

Tiivistelmä

Nyky-yhteiskunnassa kerätään ja tallennetaan tietoa ennennäkemättömin määrin. Tämä tarjoaa merkittäviä mahdollisuuksia koneoppimiselle (Machine Learning) ja tekoälylle (Artificial Intelligence). Koneoppimisen sovellukset ulottuvat lähes kaikialle, terveydenhuolto ja lääketiede mukaan lukien. Neuroverkkojen ja syväoppimismallien kehittyminen on mahdollistanut monimutkaisten ja laajojen tietomassojen analysoinnin. Lääkärit hyödyntävät usein diagnosoinnin tukena useita tietolähteitä kuten kuvia, mittaustuloksia, ja oireita. Diagnostiikkaan tai taudin luokan tunnistamiseen kuuluu siten usein monien eri lähteiden datan arvioimista yhtenäisenä kokonaisuutena. Radiologiassa syväoppimista on käytetty esimerkiksi kuvien segmentointiin ja poikkeavuuksien tunnistamiseen (Lee et al., 2017). Lupaavista tuloksista huolimatta tekoälyyn pohjautuvien järjestelmien käyttö terveydenhuollossa on vielä olematonta.

Tämän tutkielman tarkoituksena on selvittää ja tutkia miten useita modaliteettia hyödynnyttään diagnosoinnin tukena, mihin ne pohjautuvat sekä mitä ongelmia niiden käyttöön liittyy. Erityisesti tutkielma keskittyy useiden modaliteettien käyttöön liittyvään datafuusioon, eli useiden modaliteettien yhteen sovittamiseen koneoppimis arkkitehtuureissa. Modaliteetilla tarkoitetaan tietotyyppiä kuten kuvaa tai ääntä. Tämän lisäksi tutkielma käsittelee neuroverkkoja ja syväoppimista perusteiden kautta rakentaen pohjan kompleksisempien kokonaisuuksien ja datafuusion käsittelyyn.

Ohjattu oppiminen (Supervised learning) on yksi koneoppimisen osa-alue, jossa olemassa olevan datan avulla pyritään muodostamaan malli, joka pystyy esimerkiksi luokittelemaan sille syötettävää dataa. Ohjattuun oppimiseen tarvitaan siis dataa joka on jo luokiteltua, tätä kutsutaan koulutus dataksi. Lineaarinen regressio on hyvä selkeä esimerkki joka soveltuu yksinkertaisten ongelmien ratkaisemiseen. Datan ulottuvuuksien ja kompleksisuuden lisääntyessä edistyneemmät algoritmit, kuten neuroverkot, toimivat paremmin.

Neuroverkon rakenne koostuu yksinkertaisimmillaan solmuista, jotka sijaitsevat verkon eri kerroksissa. Kokonaan yhdistetyssä neuroverkossa jokainen solmu yhdistyy jokaiseen seuraavaan kerroksen solmuun. Solmujen väliselle yhteydelle sijoittuu paino sekä jokaiseen solmuun harha. Neuroverkon kouluttamisella tarkoitetaan näiden painojen ja harhojen säätelyä niin, että verkko tuottaa halutun tuloksen. Neuroverkon ensimmäinen kerros jota usein kutsutaan syötekerrokseksi (Input layer) ottaa esimerkiksi kuvan pikselien ar-

voja solmujen arvoiksi. Kuva, joka koostuu 20x20 pikselistä, tuottaa täten 400 solmun kokoisen syötekerroksen josta jokainen solmu yhdistyy jokaiseen seuraavan kerroksen solmuun. Neuroverkon viimeistä kerrosta kutsutaan tulostekerrokseksi (Output layer) ja sen solmujen määrä määräytyy esimerkiksi luokittelutehtävissä luokiteltavien kohteiden mukaan. Ensimmäisen ja viimeisen kerroksen koko on siis valmiiksi määritetty. Kerroksia jotka sijaitsevat syöte ja tuloste kerroksen välissä kutsutaan piilokerroksiksi (Hidden layer). Piilokerrosten määrää ja kokoa ei ole valmiiksi määritetty, syväoppimisesta puhuttaessa viitataan verkkoihin joissa piilokerroksia on useita.

Neuroverkon kouluttamisella tarkoitetaan prosessia jossa verkon harhat ja painot optimoidaan. Koulutukseen käytettävä data voidaan jakaa testi- ja koulutusdataan jolloin verkon suoriutumista voidaan arvioida datalla, jota ei ole käytetty sen kouluttamiseen. Eteenpäinsyöttö algoritmin avulla verkon syöte viedään kerroksien läpi aina tuloste kerrokselle asti. Yksittäisen solmun aktivaatio voidaan laskea siihen kytkettyjen solmujen painolla skaalattujen arvojen summana, johon lisätään solmun harha. Tätä summaa voidaan kutsua solmun lineaariseksi muunnokseksi, jotta aktivaatio saadaan laskettua tähän sovitetaan aktivaatiofunktio. Aktivaatiofunktion avulla lineaarisuus saadaan rikottua. Tämä mahdollistaa epälineaaristen suhteiden mallintamisen. Samalla kaavalla voidaan edetä tuloskerrokselle asti. Tulostekerroksen aktivaatiofunktio poikkeaa usein muiden kerroksien aktivaatiofunktioista, koska tämän kerroksen arvot halutaan kuvata todennäköisyyksinä ja täten niiden summan on oltava 1. Viimeisen kerroksen solmujen ja syötteen tunnisteiden (label) erotusta kutsutaan kustannusfunktioksi. Takaisinvirtausalgoritmi hyödyntää tätä ja sen tehtävä on laskea jokaisen painon ja harhan osittaisderivaatta suhteessa kustannusfunktioon. Näitä arvoja voidaan kutsua myös gradientteiksi. Gradientti suhteessa kustannusfunktioon kertoo suunnan, joka kasvattaa sen arvoa nopeimmin. Siirtämällä painoja ja harhoja kohti negatiivista gradienttiä voidaan kustannusfunktion arvo minimoida. Neuroverkkoja on kehitetty moneen tarkoitukseen. Useat näistä kuten konvoluutio ja takisinkytöntäneuroverkot perustuvat samojen algoritmien ja koulutustekniikoiden käyttöön kuin kokonaan kytketyssä eteenpäinsyöttö neuroverkossa.

Multimodaalinen koneoppiminen (Multimodal machine learning) pyrkii hyödyntämään useita modaliteetteja saman aikaisesti. Multimodaalista koneoppimista voidaan hyödyntää monissa eri käyttötarkoituksissa kuten generatiivisessä tekoälyssä. Lääketieteellisessä diagnosoinnissa multimodaalisia koneoppimista voidaan hyödyntää esimerkiksi tunnistamaan tauti. Esimerkiksi röntgenkuvien avulla voidaan kouluttaa malli tunnistamaan nivelrikko (Tiulpin et al., 2018). Malleja, jotka pohjautuvat vain yhden modaliteetin kuten

kuvan käyttöön, kutsutaan unimodaalisiksi (Unimodal) malleiksi. Röntgenkuvien lisäksi mallille voidaan syöttää myös muita modaliteetteja, jotka ovat relevantteja ongelman kannalta. Esimerkiksi lääketieteellisten kuvien lisäksi voidaan käyttää potilaskertomuksia, mittaus- tai testi tuloksia, tai geneettisiä tietoja. Datafuusio käsittelee useiden modaliteettien yhdistelemistä (Baltrušaitis et al., 2019). Datafuusiomenetelmät voidaan karkeasti jakaa kolmeen luokkaan: aikaiseen, myöhäiseen ja keskivaiheen fuusioon. Mallille syötettävät modaliteetit ovat lähtökohtaisesti erillisiä ja rakenteellisesti toisistaan poikkeavia. Fuusiomenetelmien etuliite kuvaa arkkitehtuurissa fuusion vaiheen ajankohtaa suhteessa koneoppimisalgoritmiin tai -malliin. Aikaisella fuusiolla tarkoitetaan täten siis menetelmää, jossa modaliteetit yhdistetään yhtenäiseksi ennen mallille syöttämistä. Myöhäinen fuusio taas on menetelmä, jossa jokainen modaliteetti käsitellään erillisenä ja jokainen syötetään omaan malliin. Näiden tuloksia yhdistetään ennen lopullista päätöksentekoa. Keskivaiheen fuusiolla tarkoitetaan yhdistämistä, joka tapahtuu mallin sisällä. Kaikista näistä myös löytyy variaatioita, joissa esimerkiksi jokin modaliteetti käsitellään eri tavalla.

Fuusiomenetelmän toimivuuteen vaikuttaa vahvasti ongelma, jota pyritään ratkaisemaan. Lisäksi siihen vaikuttaa datan määrä ja laatu. Aikaisen fuusion menetelmässä voidaan esimerkiksi hyödyntää valmiiksi koulutettuja malleja tai autoenkoodereita (Auto encoder) modaliteettien kuvaamiseksi yhtenäiseksi vektoriksi. Myöhäisen fuusion kohdalla taas voidaan hyödyntää unimodaaleja malleja, joiden koulutus ja käyttö eivät ole sidoksissa muihin modaliteetteihin. Myöhäisen fuusiota hyödyntävissä kokonaisuuksissa voidaan myös puuttuviin modaliteetteihin reagoida, koska yksikään osa ei ole suoraan riippuvainen muista modaliteeteista. Keskivaiheen fuusio hyödyntää neuroverkkoja ja poiketen muista menetelmistä, se voidaan kouluttaa yhtenäisenä mallina ja takaisinvirtausalgoritmiä hyödyntäen aina tuloksesta alkuperäisiin syötteisiin asti. Neuroverkkojen hyödyntäminen vaatii enemmän dataa kuin muut menetelmät ja tekee siten mallin kouluttamisesta vaikeampaa. Tieteellisessä yhteisössä ei ole konsensusta siitä, mikä menetelmästä on tehokkain. Multimodaaliseen koneoppimiseen liittyvät keskeiset haasteet voidaan jakaa viiteen luokkaan (Baltrušaitis et al., 2019):

1. *Esittäminen (Representation)* miten useat modaliteetit esitetään merkityksellisesti?
2. *Kääntäminen (Translation)* miten data määritetään modaliteetista (A) \rightarrow (B)?
3. *Kohdistaminen (Alignment)* mitkä osat yhdestä modaliteetista vastaavat suoraan toista?
4. *Fuusio (Fusion)* miten modaliteetit yhdistetään?
5. *Yhteisoppiminen (Co-learning)* miten tietoa voidaan jakaa eri modaliteettien välillä?

Multimodaaliseen koneoppimiseen pohjautuvien mallien tehokkuuden arviointia vaikeuttaa ratkaisujen vahva riippuvuus käsiteltävästä ongelmasta ja saatavilla olevasta datasta. Monet tutkimukset käyttävät valmiiksi kerättyä dataa, joka ei ole julkisesti saatavilla. Multimodaalista koneoppimista täsmälääketieteessä tarkasteleva kirjallisuuskatsaus (Kline et al., 2022), osoitti usean modaliteetin nostavan mallien tarkkuutta keskimäärin 6.4% verrattaen yhden modaliteetin malleihin. Katsaukseen päätyneissä artikkeleissa suurimpina ongelmina nousivat pienet otoskoot, luokkien erisuuruus ja retrospektiivinen data. Lisäksi useat tutkimukset käyttivät vain yhden sairaanhoitopiirin dataa. Kliinisen validoinnin sekä mallien yleistyvyyden arviointiin ei siis tämän katsauksen perusteella pystytä. Huomioitavaa on kuitenkin multimodaalisten mallien tarkkuuden paraneminen verrattaen vain yhden modaliteetin hyödyntämiseen. Katsaukseen valituista papereissa lääketieteellisistä alueista neurologia ja syöpä olivat eniten edustettuina. Sähköisten terveystiedot ja kuvantamisdata olivat eniten edustettu yhdistelmä modaliteetteja.

Tekoälyn soveltaminen lääketieteellisessä kliinisessä kontekstissa on laaja ongelma. Multimodaalisen datan käyttöön pohjautuvat mallit ovat antaneet lupaavaa näyttöä mahdollisesta tarkkuuden paranemisesta verrattaessa yhden modaliteetin malleihin. Koneoppimiseen pohjautuvien järjestelmien läpinäkyvyydellä ja päätösten selitettävyydellä on merkittävä rooli, kun ajatellaan potentiaalista kliinistä käyttöä.