



Kirjoitelma

Tietojenkäsittelytieteen kandiohjelma

# Koneoppimisen menetelmät ja käyttökohteet lääketutkimuksessa -ja kehityksessä

Heikki Pulli

22.9.2021

MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA  
HELSINGIN YLIOPISTO

**Ohjaaja(t)**

**Tarkastaja(t)**

## **Yhteystiedot**

PL 68 (Pietari Kalmin katu 5)  
00014 Helsingin yliopisto

Sähköpostiosoite: [info@cs.helsinki.fi](mailto:info@cs.helsinki.fi)

URL: <http://www.cs.helsinki.fi/>

# Sisällys

1	Johdanto	1
2	Lääketutkimuksessa -ja kehityksessä käytetyt koneoppimisen mallit	3
3	Koneoppimismallien käyttökohteet	4
4	Koneoppimismallien heikkoudet ja puutteet	5
	Lähteet	7



# 1 Johdanto

käyttökohteet 1. kohteen varmentaminen 2. Biomerkin tunnistaminen 3. kliinisistä testeistä kerätyn patologisen datan analyysia

hidasteet/puutteet 1. tulosten tulkitsemisen vajaavuus 2. tulosten toistettavuus

tärkeitä nosteita 1. Korkea laatuisten datan tarve 2. Koneoppimismallien validointi

koenoppimisesta 1. Käytetään supervised ja unsupervised malleja supervised, kun halutaan luokitella näytteitä ja ennustaa tulevia näytteitä unsupervised, kun etsitään ryppäitä datasta 2. Käytettävä malli tulee valittava käyttökohteen mukaan Eri malleja vertaillaan usean eri metriikan perusteella, esim. ryhmittely tarkkuus, kappa -arvo, logaritminen ero...

3. CPU -> GPU -> TPU

Neuroverkoista Eri neurtoverkkoja ja niiden toiminnalliset erot 1. Convolutional neural networks (CNN) 2. Recurrent neural network (RNN) 3. feedforward network 4. Deep autoencoder network (DAEN)

Datasta ja datan laadusta ja määrästä 1. Tulee olla tarkistettua, tarkkaa ja tarpeeksi kuvaavaa 2. Mallia suunniteltaessa mietittävä, mikä on tarpeeksi dataa 3. Useat käytännön koneet eivät suoriudukkaan tehtävästään hyvin, koska niiden koulutukseen tai niiden käsittelemä data vaihtelee laadultaan

Käytännön kohteet lääkkeiden löytämisessä/kehityksessä 1. Koneoppimista on jo käytetty useassa tutkimuksessa tunnistamaan ja luokittelemaan dataa näytteistä 2. NLP -mementelmiä hyödynnetään aineiston etsinnässä. Aineistoa voidaan etsiä lääke-sairaus, geeni-sairaus ja kohde-lääke konteksteissa 3. Voidaan käyttää syöpäkohtaisten lääkkeiden etsinnässä. Koulutetaan malli hyödyntäen kerättyä dataa ja ennustetaan eri lääkkeiden vaikutusta 4. Yksi suurimmista tavoitteista koneoppimisen hyödyntämisessä on ennustaa, mitkä lääkkeet tulevat menestymään kliinisissä kokeissa 5. Koneoppimista on myös hyödynnetty pienten kemiallisten yhdisteiden suunnittelussa. - Multitask -mallit ovat osoittautuneet tehokkaammiksi kuin singletask -mallit, mutta ne ovat hyvin datan laadusta riippuvaisia - Neuroverkkoja ja nykyisiä hakupuualgoritmeja voidaan käyttää kemiallisen aineen synteesin välivaiheiden suunnittelussa - De novo yhdisteiden suunnittelussa - Suuri kysymys on, kuinka pieniä molekyyliä/yhdisteitä kuvataan, esim. circular fingerprint, symmetry functions

Koneoppimisen malleja on alettu käyttämään lääketutkimuksen ja -kehityksen tuotantoprosessin eri osa-alueilla ja vaiheissa. Koneoppimisen malleja voidaan hyödyntää esim. eri solujen tunnistamiseen kuvista, kemiallisten yhdisteiden vaikutusten ennustamista eri soluihin ja parhaimmillaan jopa mallentamaan tiettyyn sairauteen tehoavaa täsmälääkettä [1]. On myös kuitenkin herännyt huoli datasta, jolla koulutetaan koneoppimisen mallit eri tilanteisiin. Huoli on noussut koskien datan laatua, jolla koneet koulutetaan. Tämä myös koneoppimisen mallien käyttöönottoa, koska koskaan ei täysin tiedetä, miten koulutettu kone on tulokseensa päätenyt.

## **2 Lääketutkimuksessa -ja kehityksessä käytetyt koneoppimisen mallit**

### 3 Koneoppimismallien käyttökohteet



## 4 Koneoppimismallien heikkoudet ja puutteet



# Lähteet

- [1] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer ja S. Zhao. "Applications of machine learning in drug discovery and development". eng. *Nature reviews. Drug discovery* 18.6 (2019), s. 463–477. ISSN: 1474-1776.

