessica Torres Soto, Benjamin S. Glicksberg, Khader Shameer, Riccardo Miotto, Mohsin Ali, Euan Ashley ja Joel T. Dudley. 2018. “Artificial Intelligence in Cardiology”. *Journal of the American College of Cardiology* 71 (23): 2668–2679. <https://doi.org/10.1016/j.jacc.2018.03.521>. eprint: <https://www.jacc.org/doi/pdf/10.1016/j.jacc.2018.03.521>. <https://www.jacc.org/doi/abs/10.1016/j.jacc.2018.03.521>.

Li, Pengpai, Yongmei Hu ja Zhi-Ping Liu. 2021. “Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods”. *Biomedical Signal Processing and Control* 66:102474. ISSN: 1746-8094. <https://doi.org/https://doi.org/10.1016/j.bspc.2021.102474>. <https://www.sciencedirect.com/science/article/pii/S1746809421000719>.

Liu, Xinwen, Huan Wang, Zongjin Li ja Lang Qin. 2021. “Deep learning in ECG diagnosis: A review”. *Knowledge-Based Systems* 227:107187. ISSN: 0950-7051. <https://doi.org/https://doi.org/10.1016/j.knosys.2021.107187>. <https://www.sciencedirect.com/science/article/pii/S0950705121004494>.

Mayampurath, Anoop, Raffi Hagopian, Laura Venable, Kyle Carey, Dana Edelson, Matthew Churpek, American Heart Association’s Get With the Guidelines-Resuscitation Investigators ym. 2022. “Comparison of machine learning methods for predicting outcomes after in-hospital cardiac arrest”. *Critical Care Medicine* 50. <https://doi.org/https://doi.org/10.1097/CCM.0000000000005286>.

Miotto, Riccardo, Fei Wang, Shuang Wang, Xiaoqian Jiang ja Joel T Dudley. 2017. “Deep learning for healthcare: review, opportunities and challenges”. *Briefings in Bioinformatics* 19, numero 6 (toukokuu): 1236–1246. ISSN: 1477-4054. <https://doi.org/10.1093/bib/bbx044>. eprint: <https://academic.oup.com/bib/article-pdf/19/6/1236/27119191/bbx044.pdf>. <https://doi.org/10.1093/bib/bbx044>.

Nakamura, Tomofumi, ja Tetsuo Sasano. 2022. “Artificial intelligence and cardiology: Current status and perspective”. *Journal of Cardiology* 79 (3): 326–333. ISSN: 0914-5087. <https://doi.org/https://doi.org/10.1016/j.jjcc.2021.11.017>. <https://www.sciencedirect.com/science/article/pii/S0914508721003361>.

Oswald, Gadi Jaya Sathwika ja Arnab Bhattacharya. 2022. "Prediction of CardioVascular Disease (CVD) Using Ensemble Learning Algorithms". Teoksessa *5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD)*, 292–293. CODS-COMAD 2022. Bangalore, India: Association for Computing Machinery. ISBN: 9781450385824. <https://doi.org/10.1145/3493700.3493747>. <https://doi.org/10.1145/3493700.3493747>.

Silverio, Angelo, Pierpaolo Cavallo, Roberta De Rosa ja Gennaro Galasso. 2019. "Big Health Data and Cardiovascular Diseases: A Challenge for Research, an Opportunity for Clinical Care". *Frontiers in Medicine* 6 (helmikuu): 36. <https://doi.org/10.3389/fmed.2019.00036>.

Sitar-Taut, Adela Viviana, D Zdrengea, D Pop ja Dan-Andrei Sitar-Taut. 2009. "Using Machine Learning Algorithms in Cardiovascular Disease Risk Evaluation". *Journal of Applied Computer Science & Mathematics* 3 (tammikuu).

THL. 2020. "Sydän- ja verisuonitaudit". Viitattu 2. kesäkuuta 2020. <https://thl.fi/fi/web/kansantaudit/sydan-ja-verisuonitaudit>.

Tiainen, Marjaana, Johanna Hästbacka, Olli Takkunen ja Risto O Roine. 2006. "Viilennys-hoito parantaa kammiovärinästä elvytetyn potilaan ennustetta". *Duodecim* 122 (3): 295–304. ISSN: 0012-7183.

WHO. 2021. "Cardiovascular diseases (CVDs)". Viitattu 11. kesäkuuta 2021. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).

Yusuf, S, S Hawken, S Ounpuu, T Dans, A Avezum, F Lanas, M McQueen ym. 2004. "Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study". *Lancet* 364, numero 9438 (syyskuu): 937–952. [https://doi.org/https://doi.org/10.1016/s0140-6736\(04\)17018-9](https://doi.org/https://doi.org/10.1016/s0140-6736(04)17018-9). <https://pubmed.ncbi.nlm.nih.gov/15364185/>.