Can DL models be trusted in the clinical practice? Our lessons learned.

SASG meeting – July 3 2023

Bart Elen

vito

REALM

vito.be

INTRO VITO

SUSTAINABLE CHEMISTRY

SUSTAINABLE LAND USE

SUSTAINABLE HEALTH

SUSTAINABLE ENERGY

SUSTAINABLE MATERIALS

vito REALM

vito.be

Our European projects on medical AI evaluation

Vivaldy - **V**er**I**fication and **V**alidation of **A**i-enabL**e**D s**Y**stems

Three year (2020–2023) PENTA project

https://vivaldy-penta.eu/

REALM - Real-world-data Enabled Assessment for heaLth regulatory decision-Making

Four years (2023-2027) Horizon Europe project

https://realm-ai.eu/

TEF-Health - Testing and Experimentation Facility for Health AI and Robotics

Five year (2023-2028) Digital Europe project

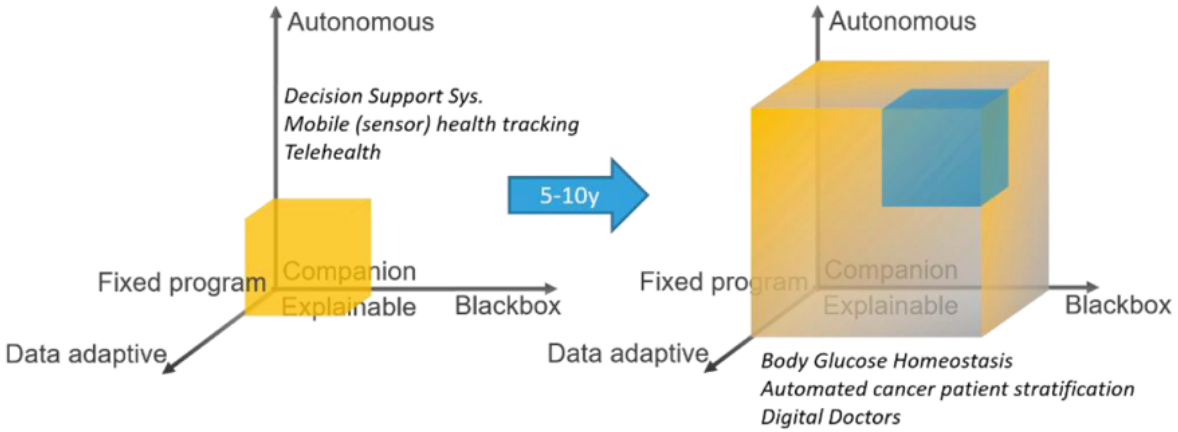
https://www.tefhealth.eu/

NoBoCap - Notified Body Increased Capacity

Three year (2023-2026) EU4Health project

Software-assisted healthcare is becoming a reality

The rise of AI combined with affordable data collection will bring autonomous, self-learning, and potentially complex (black-box) decision-making software solutions into the clinic.

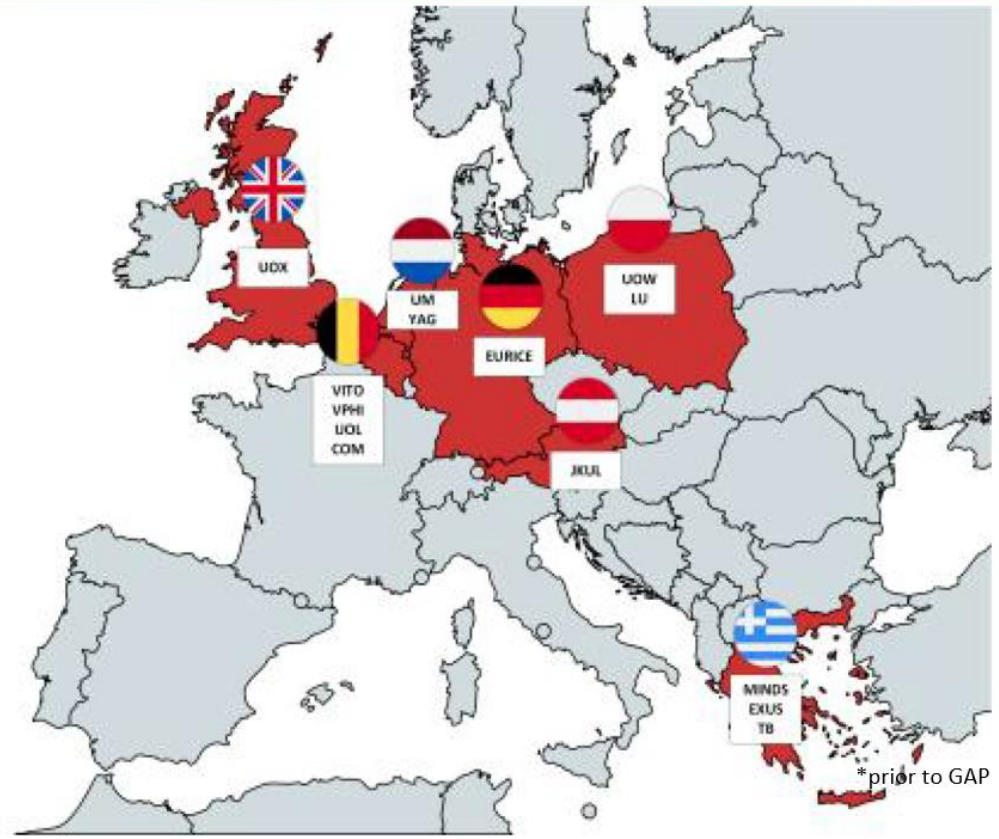


Their certification and post-market monitoring will require **independent datasets** and state-of-the-art ethical and societally grounded **algorithms in the regulatory toolbox** to ensure their robustness, trustworthiness, and explainability.

The European MDSW Market is anticipated to reach US$25 billion by 2027 from US$ 5,5 billion in 2019 with a compound annual growth rate (CAGR) of 21.6%.

Team

1. Maastricht University, NL (Coordinator)
2. VITO B.V., BE
3. Comunicare, BE
4. University of Liege, BE
5. Virtual Physiological Human Inst., BE
6. Metaminds Innovations PC, GR
7. Tragbeat, GR
8. Exus Software Monoprosopi Etairia Periorismenis, GR
9. University of Oxford, UK
10. University of Bristol, UK
11. EURICE European Research and Project Office GmbH, DE
12. University of Warsaw, PL
13. Lazarski University, PL
14. Yaghma, NL
15. University of Antwerp, BE



*prior to GAP

- Technology Development
- Ethical, Legal & Societal
- Dissemination, Project Management & Outreach
- Demonstrator

Questions we aim to address

- Which **real-world data (RWD)** and **real-world evidence (RWE**) should be **required** to permit marketing as Medical Device Software (MDSW)?
- How can we **ensure the safety and effectiveness of artificial intelligence/machine learning (AI/ML)-based MDSW** (or IVD-Software) that may drift over time as they are applied in real life and exposed to new data?
- Should software solutions be **reimbursed** by national healthcare services ?

MONA

Smart eye screening solutions

https://mona.health

02 Apr 2019 | 15:00 GMT

How IBM Watson Overpromised and Underdelivered on AI Health Care

After its triumph on *Jeopardy!*, IBM's AI seemed poised to revolutionize medicine. Doctors are still waiting

By **Eliza Strickland**

MIT Technology Review

ARTIFICIAL INTELLIGENCE

Google's medical AI was super accurate in a lab. Real life was a different story.

If AI is really going to make a difference to patients we need to know how it works when real humans get their hands on it, in real situations.

By Will Douglas Heaven April 27, 2020

But, after its publication, the authors of the study noticed a bias in their algorithm — it was more likely to label an image as malignant cancer if there was a ruler in the image. Dermatologists often use a ruler to measure the size of a skin lesion in the photo if they're particularly concerned about it, so a photo with a ruler in it is more likely to be cancerous.

There seems to be little guarantee that the good results found with the AI lab evaluations will always be obtained in the clinical practice

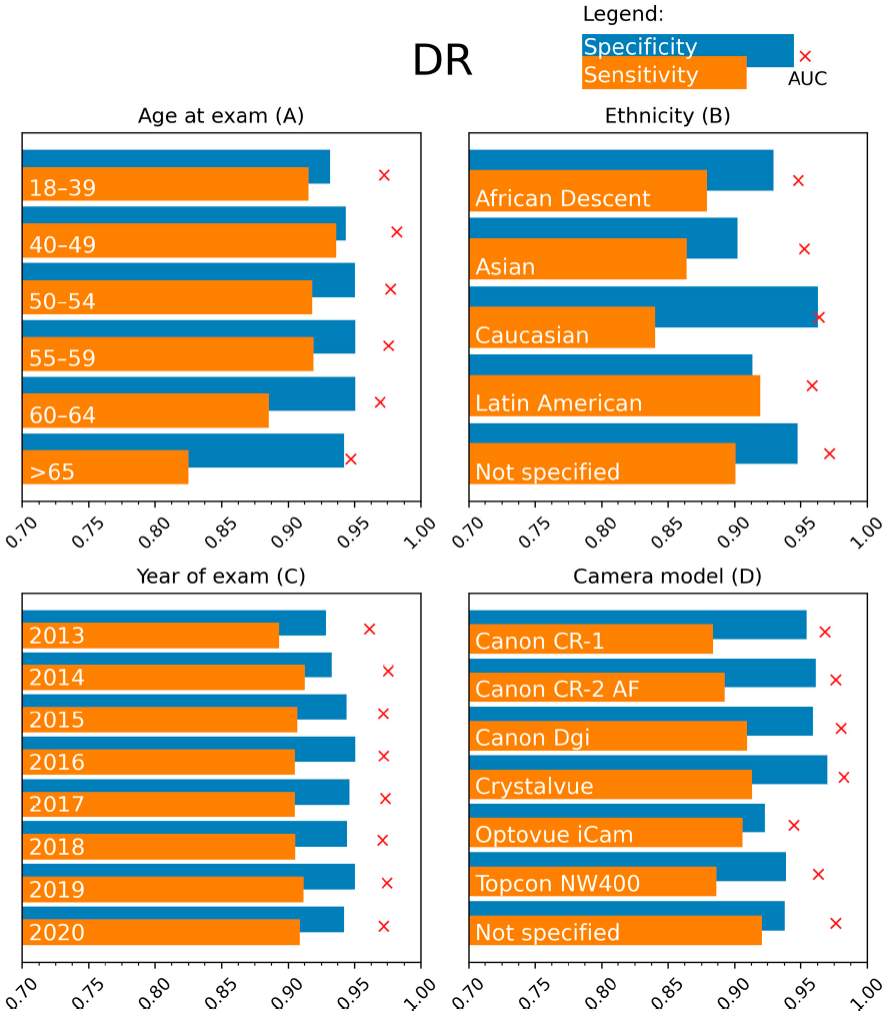
It's disturbingly easy to trick AI into doing something deadly

How "adversarial attacks" can mess with self-driving cars, medicine, and the military.

By Sigal Samuel | Apr 8, 2019, 9:10am EDT

Is an elaborated evaluation the solution?

Data from > 100.000 patient encounters from > 300 medical centers used for model training and evaluation



Journal of *Clinical Medicine*

MDPI

Article

Artificial Intelligence Software for Diabetic Eye Screening: Diagnostic Performance and Impact of Stratification

Freya Peeters [1,2,*,†], Stef Rommes [3,4,†], Bart Elen [4], Nele Gerrits [4,‡], Ingeborg Stalmans [1,2], Julie Jacob [1,2,§] and Patrick De Boever [4,5,§]

vito

REALM

vito.be

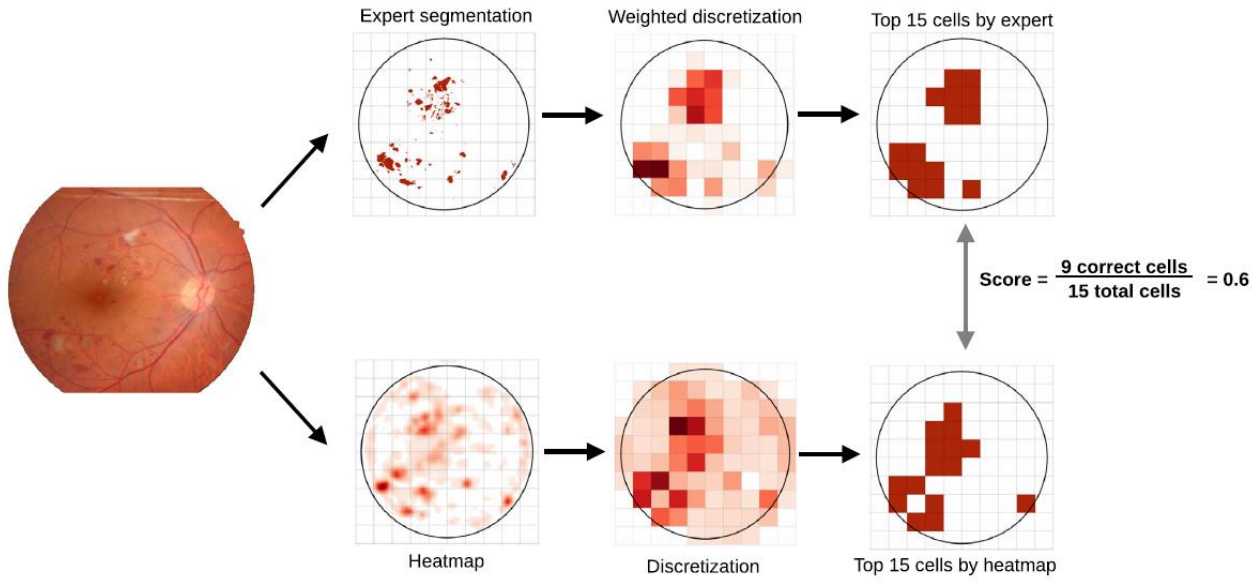
Is explainable AI the solution?

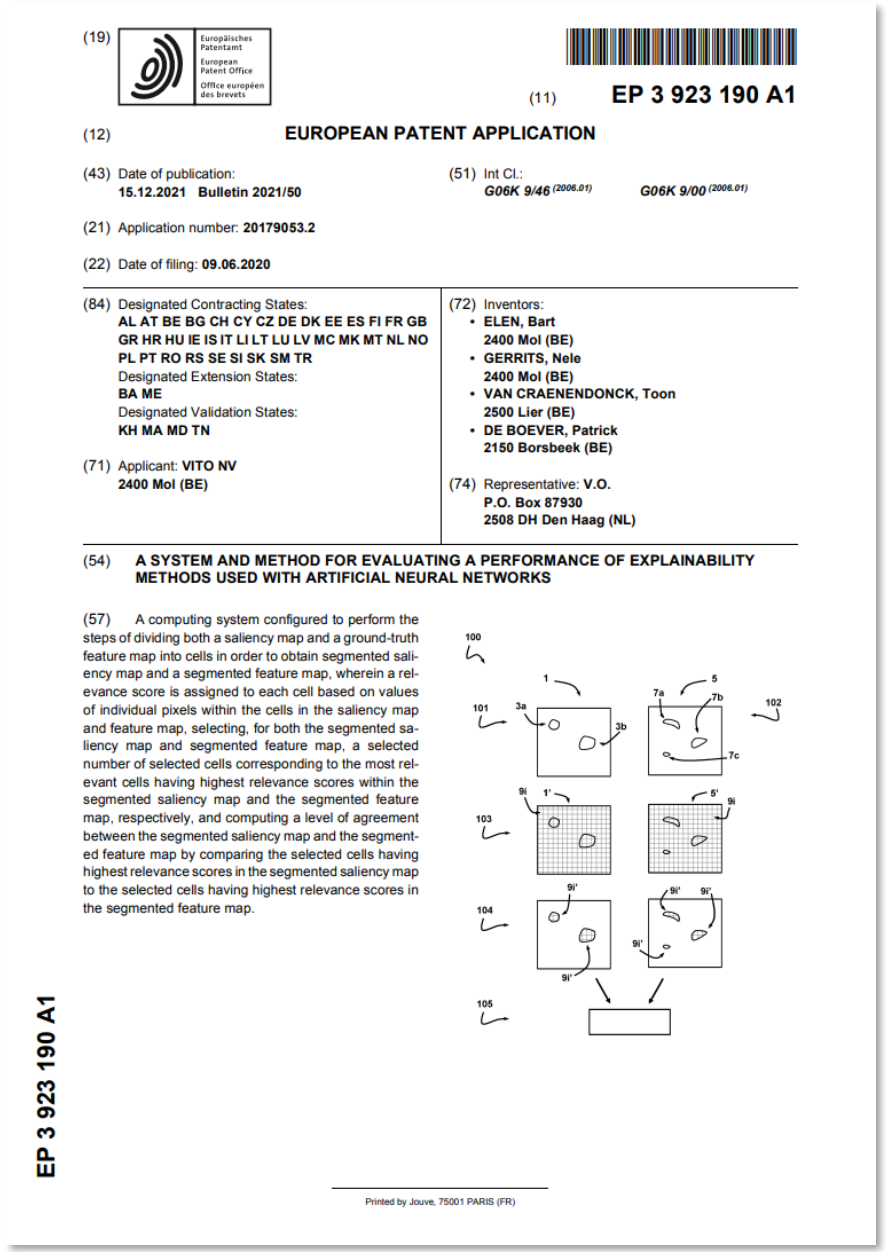
tvst

Special Issue

Systematic Comparison of Heatmapping Techniques in Deep Learning in the Context of Diabetic Retinopathy Lesion Detection

Toon Van Craenendonck[1], Bart Elen[1], Nele Gerrits[1], and Patrick De Boever[1–3]





(19) Europäisches Patentamt / European Patent Office / Office européen des brevets

(11) EP 3 923 190 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication: 15.12.2021 Bulletin 2021/50

(51) Int Cl.: G06K 9/46 (2006.01) G06K 9/00 (2006.01)

(21) Application number: 20179053.2

(22) Date of filing: 09.06.2020

(84) Designated Contracting States: AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States: BA ME
Designated Validation States: KH MA MD TN

(71) Applicant: VITO NV
2400 Mol (BE)

(72) Inventors:
- ELEN, Bart
  2400 Mol (BE)
- GERRITS, Nele
  2400 Mol (BE)
- VAN CRAENENDONCK, Toon
  2500 Lier (BE)
- DE BOEVER, Patrick
  2150 Borsbeek (BE)

(74) Representative: V.O.
P.O. Box 87930
2508 DH Den Haag (NL)

(54) A SYSTEM AND METHOD FOR EVALUATING A PERFORMANCE OF EXPLAINABILITY METHODS USED WITH ARTIFICIAL NEURAL NETWORKS

(57) A computing system configured to perform the steps of dividing both a saliency map and a ground-truth feature map into cells in order to obtain segmented saliency map and a segmented feature map, wherein a relevance score is assigned to each cell based on values of individual pixels within the cells in the saliency map and feature map, selecting, for both the segmented saliency map and segmented feature map, a selected number of selected cells corresponding to the most relevant cells having highest relevance scores within the segmented saliency map and the segmented feature map, respectively, and computing a level of agreement between the segmented saliency map and the segmented feature map by comparing the selected cells having highest relevance scores in the segmented saliency map to the selected cells having highest relevance scores in the segmented feature map.

EP 3 923 190 A1

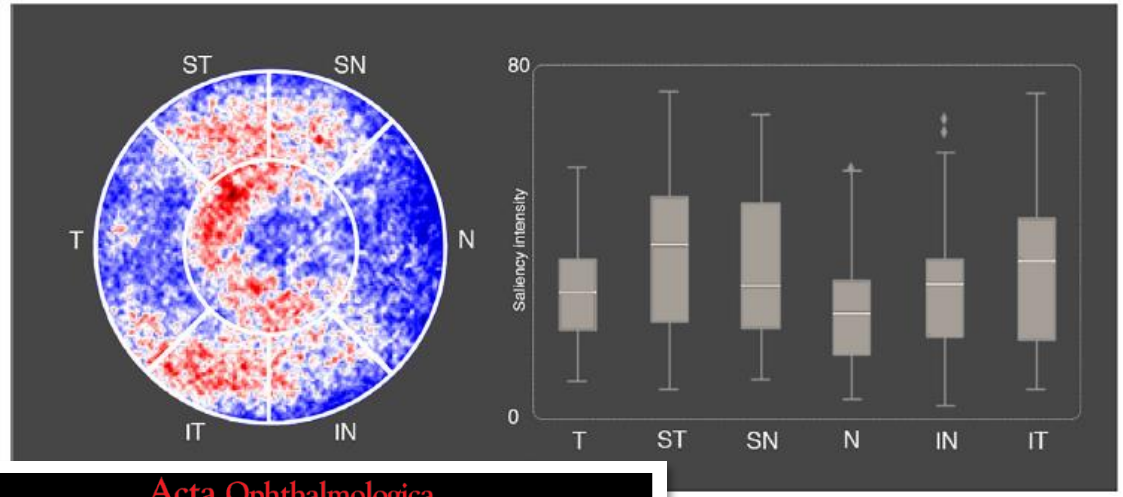
Printed by Jouve, 75001 PARIS (FR)

vito

REALM

vito.be

Is explainable AI the solution?



Acta Ophthalmologica

Acta Ophthalmologica 2019

Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning

Ruben Hemelings,[1,2] Bart Elen,[2] João Barbosa-Breda,[1] Sophie Lemmens,[1] Maarten Meire,[3] Sayeh Pourjavan,[4] Evelien Vandewalle,[1,5] Sara Van de Veire,[6] Matthew B. Blaschko,[7] Patrick De Boever[2,8] and Ingeborg Stalmans[1,5]

AMERICAN ACADEMY OF OPHTHALMOLOGY®

Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy

Rory Sayres, PhD,[1] Ankur Taly, PhD,[1] Ehsan Rahimy, MD,[2] Katy Blumer, BS,[1] David Coz, BS,[1] Naama Hammel, MD,[1] Jonathan Krause, PhD,[1] Arunachalam Narayanaswamy, PhD,[1] Zahra Rastegar, MD, PhD,[1] Derek Wu, BS,[1] Shawn Xu, BS,[3] Scott Barb, MD,[4] Anthony Joseph, MD,[5] Michael Shumski, MD, MSE,[6] Jesse Smith, MD,[7,8] Arjun B. Sood, MD,[9] Greg S. Corrado, PhD,[1] Lily Peng, MD, PhD,[1,] Dale R. Webster, PhD[1,*]*

Unassisted

Grades Only

Grades + Heatmap

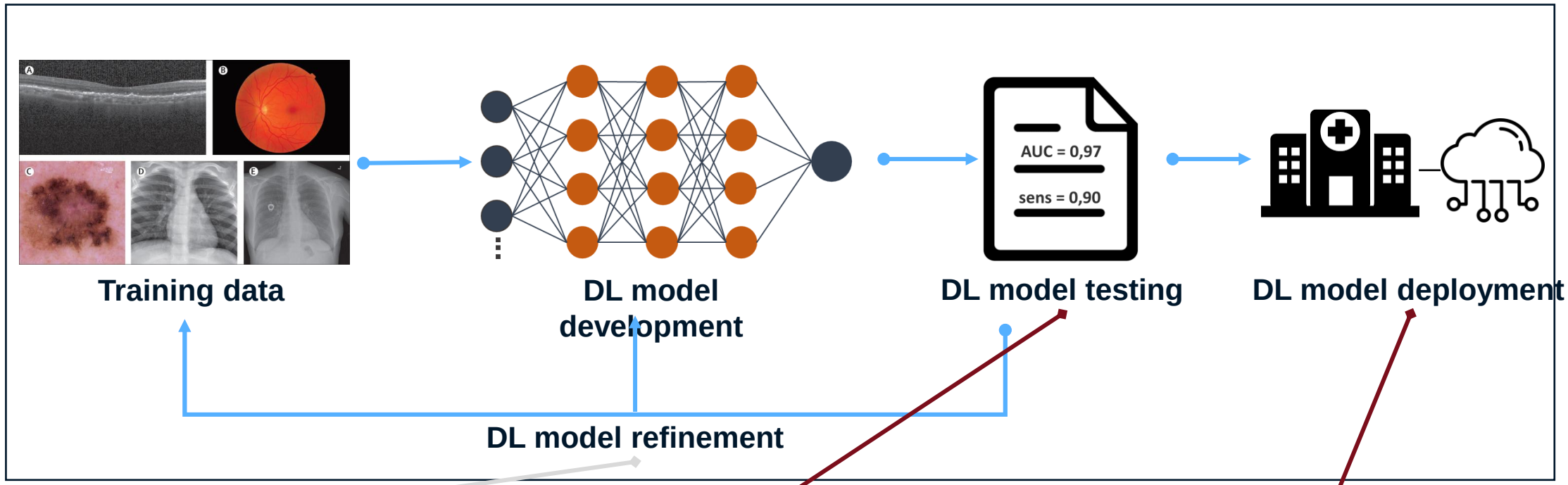
vito

REALM

vito.be

Our research in REALM

AI product life cycle



Innovation 3: Efficient DL model retraining

Reduce need for manual labeling

Innovation 2: Better predict 'real-world' model performance

- Better evaluation methodology
- Model robustness scores
- Estimate performance on fine grained subpopulations using RWD and synthetic data

Innovation 1: smart model monitoring

- Reduce need for manual monitoring
- Be compliant with PMS requirements from regulatory bodies

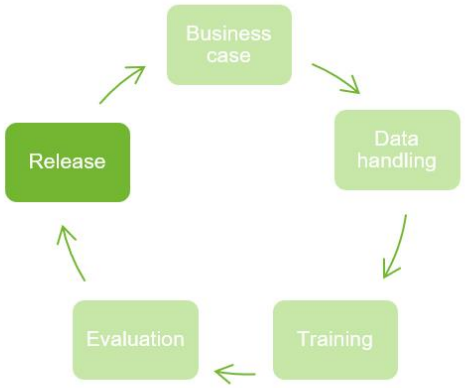
Release

Model performance is monitored after model deployment in clinical practice

AI predictions are checked and corrected by medical experts of AI company

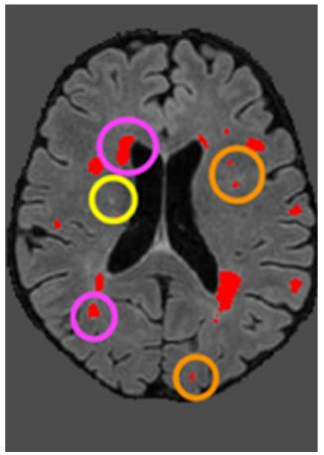
Reduce workload by selecting the cases which will benefit most from being checked

How? - Most relevant cases are identified by quantifying model uncertainty



MS lesion detection

Order MS lesions according to model uncertainty



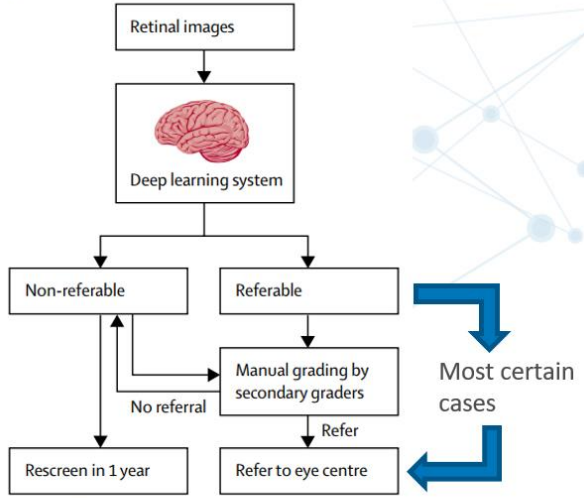
Using uncertainty to determine which lesions to check first gives significantly better results:

F1-score = 76% → 88%

With 50% of lesions checked

DR screening

Don't check most certain diagnoses in semi-automated* screening approach (* most cost-effective according to UK and Singapore studies)



the semi-automated framework can be **further enhanced, and its cost can be reduced** when ***uncertainty is incorporated***

Can DL models be trusted in the clinical practice? Our lessons learned.

SASG meeting – July 3 2023

Bart Elen

vito

REALM

vito.be