

Artificial intelligence

Citation for published version (APA):

Groot Lipman, K. B. W. (2024). *Artificial intelligence: the key to standardizing respiratory disease evaluation*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20240122kg>

Document status and date:

Published: 01/01/2024

DOI:

[10.26481/dis.20240122kg](https://doi.org/10.26481/dis.20240122kg)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

This PhD thesis investigates the application of artificial intelligence (AI) in quantifying disease status and response to therapy for patients with asbestosis, pleural plaques, COVID-19, and pleural mesothelioma. The thesis is divided into two parts: Enhancing Disease Quantification at Baseline and Evaluating Therapeutic Response in Pleural Mesothelioma.

In Part I, our first study in *Chapter 2* explores using AI to assist in deciding which asbestosis patients (a lung disease caused by asbestos exposure) should receive government support. We analyzed 523 cases in the Netherlands, using AI to review chest CT scans and lung function tests. The AI's decisions were compared with those of a panel of lung doctors. Results showed the AI system was quite accurate, even more so when combined with lung function test data. This research suggests AI could be a valuable tool in streamlining and improving the fairness of the support application process for asbestosis patients.

Chapter 3 tested the AI model developed in *Chapter 2* in a real-life setting without having an impact on the decision. We included all applicants seeking asbestosis compensation in a Dutch nationwide cohort from September 2020 to July 2022. The AI's assessments were compared with the evaluations of the three pulmonologists. If the AI was unsure, two more reviewers joined the assessment. The results showed that the AI was quite accurate, but it didn't hit our target for sensitivity – the ability to correctly identify those with asbestosis. The AI did well in terms of specificity – correctly identifying those without the disease – and overall accuracy. However, because it didn't meet our sensitivity goal, we believe more work is needed.

In this study of *Chapter 4*, we focused on pleural plaques (PP), which

are signs of long-term asbestos exposure, and their impact on lung function. We also aimed to speed up the process of measuring PP by using AI. We trained an AI model to identify PP in CT scans of patients who had been exposed to asbestos. This model was compared with the work of radiologists. We also looked at how the volume of PP related to different lung function tests. The AI was trained on 422 CT scans and tested its accuracy in predicting PP volume. The results showed a strong correlation between the AI's predictions and the actual measurements. We also found that larger PP volumes are associated with a decrease in some lung function measures, giving us new insights into the effects of asbestos exposure on lung health.

Chapter 5 evaluates an AI model for analyzing COVID-19 in CT scans. The model was trained to identify lung infections and estimate COVID-19 likelihood. To test the performance in real-world situations, we used a different set of 400 CT scans from various centers. The results showed that while the model was excellent at identifying lung contours both in our tests and the external ones, it struggled with detecting lung infections when used outside the initial testing environment. Additionally, its effectiveness in determining the likelihood of COVID-19 also dropped in these external tests. The takeaway from this study is that while the AI model showed promise, its performance varied significantly in different settings. This highlights the importance of testing AI models in various real-world conditions, especially for clinical tools, to ensure they are reliable and effective in all potential environments.

In Part II of this thesis, we developed in **Chapter 6** a new way to measure how well treatments work for pleural mesothelioma (PM), a type of cancer linked to asbestos exposure. We created an AI algorithm that can automatically measure the volume of PM tumors in CT scans. This AI tool was designed to make it easier and more accurate to track changes in tumor size over time, which is important for understanding how well treatments are working. The AI's performance was impressive in our tests. It matched the expert-segmented tumor volumes with 89% accuracy in an internal test set. When we used it in a large European dataset, the AI showed 98% overlap with expert corrections,

demonstrating its high reliability. In a side-by-side comparison of CT scans of a smaller phase II trial, radiologists often preferred the AI's segmentation over the manual volume extraction of a radiologist.

We also used the AI to see if the size of the tumor at the beginning of treatment could predict how patients would fare. We divided the tumors into four size groups and found that the initial tumor size was significantly linked to the patient's overall survival.

Furthermore, we introduced new criteria called ARTIMES to evaluate how PM tumors respond to treatment. These criteria use both a fixed size change and a percentage increase to determine if the cancer is getting worse. In our studies, ARTIMES spotted tumor growth about 7 weeks earlier than traditional methods. It also proved to be more effective in predicting patients' overall survival.

Samenvatting

Dit proefschrift onderzoekt de toepassing van kunstmatige intelligentie (AI) bij het kwantificeren van ziektestatus voor patiënten met asbestose, pleurale plaques, en COVID-19. Verder stellen we een nieuwe methode voor om respons op therapie bij pleuraal mesothelioom te bepalen. Het proefschrift is verdeeld in twee delen: verbetering van ziektekwantificering op de eerste scan en evaluatie van de respons op de behandeling bij pleuraal mesothelioom.

In deel I onderzoeken we in *Hoofdstuk 2* het gebruik van AI om te helpen bepalen welke patiënten asbestose hebben (een longziekte veroorzaakt door asbestblootstelling), en daardoor in aanmerking komen voor overheidssteun. We analyseerden 523 casussen in Nederland, waarbij AI werd gebruikt om CT-scans en longfunctietests te beoordelen. De beslissingen van de AI werden vergeleken met die van een panel van longartsen. De resultaten toonden aan dat het AI-systeem vrij nauwkeurig was, en zelfs verbeterde toen het gecombineerd werd met longfunctietestgegevens. Dit onderzoek suggereert dat AI een waardevol hulpmiddel kan zijn bij het stroomlijnen en verbeteren van de consistentie van het aanvraagproces voor overheidssteun aan asbestosepatiënten.

Hoofdstuk 3 testte het AI-model ontwikkeld in *Hoofdstuk 2* in de echte beoordelingopzet zonder invloed op de uiteindelijke beslissing. We hebben alle aanvragers voor asbestosecompensatie in een landelijke Nederlandse cohort van september 2020 tot juli 2022 opgenomen. De beoordelingen van de AI werden vergeleken met de evaluaties van de drie longartsen. Als de AI onzeker was, sloten twee extra beoordelaars zich aan bij de beoordeling. De resultaten toonden aan dat de AI vrij nauwkeurig was, maar ons doel voor gevoeligheid – het correct identificeren van mensen met asbestose – niet haalde. De AI deed het goed

qua specificiteit – het correct identificeren van mensen zonder de ziekte – en algehele nauwkeurigheid. Omdat het echter ons gevoeligheidsdoel niet bereikte, is er meer werk nodig.

In de studie van **Hoofdstuk 4** hebben we ons gericht op pleurale plaques (PP), die kunnen voorkomen na langdurige blootstelling aan asbest. Mensen met PP komen niet in aanmerking voor overheidssteun, omdat er niet duidelijk is aangetoond dat PP een negatieve impact op de kwaliteit van leven heeft. Om dit te onderzoeken, beoogden we het proces van het meten van PP versnellen door het gebruik van AI. We trainden een AI-model om PP te identificeren in CT-scans van patiënten die blootgesteld waren aan asbest. Dit model werd vergeleken met het werk van radiologen. We keken ook naar hoe het volume van PP gerelateerd was aan verschillende longfunctietests. De AI werd getraind op 422 CT-scans en testte de nauwkeurigheid bij het voorspellen van PP-volume. De resultaten toonden een sterke correlatie tussen de voorspellingen van de AI en de werkelijke metingen van de radiologen. We vonden ook dat grotere PP-volumes geassocieerd zijn met een afname van sommige longfunctiemetingen, waardoor we nieuwe inzichten kregen in de effecten van pleurale plaques op de longgezondheid.

Hoofdstuk 5 evalueert een AI-model voor het analyseren van COVID-19 in CT-scans. Het model werd getraind om longinfecties te identificeren en de waarschijnlijkheid van COVID-19 te schatten. Om de prestaties in een realistische situatie te testen, gebruikten we een andere set van 400 CT-scans uit verschillende centra. De resultaten toonden aan dat het model uitstekend was in het identificeren van longcontouren, maar het moeite had met het detecteren van longinfecties wanneer deze in een nieuw ziekenhuis werd gebruikt. Bovendien daalde de effectiviteit in het bepalen van de waarschijnlijkheid van COVID-19 ook. De conclusie van deze studie is dat hoewel het AI-model belovend was, de prestaties aanzienlijk varieerden in nieuwe omgevingen. Dit benadrukt het belang van het testen van AI-modellen in verschillende omstandigheden om ervoor te zorgen dat ze betrouwbaar en effectief zijn in meerdere ziekenhuizen.

In deel II van dit proefschrift hebben we in *Hoofdstuk 6* een nieuwe manier ontwikkeld om te meten hoe goed behandelingen werken voor pleuraal mesothelioom (PM), een soort kanker veroorzaakt door asbestblootstelling. We creëerden een AI-algoritme dat automatisch het volume van PM-tumoren in CT-scans kan meten. Dit AI-hulpmiddel is ontworpen om het gemakkelijker en nauwkeuriger te maken om veranderingen in tumorgrootte over tijd bij te houden, wat belangrijk is om te begrijpen hoe goed behandelingen werken. De prestaties van de AI waren indrukwekkend in onze test. Het kwam overeen met de door experts bepaalde tumorvolumes met 89% nauwkeurigheid in een onafhankelijke interne testset. Toen we de AI inzetten in een grote Europese dataset, toonde het 98% overlap na correcties van medische experts, wat een hoge betrouwbaarheid aantoont. In een vergelijking van CT-scans van een kleinere fase II-studie gaven radiologen vaak de voorkeur aan de volume-bepaling van de AI ten opzichte van de handmatige volume-bepaling van een radioloog.

We gebruikten ook de AI om te zien of de grootte van de tumor aan het begin van de behandeling kon voorspellen hoe het met patiënten zou gaan. We verdeelden de tumoren in vier groepen van tumorvolumes en ontdekten dat de aanvangsgrootte van de tumor significant gekoppeld was aan de algehele overleving van de patiënt.

Verder introduceerden we nieuwe criteria genaamd ARTIMES om te evalueren hoe PM-tumoren reageren op behandeling. Deze criteria gebruiken zowel een absolute verandering in volume als een percentage toename om te bepalen of de kanker erger wordt. In onze studies ontdekte ARTIMES tumorgroei ongeveer 7 weken eerder dan traditionele methoden. Het bleek ook effectiever te zijn in het voorspellen van de algehele overleving van patiënten.