Some of the main hurdles to AI development in radiological imaging
- a clinician's perspective

Professor Nicola H Strickland

Consultant Radiologist

Imperial College Healthcare NHS Trust, London

The main hurdles

- Data
- Accuracy
- Regulation and reassurance
- Accountability
- Integration into clinical workflow
- Collaboration with computer scientists
- “Black box” scepticism

The main hurdles

- **Data**
- Accuracy
- Regulation and reassurance
- Accountability
- Integration into clinical workflow
- Collaboration with computer scientists
- “Black box” scepticism

Data

- amount
- access
- anonymisation
- uniformity and quality
- curation (labelling)

Data - some examples: challenges

- **<u>Data access</u> “data are the new gold”**

- PACS data being largely wasted
- no anonymous national database
  - motivation, logistics, federated, cost
- IHE TCE★ profile implementation limited
- access to hospital data behind firewalls
  - inequitable – discriminates in favour of big companies
- different cohorts for training, validating, testing
  - overfitting issues ➔inaccuracy in clinical practice

★*Teaching file and clinical trial export*

anonymised or robustly pseudo-anonymised data may be used without consent

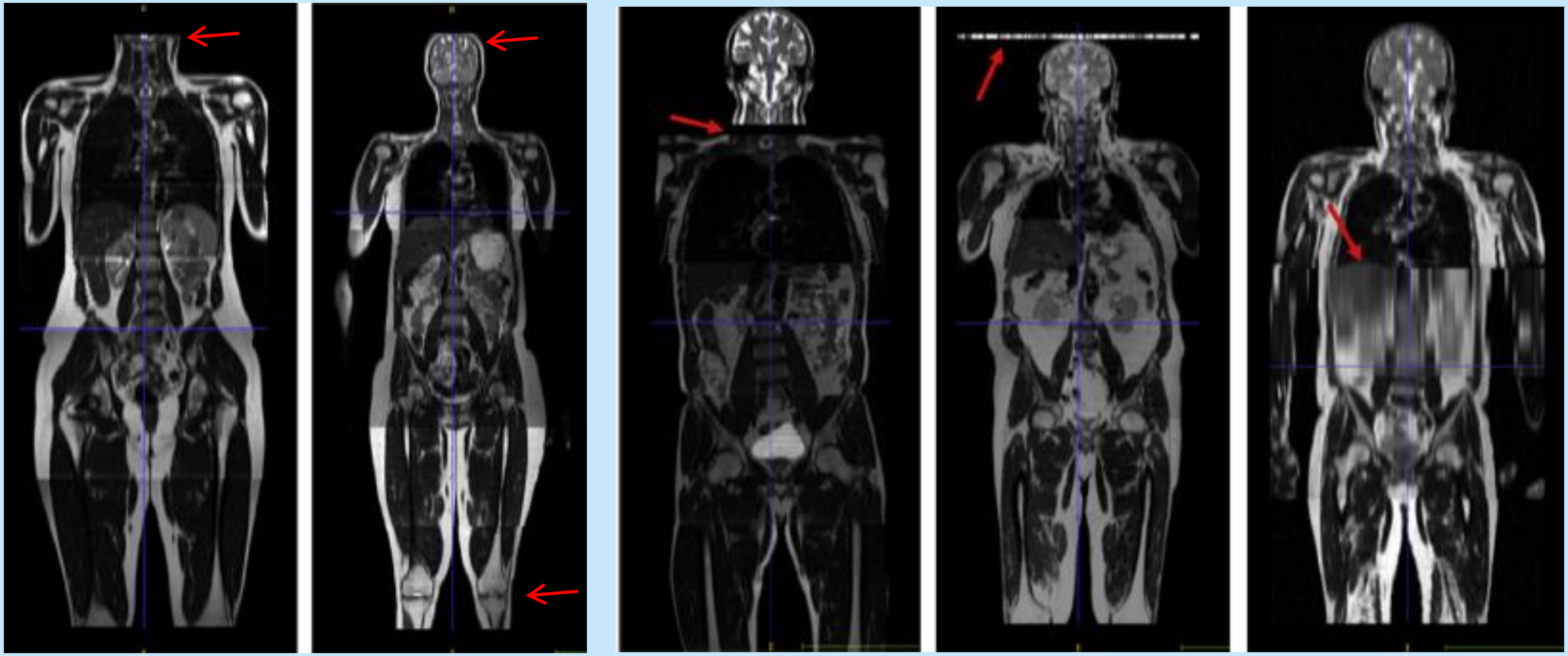
- ethical unease in the UK
- not in India, China

Data - some examples: challenges

- ## Data quality

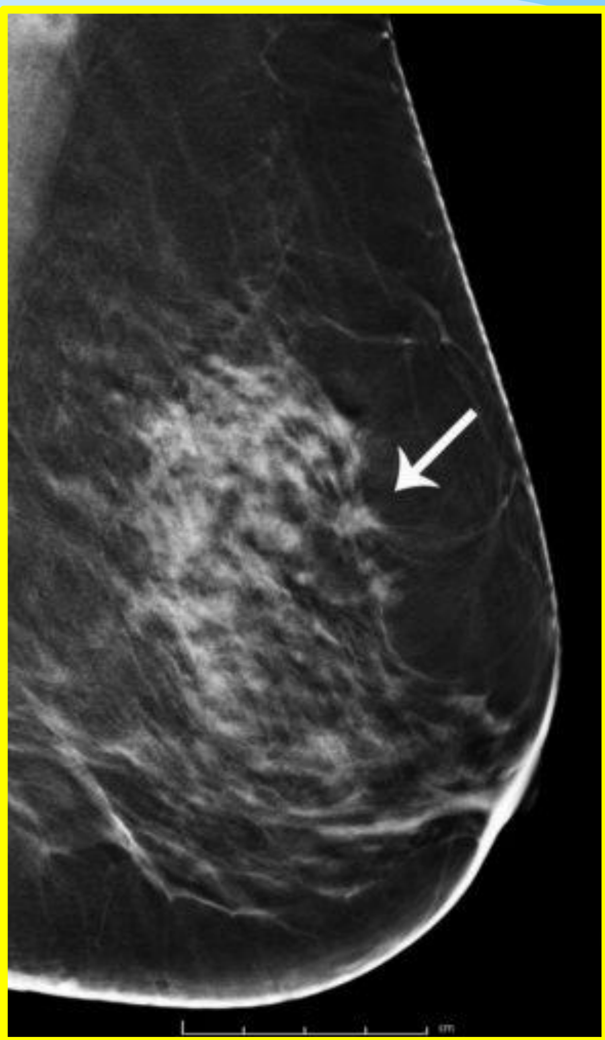
Lavdas I, Glocker B, Rueckert D, Taylor SA, Aboagye EO, Rockall AG. Machine learning in whole-body MRI: experiences and challenges from an applied study using multicentric data. Clinical Radiology (2019) 74: *346-356*

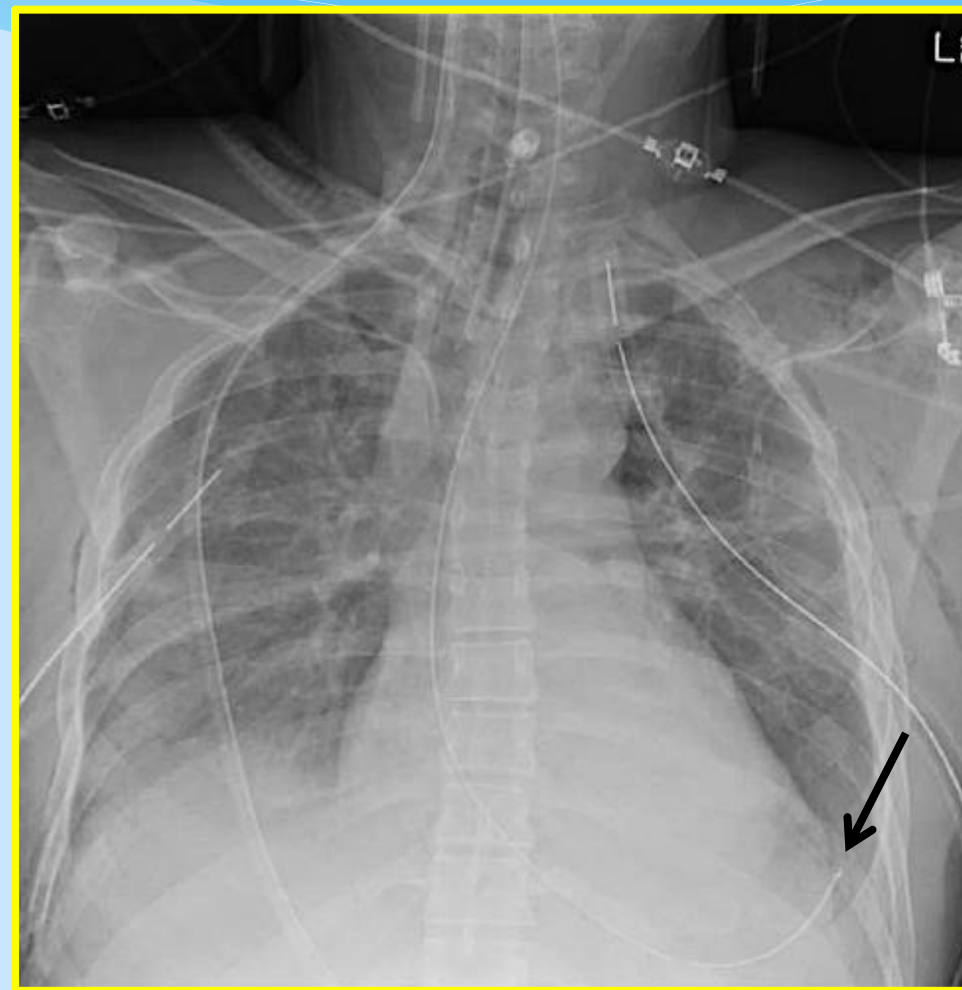
  - **training data**, from which task-specific features are learned, should be **similar to unseen test data**
  - **homogeneous data:** slice width, ?protocol ?contrast ?sequences used (different machine manufacturers)



Data - some examples: challenges

- **Data curation**





The main hurdles

- Data
- **Accuracy – at least equal to that of radiologists**
- Regulation and reassurance
- Accountability
- Integration into clinical workflow
- Collaboration with computer scientists
- “Black box” scepticism

Accuracy: very difficult to measure

- **difficulties establishing 'absolute truth'**
  - often need prolonged follow-up studies
- **published literature: radiological reporting discrepancy rate varies between 2-30%**
- **depends upon:**
  - selection bias
  - case mix
  - imaging modality
  - criteria used to define discrepancy
  - inter- and intra-observer variation in scoring/assessing
- **a good review article:**

Richard Fitzgerald. 'Radiological error: analysis, standard setting, targeted instruction and teamworking'. Eur Radiol (2005) 15:1760-1767

Accuracy: retinal scans

- optical coherence tomography, OCT (3D retinal images)

  Google DeepMind

- >94% accuracy compared with 8 eye experts
- this AI technology can be applied to different types of eye scanners – **vendor independent**
- exemplar of AI (deep learning) development
- collaborative project with Moorfields Eye Hospital, London
- huge dataset 14,884 scans
- data were:
  - cleaned
  - curated (annotated)

Accuracy: optical coherence tomography (OCT)

ARTICLES
https://doi.org/10.1038/s41591-018-0107-6
nature medicine

Clinically applicable deep learning for diagnosis and referral in retinal disease

Jeffrey De Fauw[1], Joseph R. Ledsam[1], Bernardino Romera-Paredes[1], Stanislav Nikolov[1], Nenad Tomasev[1], Sam Blackwell[1], Harry Askham[1], Xavier Glorot[1], Brendan O'Donoghue[1], Daniel Visentin[1], George van den Driessche[1], Balaji Lakshminarayanan[1], Clemens Meyer[1], Faith Mackinder[1], Simon Bouton[1], Kareem Ayoub[1], Reena Chopra[2], Dominic King[1], Alan Karthikesalingam[1], Cían O. Hughes[1,3], Rosalind Raine[3], Julian Hughes[2], Dawn A. Sim[2], Catherine Egan[2], Adnan Tufail[2], Hugh Montgomery[3], Demis Hassabis[1], Geraint Rees[3], Trevor Back[1], Peng T. Khaw[2], Mustafa Suleyman[1], Julien Cornebise[1,3,4], Pearse A. Keane[2,4*] and Olaf Ronneberger[1,4*]

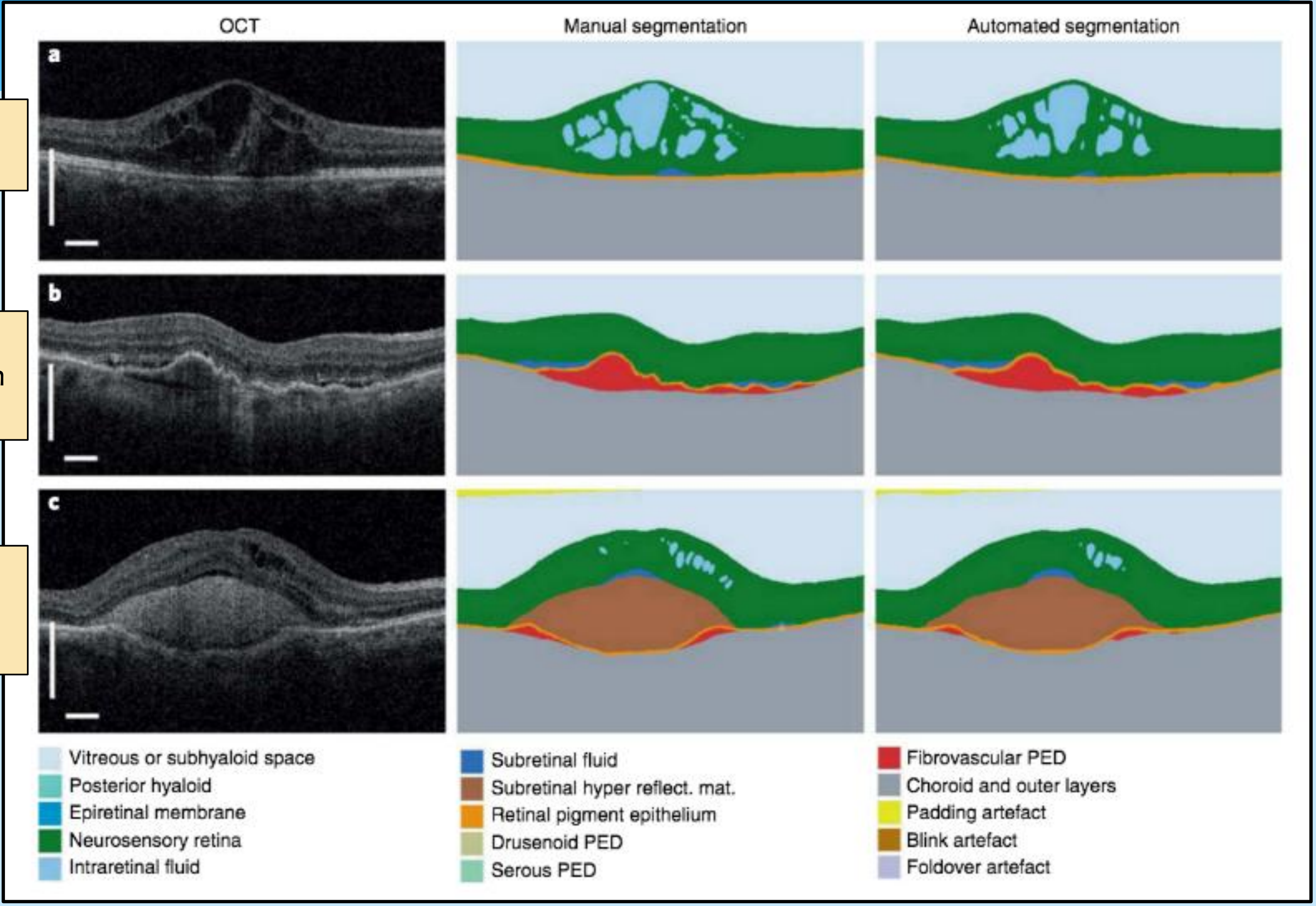
Results have been published: Nature Medicine (2018) 24: 1342–1350

Results of the segmentation network: maps the disease features

diabetic macular oedema

choroidal neovascularization due to AMD

neovascular AMD with subretinal haemorrhage



Nature Medicine (2018) 24: 1342–1350

Accuracy: screening / simple questions

- **as first or second reader**
  - **lung nodules, liver nodules** (e.g. Arterys, Optellum products)
    - lung cancer CT **screening**
    - automatic **detection, segmentation and measurement**
    - **benign vs malignant**
    - **follow-up** tracking of nodules

  - **mammograms** (Kheiron, Hologic products)
    - trials with AI as third reader in UK breast **screening**
      - very good quantitative data exist for accuracy
    - "smart mapping" from 2D to 3D for **suspicious areas**
    - AI breast **density** assessment – trained on BIRADS categories

Accuracy: recognizing vertebral #s
- to detect and treat osteoporosis

- **extracting more information from the data already acquired**

(Zebra Medical Vision)

Compression Fractures Detection on CT

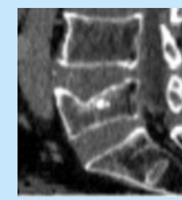
Amir Bar[1,2], Lior Wolf[1], Orna Bergman Amitai[2], Eyal Toledano[2], and Eldad Elnekave[2]

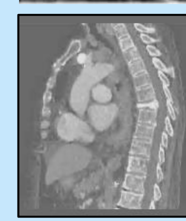
[1]The Blavatnik School of Computer Science, Tel Aviv University
[2]Zebra Medical Vision

ABSTRACT

The presence of a vertebral compression fracture is highly indicative of osteoporosis and represents the single most robust predictor for development of a second osteoporotic fracture in the spine or elsewhere. Less than one third of vertebral compression fractures are diagnosed clinically. We present an automated method for detecting spine compression fractures in Computed Tomography (CT) scans. The algorithm is composed of three processes. First, the spinal column is segmented and sagittal patches are extracted. The patches are then binary classified using a Convolutional Neural Network (CNN). Finally a Recurrent Neural Network (RNN) is utilized to predict whether a vertebral fracture is present in the series of patches.

Keywords: compression fracture, osteoporosis, convolutional neural networks, recurrent neural networks

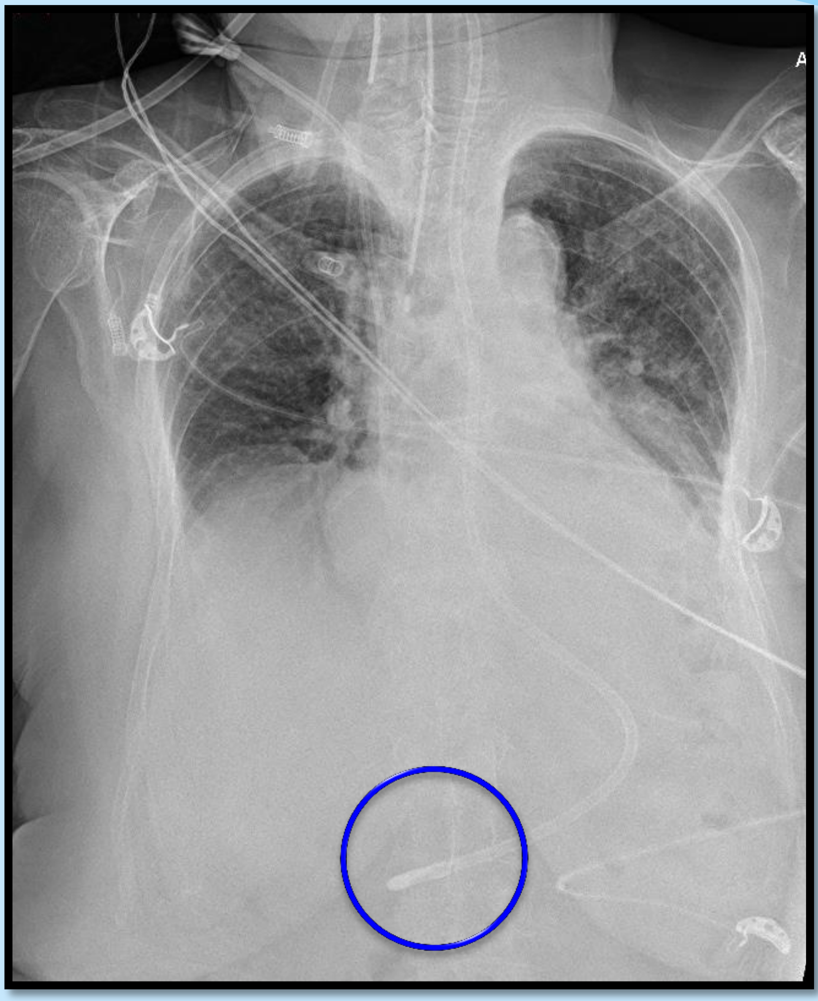


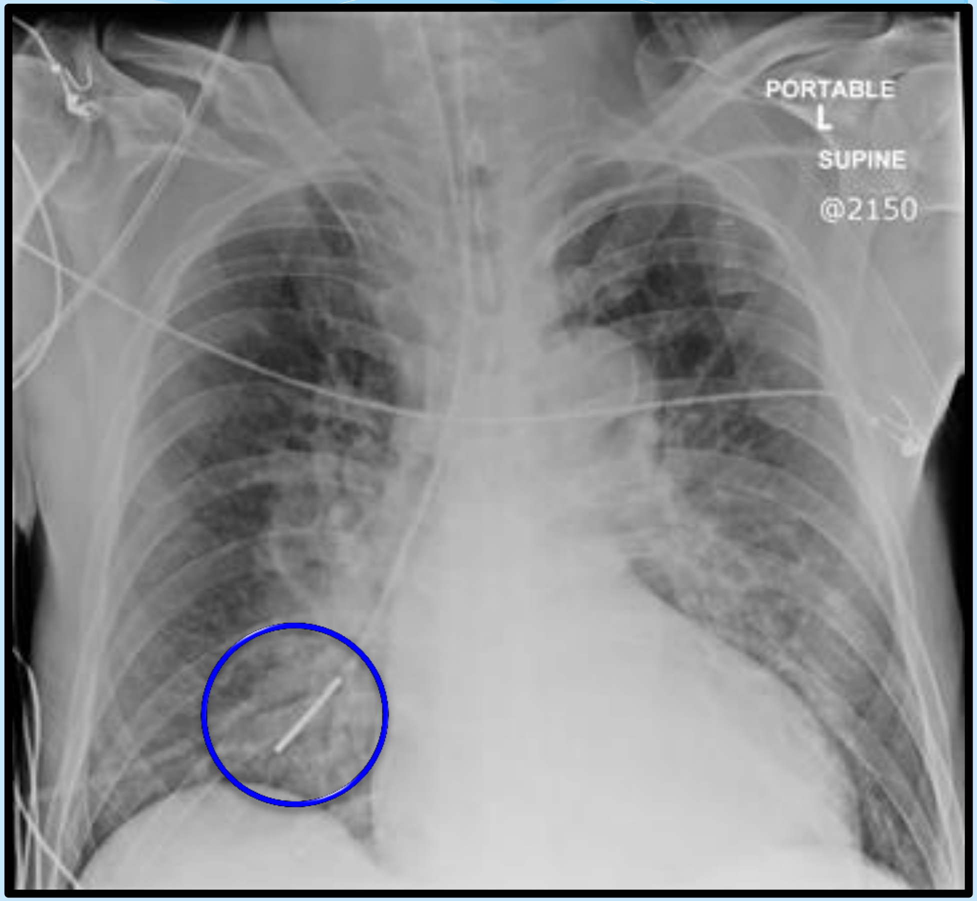


https://www.zebra-med.com/research-publications/e5148770d224518052ada5b254fa5fecd4b2/

Accuracy: what is acceptable?

e.g. nasogastric tube position – zero error tolerance?





The main hurdles

- Data
- Accuracy
- **Regulation and reassurance**
  - that the AI has been properly tested
- Accountability
- Integration into clinical workflow
- Collaboration with computer scientists
- “Black box” scepticism

AI regulation – comparison with new drugs

- **New drugs**:
  - they don't evolve whilst in use – unlike AI
  - by law, must have a product license, from a medicines' regulator, before going on the market
  - in UK: MHRA (Medicines & Healthcare products Regulatory Agency)
  - trained MHRA assessors review all available evidence from pre-clinical research and clinical trials
  - MHRA also inspects manufacturing factory – supplies of uniform and high standard

AI regulation – comparison with new drugs

- **AI software:**
- **CE Mark** (Conformité Européene)
  - MHRA works with Notified Bodies from anywhere in Europe to approve these AI algorithms, but the algorithms are not actually tested independently
  - Notified Bodies look at the **controls and clinical governance** in place in the companies making the AI algorithms

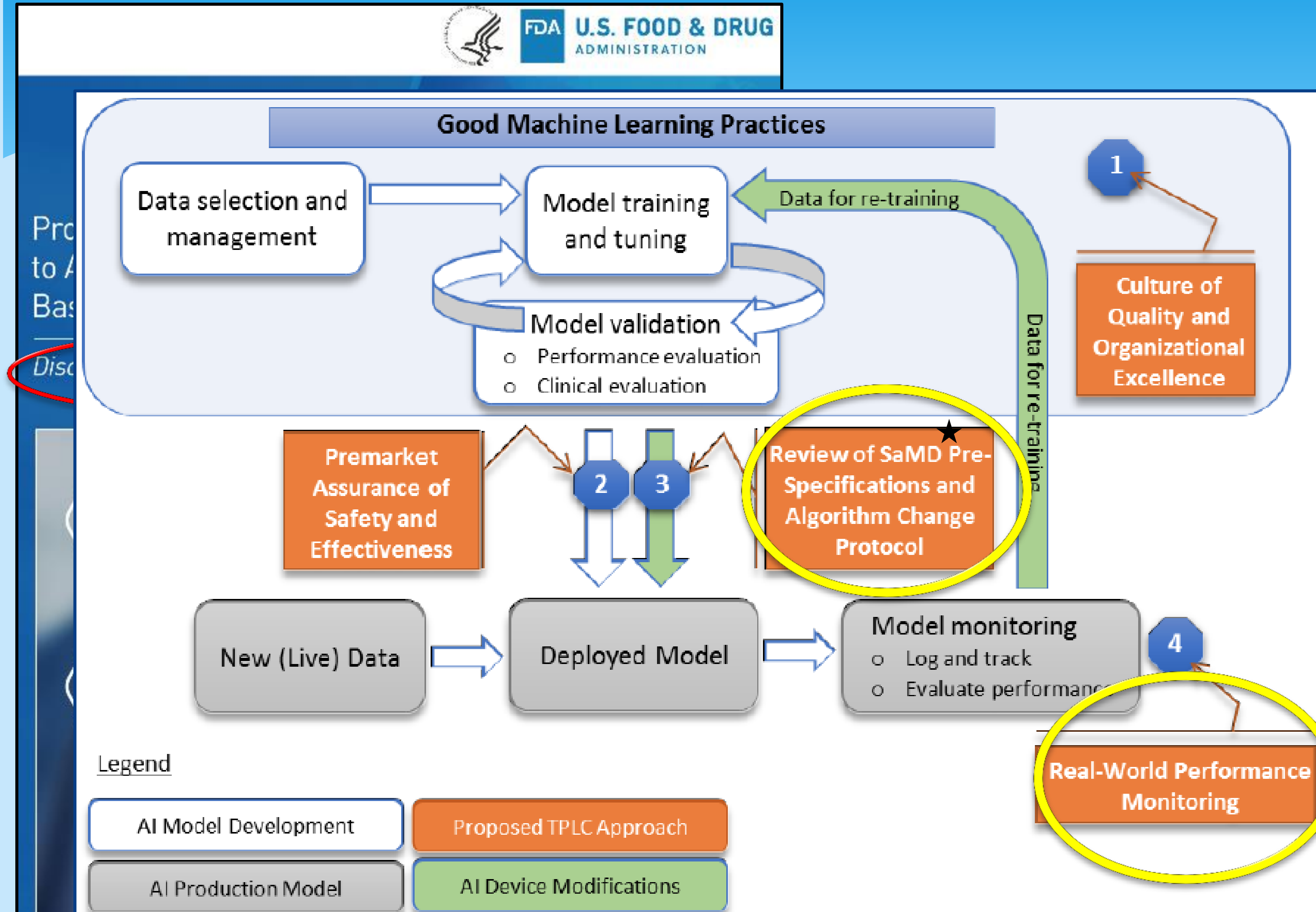
The problems with AI regulation

Why it is currently inadequate

- **lack of rigour in regulatory testing**
- **AI algorithms coming to market with CE mark and/or FDA approval, without:**
  - having been independently tested
  - publication in peer-reviewed literature
- **huge problems associated with the testing process:**
  - **data:** amount, access, quality
  - **resource**
  - **workforce**
- **AI/ML spectrum of continuous learning**

locked algorithm ⟶ adaptive algorithm

FDA Regulation



★ SaMD – software as medical device

Reassurance that AI software has been properly tested, before introduction into clinical practice

- **By whom? How? Where?**
- **Who is going to be involved in the UK?**
  - MHRA expert panel - (Medicines & Healthcare products Regulatory Agency)
  - CQC – (Care Quality Commission)
  - NICE – (National Institute for Health and Care Excellence)
  - NHS X: keen to test these technologies in the NHS context and gather evidence of accuracy, efficacy and value

‘Proper testing’ of AI: example of concerns

- thrombotic stroke detection alerts with CE mark (viz.AI)
  - **large vessel occlusions, LVOs**
  - **analyses data directly on CT scanner→notifies mobile device of neurorad/stroke physician**
  - **6 mins** (versus 52 mins)
  - *but:* analysed **only 300 CTA studies vs 2 neurorads!**

    90% sensitivity and specificity
  - **no peer reviewed publication**

The main hurdles

- Data
- Accuracy
- Regulation and reassurance
- **Accountability**
- Integration into clinical workflow
- Collaboration with computer scientists
- “Black box” scepticism

Accountability:
who takes the blame when AI is wrong?

- the Radiologist?
  - no!
  - ?blamed for *NOT* using AI algorithm if available
- **the hospital** – same as now
- urgent need to **educate the public about error**
  - radiology reporting is an opinion – not an exact result "cancer/not cancer"
- the unique feature of **AI: constant "learning" changes its performance** (? for the better)
  - lock down the algorithm?

The main hurdles

- Data
- Accuracy
- Regulation and reassurance
- Accountability
- **Integration into clinical workflow**
- Collaboration with computer scientists
- “Black box” scepticism

Integration into workflow:

- **AI products *must* be seamlessly integrated into RIS/PACS/EPR (and radiotherapy planning)**
  - otherwise won't be used
  - cf stand alone MPR/other software, CADs
- **vendor neutral interfacing standards do now exist, so no excuse for not using them**
- **where in the pathway should the AI algorithm be integrated?**
  - e.g. between image acquisition device and PACS, with on/off toggle on PACS

Between image acquisition and PACS

- **as first or second reader**
  - **lung nodules, liver nodules**
    - lung cancer CT **screening**
    - automatic **detection, segmentation and measurement**
    - **benign vs malignant**
    - **follow-up** tracking of nodules

  - **mammograms**
    - trials with AI as third reader in UK breast **screening**
    - “smart mapping” from 2D to 3D for **suspicious areas**
    - AI breast **density** assessment – trained on BIRADS categories

Integration: organizing hospital workflow

NEWS | ARTIFICIAL INTELLIGENCE | OCTOBER 08, 2018

CHI Franciscan Launches Washington State's First AI-Powered Hospital Mission Control Center

Advanced analytics software platform will enhance patient safety, speed delivery of care and support quality outcomes

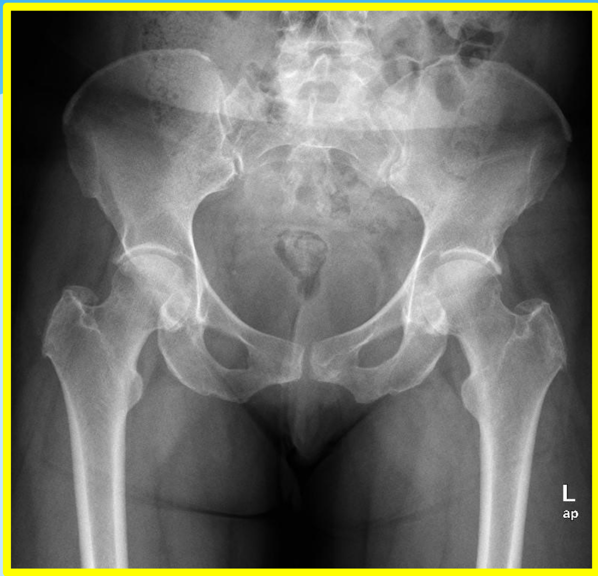
The Mission Control Center will use AI and predictive analytics to optimize care coordination, speed care delivery and improve the patient experience, while maintaining patient privacy. The system works by looking at each individual hospital as part of a larger system, continually examining real-time data and using machine learning to recommend actions that can predict and prevent risk, balance staff workload and streamline the discharge process so patients can get home sooner.

Integration: radiography technique improvement

- **best CT or MR protocol for specific patient**
  - cardiac CT and MR
    - ECG analysis
    - clinical question
    - pertinent data from EPR

- **artefact correction**
    - movement/breathing
    - metallic implant

Integration: application to radiology worklists



- **Recognize normals**

➢ de-prioritise normals to bottom of the worklist

➢ allow radiologists to concentrate on the abnormals

➢ avoid patients with serious pathology waiting weeks for diagnosis

➢ speed up reporting of the normal studies

The main hurdles

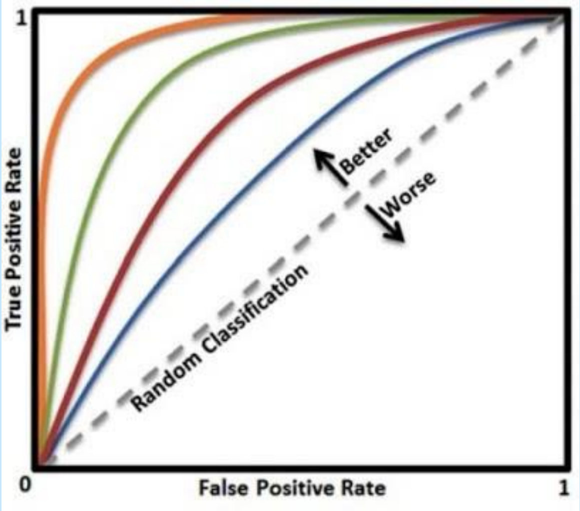
- Data
- Accuracy
- Regulation and reassurance
- Accountability
- Integration into clinical workflow
- **Collaboration with computer scientists**
- “Black box” scepticism

Collaboration with computer scientists

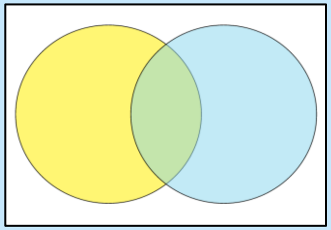
- **not happening enough!**
  - cf the early days of PACS
- **a 2-way process**

1. **Clinicians must empower themselves to understand the major concepts of AI ➔thereby understand the hurdles**
   - medical student and post-grad syllabi
   - basic statistics and terminology
   - equip themselves to be able to judge AI in clinical practice

ROC curves:



DICE indices:



AI contour
expert-drawn contour

Collaboration with computer scientists

- **a 2-way process**

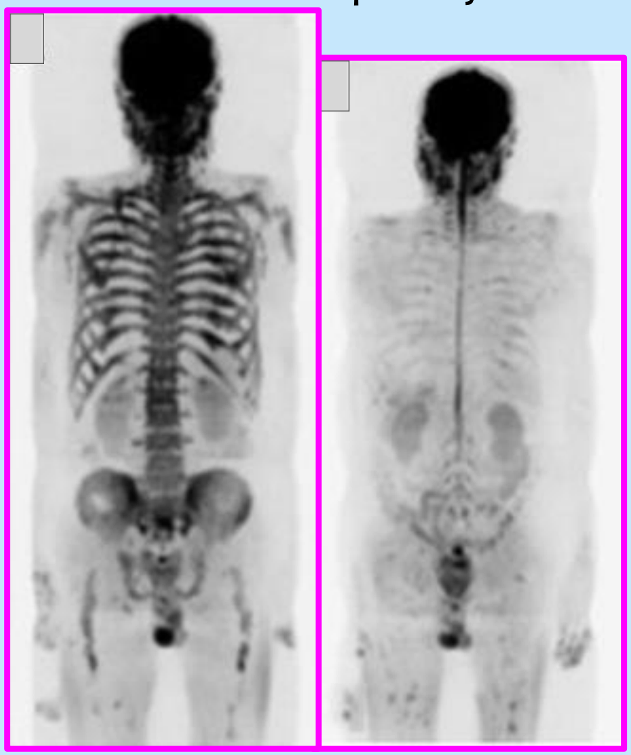
1. **Clinicians must empower themselves to understand the major concepts of AI → thereby understand the hurdles**
   - medical student and post-grad syllabi
   - basic statistics and terminology
   - equip themselves to be able to judge AI in clinical practice

2. **Computer scientists must work with clinicians to understand the clinical needs from AI**
   - **removal of "drudgery"**
   - **longer term goal: radiogenomics**
   - **LMIC versus HIC needs**

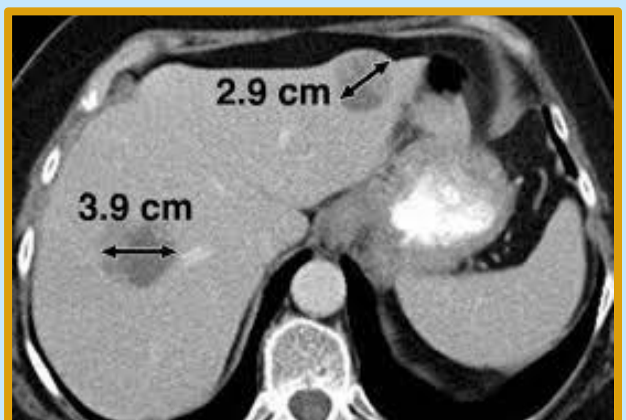
Drudgery: disease related quantitation

- with prepopulation of reports
  - metastatic burden
  - metastatic size,
  - progression of IPF

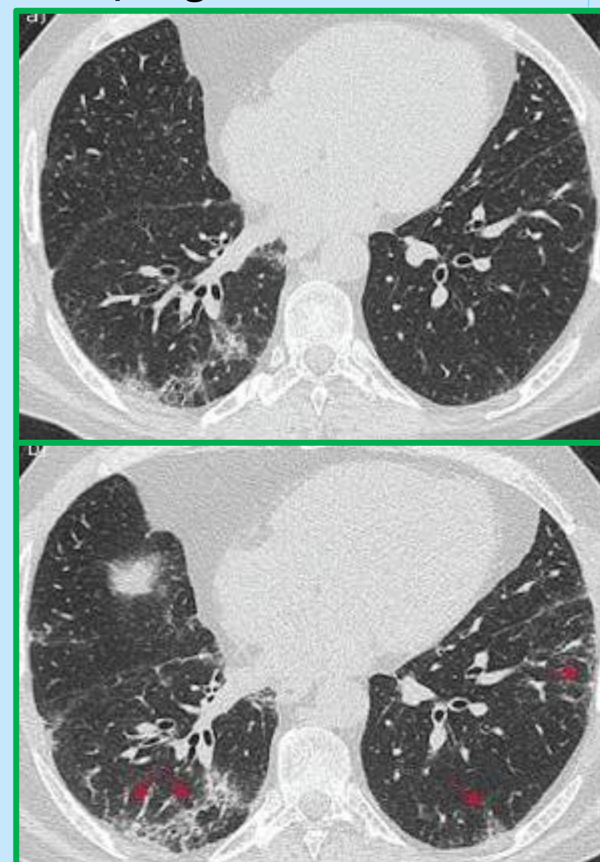
- WB MRI multiple myeloma



- CT RECIST measurements



- progression of IPF



What we "<u>need</u>" from AI depends upon the healthcare setting

- LMIC versus HIC
- AI as the <u>**only**</u> reporter
- something much better than nothing in some settings
- in these settings it doesn't matter that:
  - AI trained for one task only
  - AI fails to make associations as human brain does

AI in LMIC:
gestational age vs fetal maturity

automated in utero ultrasound

Med Image Anal. 2015 Apr;21(1):72-86. doi: 10.1016/j.media.2014.12.006. Epub 2015 Jan 3.

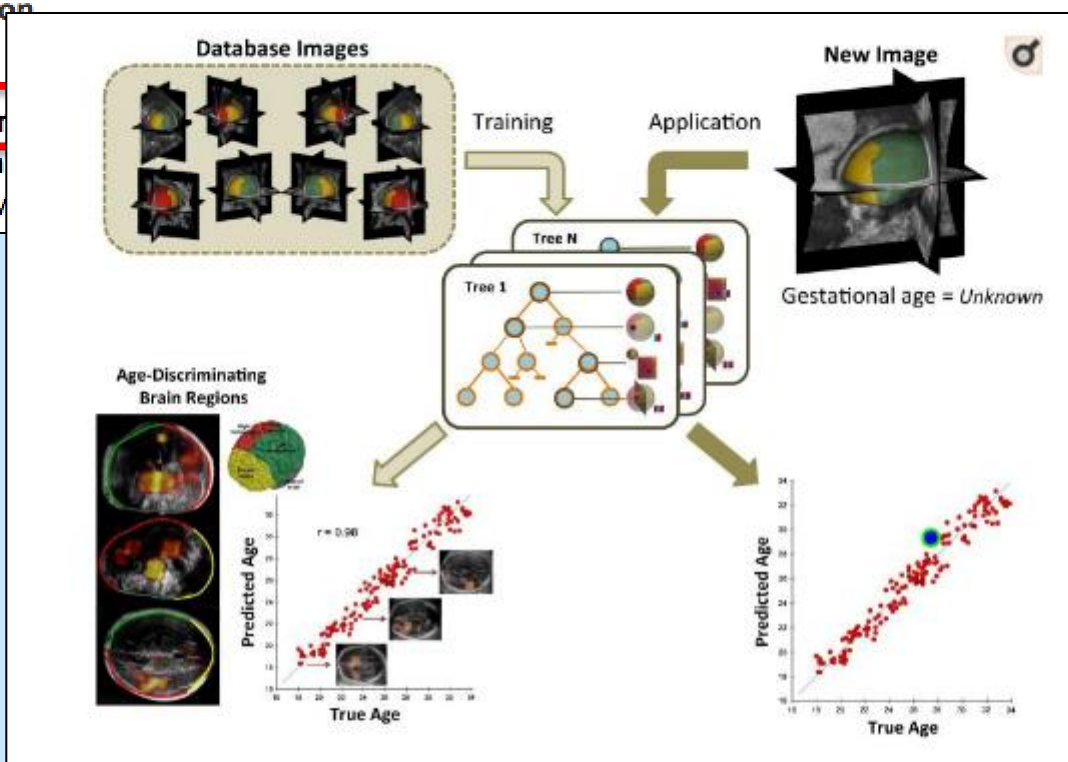
Learning-based prediction of gestational age from ultrasound images of the fetal brain.

Namburete AI[1], Stebbing RV[2], Kemp B[3], Yaqub M[2], Papageorghiou AT[3], Alison Noble J[1].

Author information

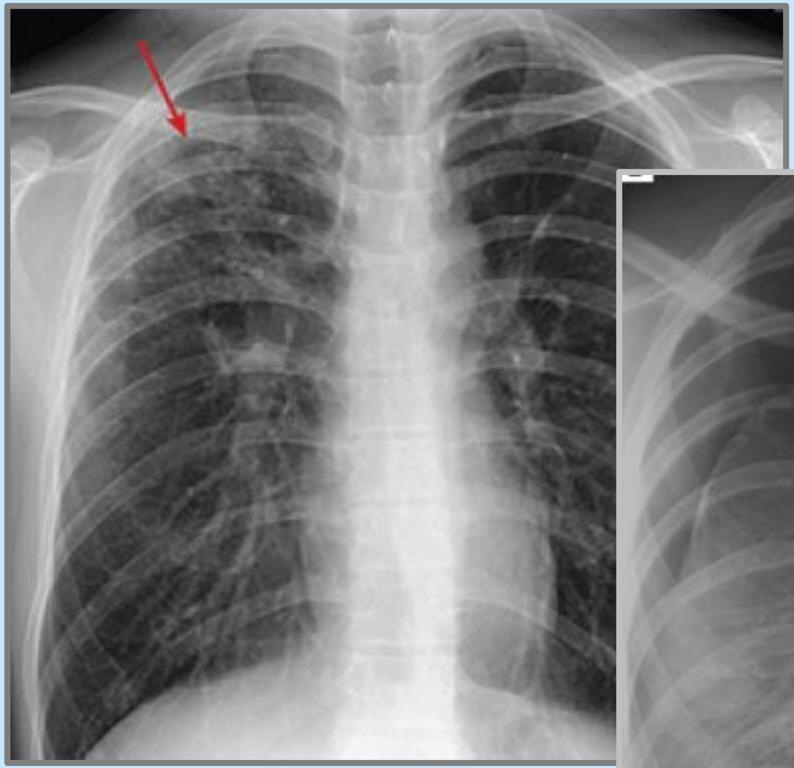
Abstract

We propose an autom … ration of a fetus based on 3D ultrasound (US) brain … terns in conjunction with clinical measurements to dev … potential of US images. The



AI in LMIC:
detection of specific CXR abnormalities

specific important feature detection on CXR, not a full CXR report – e.g:

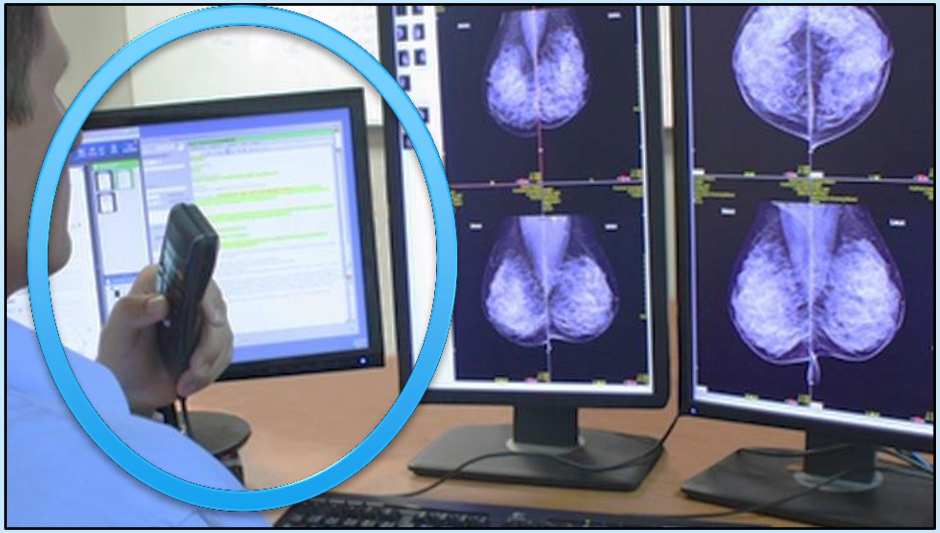


The main hurdles

- Data
- Accuracy
- Regulation and reassurance
- Accountability
- Integration into clinical workflow
- Collaboration with computer scientists
- **“Black box” scepticism**

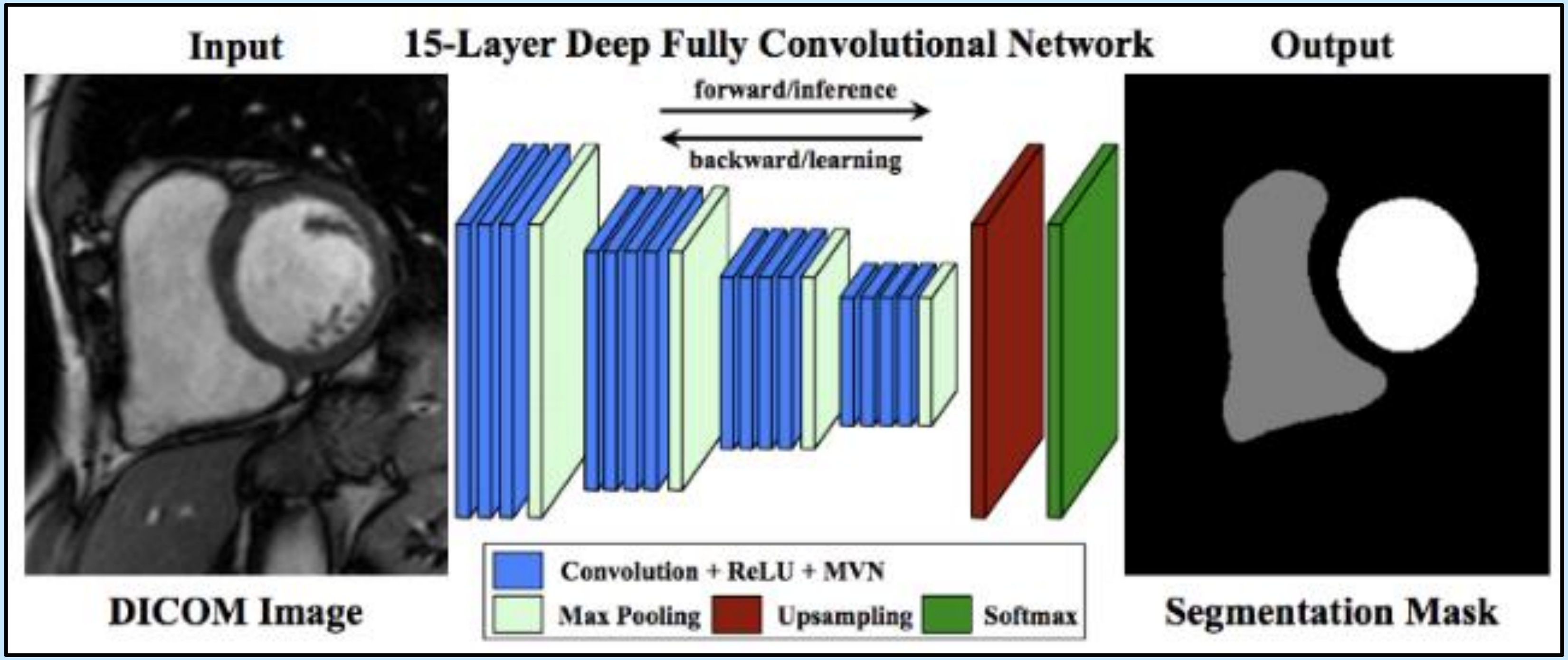
"Black box": speech recognition

- 20 years use in radiology reporting
  - natural language processing
  - neural networks
- seamlessly integrated into clinical radiological practice
- continues to learn whilst in use
- 5% error rate
- transparent outcome
- confidently over-ride



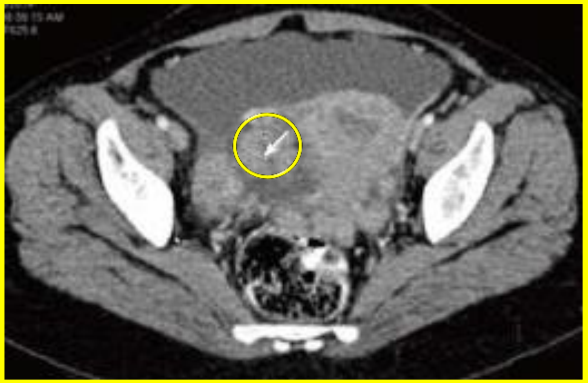
"Black box": Cardiac MR/CT segmentation

- automated LV and RV segmentation
- Clinicians accept because:
  - visually accurate
  - high DICE index $\geq$ 0.9



"Black box": radiomics, radiogenomics

- **Image-based precision 'personalised' medicine in:**
  - **diagnosis**
  - **prognosis assessment**
  - **therapy response prediction**



AI data-mining extraction of quantitative features in the imaging data not appreciated by the naked eye → combined with other patient data (genomics, clinical features) → discover patterns in large data sets → "the answer"

- **Clinicians understandably sceptical:**
  - have no way of checking accuracy of algorithm (even in longterm)
  - no understanding of the "quantitative features"
  - loss of control = scary
  - all taken on "trust"

Ref: Gillies RJ et al. Radiomics: images are more than pictures, they are data. Radiology, 2016. Vol 278, issue 2

Conclusion:

I have used clinical examples to give
a Clinician's perspective on

<u>The main hurdles to AI development in radiological imaging</u>:

- Data
- Accuracy
- Regulation and reassurance
- Accountability
- Integration into clinical workflow
- Collaboration with computer scientists
- "Black box" scepticism