



Universiteit
Leiden
The Netherlands

From code to clinic: theory and practice for artificial intelligence prediction algorithms

Hond, A.A.H. de

Citation

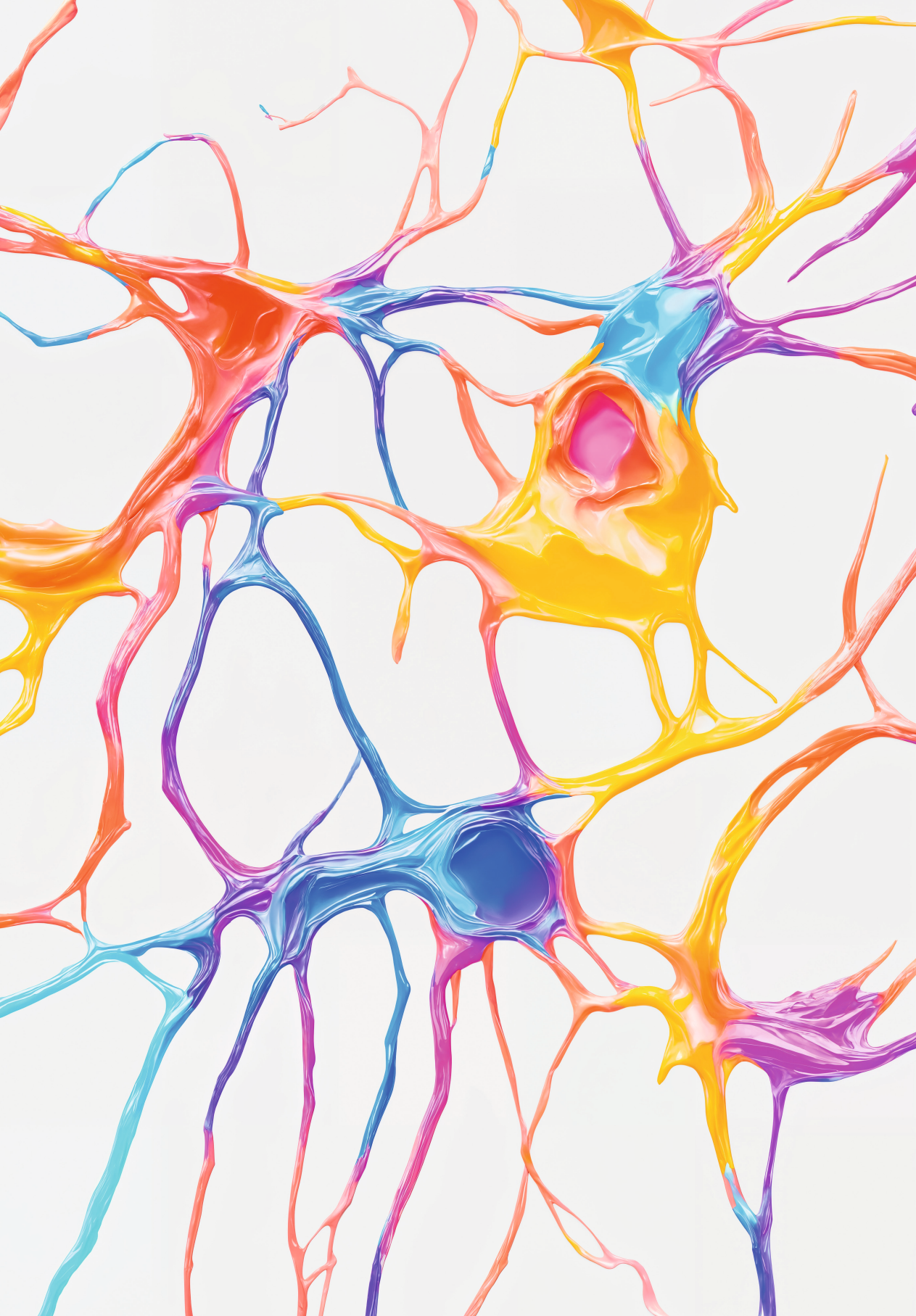
Hond, A. A. H. de. (2023, October 11). *From code to clinic: theory and practice for artificial intelligence prediction algorithms*. Retrieved from <https://hdl.handle.net/1887/3643729>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3643729>

Note: To cite this publication please use the final published version (if applicable).



11

Samenvatting geschreven samen met GPT-3 [1]

Dit proefschrift onderzoekt de potentie van kunstmatige intelligentie (*Artificial Intelligence* of AI in het Engels) voor de gezondheidszorg. Het eerste deel van dit proefschrift richt zich op het verantwoord ontwikkelen en valideren van klinische voorspelalgoritmen op basis van AI, waarbij de belangrijkste overwegingen in kaart worden gebracht. Het tweede deel van dit proefschrift onderzoekt de sterke en zwakke punten van klassieke statistiek en machine learning-technieken voor het ontwikkelen van voorspelalgoritmen. Ook beschrijft dit deel het voorspellend vermogen, de potentie, en de uitdagingen van AI-voorspelalgoritmen voor de klinische praktijk.

Hoofdstuk 2 geeft een overzicht van richtlijnen en kwaliteitscriteria voor de ontwikkeling, evaluatie, en implementatiecyclus van AI-voorspelalgoritmen. Dit hoofdstuk behandelt verschillende onderwerpen, zoals datakwaliteit, externe validatie, en effectiviteitsstudies. Dit hoofdstuk concludeert dat er meer onderzoek nodig is naar de laatste drie fasen van de AI-ontwikkelcyclus (ontwikkeling van de benodigde software, effectbeoordeling van de AI, en implementatie en gebruik van de AI in de dagelijkse praktijk).

Hoofdstuk 3 bediscussieert de belangrijkste overwegingen bij het kwantificeren en valideren van het voorspellend vermogen van algoritmen. Dit hoofdstuk gaat eerst in op de verschillen in discriminatiematen van de statistiek en machine learning disciplines. We roepen op tot één gemeenschappelijke aanpak om samenwerking tussen verschillende disciplines te bevorderen. Daarnaast toont dit hoofdstuk aan dat labels (bijv. ‘goed’ of ‘slecht’) voor specifieke *Area Under the Receiver Operating Characteristic Curve*-waarden grotendeels arbitrair zijn. Andere maten, zoals Net Benefit, moeten worden gerapporteerd om de waarde van een algoritme voor besluitvorming te evalueren. Ten slotte geeft dit hoofdstuk een overzicht van vier soorten generaliseerbaarheid voor voorspelalgoritmen en sporen we aan om algoritmen te valideren in overeenstemming met het beoogd gebruik van de algoritmen.

In het tweede deel van dit proefschrift worden verschillende klinische use cases besproken. **Hoofdstuk 4** onderzoekt of machine learning-algoritmen de ziekenhuisopname van patiënten op de spoedeisende hulp nauwkeuriger kan voorspellen dan logistische regressie. De resultaten tonen aan dat machine learning-algoritmen een uitstekend maar vergelijkbaar voorspellend vermogen

hebben als de logistische regressie. Deze resultaten impliceren dat machine learning weinig voordeel oplevert voor het voorspellen van ziekenhuisopname in vergelijking met conventionele algoritmen.

Hoofdstuk 5 bestudeert het voorspellend vermogen van algoritmen die ernstige astma-exacerbaties voorspellen bij patiënten met milde tot matige chronische astma op basis van thuismonitoringsgegevens. We vergelijken twee machine learning-algoritmen, een statistisch algoritme, en een eenvoudige klinische regel. De resultaten laten zien dat het logistische regressie algoritme een beter voorspellend vermogen heeft dan het XGBoost-algoritme en de eenvoudige klinische regel. Het produceert echter nog steeds veel fout positieven bij hoge sensitiviteit. De studie concludeert dat machine learning-algoritmen mogelijk niet beter presteren dan klassieke regressie algoritmen bij het voorspellen van astma-exacerbaties op basis van thuismonitoringsgegevens.

Hoofdstuk 6 vergelijkt machine learning en regressie algoritmen met een *competing risk* in 11 datasets van patiënten die een artroplastische operatie ondergingen. De resultaten tonen aan dat de machine learning- en regressie-algoritmen vergelijkbaar presteren. De bevindingen benadrukken de beperkingen van machine learning-technieken en de noodzaak voor relevante voorspellers voor dit voorspelprobleem.

Hoofdstuk 7 beschrijft de ontwikkeling van een algoritme voor de identificatie van patiënten met kanker die risico lopen op een depressie bij de start van een chemo- of radiotherapiebehandeling. Het algoritme dat het beste presteerde (LASSO logistische regressie met gestructureerde data) had een redelijke discriminatie en kalibratie. Gestructureerde data blijkt belangrijker te zijn voor het voorspellen van het risico op depressie dan ongestructureerde tekst in deze specifieke klinische workflow. De resultaten impliceren dat dit algoritme potentie heeft als een vroege triage-tool, maar dit moet verder gevalideerd worden.

Hoofdstuk 8 beschrijft een externe validatie studie voor een gecertificeerd machine learning-algoritme dat heropname of overlijden voorspelt binnen 7 dagen na ontslag van de IC. Uit de resultaten blijkt dat de externe validiteit niet

als vanzelfsprekend beschouwd kan worden en dat updating nodig is om de prestaties van het algoritme op de nieuwe locatie te verbeteren.

De conclusie stelt dat discipline overstijgende samenwerking, uitwisseling van kennis en resultaten, en validatie van AI voor de zorgpraktijk essentieel zijn om het potentieel van AI voor de zorg te verwezenlijken.

REFERENTIE

1. Ouyang, L., J. Wu, X. Jiang, et al., *Training language models to follow instructions with human feedback*. arXiv preprint arXiv:2203.02155, 2022.