USE CASE Widening the AI deployment bottleneck:
Clinical AI Fellowship Programme - a first of its kind

DR ALEXANDER T DENG

Programme Lead for the Fellowships in Clinical Artificial Intelligence

Guy's and St Thomas' NHS Foundation Trust & AI Centre

Guy's and St Thomas' NHS Foundation Trust

AI CENTRE for Value Based Healthcare

Fellowships in Clinical Artificial Intelligence:

Equipping healthcare leaders to adopt clinical AI

Dr Alexander T Deng

Programme Lead

GSTT Clinical Scientific Computing

CSC Clinical Scientific Computing

Developing people, platforms and policy for digital health

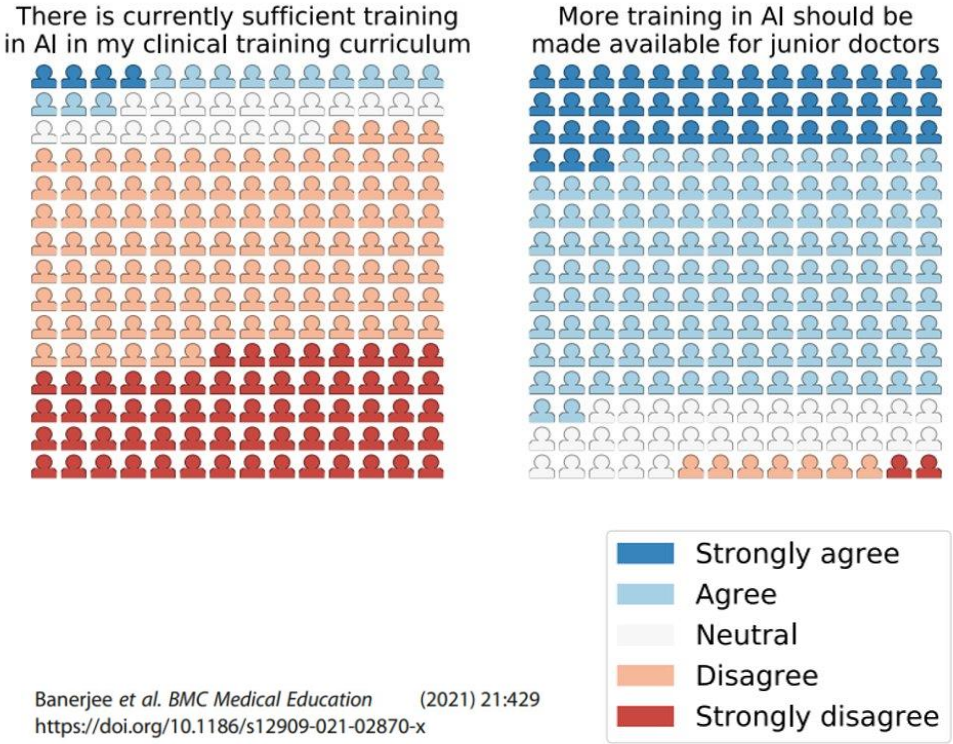
Ambition

The Topol Review

Preparing the healthcare workforce to deliver the digital future

The Digital Medicine and AI & Robotics Panels recommend:

- The NHS should create or increase the numbers of clinician, scientist, technologist and knowledge specialist posts with dedicated, accredited time, with the opportunity of working in partnership with academia and/or the health tech industry to design, implement and use digital, AI and robotics technologies. (DM4/AIR5)



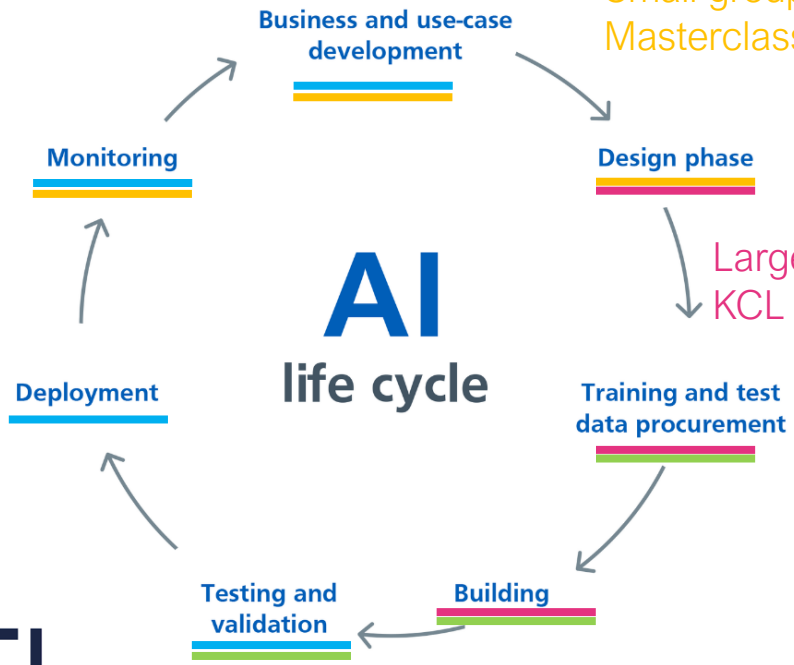
Banerjee *et al. BMC Medical Education* (2021) 21:429
https://doi.org/10.1186/s12909-021-02870-x

Multimodal curriculum design

The post lasts 1 year, at 2 days per week (0.4 FTE)

Immersive clinical project with 1:1 clinical AI supervision

Small group teaching: Masterclass workshops



Large-group teaching: KCL Innovation Scholars programme

Developing healthcare workers' confidence in AI

Report 2 of 2

October 2022

NHS AI Lab & Health Education England

FACULTY OF CLINICAL INFORMATICS

Core Competency Framework for Clinical Informaticians

Self-directed learning: Programming in Datacamp

Partners

AI CENTRE for Value Based Healthcare

Lucy O'Neill
Head of AI Programmes
London AI Centre

Dr Alexander Deng
Programme Lead
Medical Doctor
Medical Educationalist

NHS
Health Education England
Digital Transformation

Health Innovation Network South London

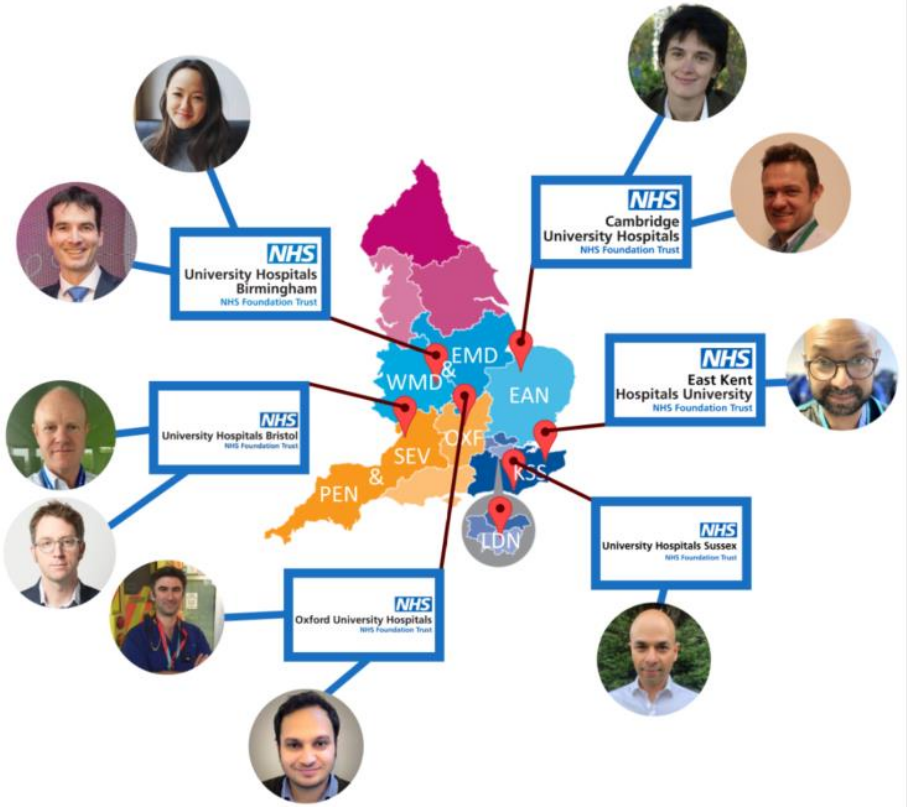
Medicines & Healthcare products Regulatory Agency

NICE
National Institute for Health and Care Excellence

NHS Digital

NHS
Health Education England

Scaling up and spread



Value delivered

Health Innovation Network South London

"Fellows and supervisors see significant benefit and value in the **real world experience** the programme offers and feel the broad structure and approach of the programme is fit for purpose.

In particular, **project placements and masterclass sessions are seen as offering significant value**."

Artificial intelligence for chest X-ray interpretation

Dr Sarah Blake, Dr Neelan Das, East Kent Hospitals University Trust

The project

A deep learning algorithm called qXR developed by Qure.AI

Provides automated report of chest x-rays (CXR) classifying into normal vs abnormal scans

Trained on 2.3 million chest x-rays

The aim was to deploy qXR in East Kent Hospital University Trust (EKHUT) to improve accuracy of reporting in ED

Examples of qXR images and reports

Evaluation

The deployment date is set for 1st May 2023

qXR will require ongoing monitoring to assess discrepancies. This involves weekly audit of reports and self-reporting to a dedicated email. Selected Scans will then be discussed at a monthly meeting.

Validating qXR

An audit of CXRs in ED showed evidence of misinterpretation of CXR resulting in incorrect patient care

A validation study on 1000 chest x-rays taken at EKHUT was performed

Ground truth is the radiologists report

It shows acceptable accuracy for use, with a sensitivity 98% and a specificity 65%

Deployment

We created a survey for clinical staff to evaluate perceptions of AI and improve engagement with qXR

I ran training sessions on using qXR for the Emergency (ED) and Radiology departments

Deploying qXR required close working with procurement, legal, IT and clinical safety teams to approve use of qXR in a pilot study in ED on 2 scanners

Scan to read more about qXR at EKHUT

The future of qXR...

"Supersurers" will be trained to continue evaluation after wider deployment.

The clinical impact of qXR will be measured to evaluate changes in patient outcomes

If evaluation of qXR performance is good, the software will be integrated into all scanners in the trust.

Guy's and St Thom

Automated extraction of dementia related volumetric assessments from radiology reports using natural language processing

Dr Adam Paynter, Dr Angus Roberts, Prof Sig Sanbrook, Dr Ashwin Sankaranarayan & Prof. Robert Stewart, King's College London

Introduction:

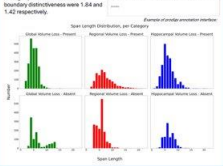
Brain MRI scans are performed on most patients coming to NHS memory clinics. We demonstrate the use of a machine learning algorithm to extract and label text relating to brain volume from a large real-world database of brain MRI reports performed on patients referred from multiple memory clinics in South London (SLAM Image Bank).

Annotation and Statistics:

9154 span examples were generated following annotation. Each category contained between 951 and 2165 examples, and span length ranged from 1 to 22 tokens with a median span length of 6.

Average span distinctiveness and boundary distinctiveness were 1.84 and 1.42 respectively.

Example of annotation interface



Span Length Distribution, per Category

Materials and Methods:

Due to the presence of nested/overlapping terms of interest and long phrase a span categorisation was used. This involves tagging specific contiguous groups of tokens as referring to a particular category, and six target categories were identified;

1) presence of global volume loss
2) absence of global volume loss
3) presence of regional volume loss
4) absence of regional volume loss
5) presence of hippocampal / medial. temporal lobe volume loss
6) absence of hippocampal /MTL volume loss

4029 brain MRI reports were extracted from the radiology information system (RIS) between February 2008 to September 2021. These plain text reports were then annotated using the prodigy annotation software.

Results:

3206 of the reports (80% of the total) formed the corpus for the training of a neural network model using the open source spaCy python library, with a further 805 reports (20%) kept as a separate test set.

5-fold cross validation was used to assess model performance during hyperparameter tuning. Against the test set over all six categories the model's F-score was 0.88 (precision = 0.92, recall = 0.84), with variability between the different categories:

Category	Precision	Recall	F Score
Global Volume Loss – Present	0.93	0.94	0.94
Global Volume Loss – Absent	0.94	0.95	0.94
Regional Volume Loss – Present	0.80	0.60	0.68
Regional Volume Loss – Absent	0.95	0.81	0.88
Hippocampal Volume Loss – Present	0.91	0.89	0.90
Hippocampal Volume Loss – Present	0.96	0.90	0.94

Example of Model output

Discussion:

The relative underperformance of the 'regional volume loss' category is expected given the underlying problem; if regional volume loss is present, then any portion of the brain volume may be mentioned making the category very heterogenous. This is reflected in the data, with the RVL category having the broadest range of span length (2-22) and the lowest span distinctiveness (1.38).

Conclusions:

These results support the feasibility and accuracy of natural language processing applied to routine radiology reports and the potential for automated generation of novel meta-data from dementia assessments in electronic health records. This approach is planned to be implemented as part of routine data processing.

Predicting high intensity users in mental healthcare

Fellow: James Lai[1, 2, 3]
Supervisor: Prof Rob Stewart[1, 2]
[1]South London and Maudsley Hospital; [2]Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London; [3]Emergency Department, Royal Free London NHS Trust

Background

Mental healthcare is costly with estimates around £22.5 billion per year.

Previous work used routinely recorded data captured in electronic healthcare records (EHR) to predict service outcomes. These models were trained on a single year of patient data and tested on the subsequent year.

The best performing model was able to predict high intensity service use following first referral to mental healthcare.

This project aims to expand on the previous work, to develop classification models using supervised machine learning to predict high service use at 12 months following first presentation to mental health services and evaluate if model performance remains accurate over time.

	Train (2007 – 2011) N = 12540	Test (2007 – 2011) N = 5803	Future (2012 – 2017) N = 16957
Age (IQR)	42 (29-55)	42 (29-57)	41 (28-55)
Female (%)	5845 (43.1)	2491 (42.9)	7262 (42.8)
Lives alone (%)	2044 (15.1)	889 (15.3)	2399 (14.1)
White (%)	9665 (71.3)	4114 (70.8)	3749 (57.4)
Black (%)	1720 (12.7)	719 (12.4)	1980 (11.6)
Asian (%)	816 (6.0)	389 (6.7)	1059 (6.2)
Risk (%)	730 (5.4)	309 (5.3)	2770 (16.3)
High intensity user (%)	1345 (9.9)	604 (10.4)	1723 (10.1)

Table 1. Patient demographics across the training, test and future validation cohorts. Data represented as median values with interquartile range (IQR) or as absolute counts and percentages.

Figure 1. Receiver operating characteristic (ROC) curves for model performance in training (left), test (middle) and future validation (right) cohorts. Models described include: Logistic regression (green); decision tree (purple); random forest (blue); and boosted ensemble (red)

	Train (2007 – 2011) N = 12540	Test (2007 – 2011) N = 5803	Future (2012 – 2017) N = 16957
Logistic Regression	0.80	0.80	0.82
Decision Tree	0.89	0.79	0.79
Random Forest	0.93	0.82	0.84
Boosted Ensemble	0.85	0.82	0.83

Table 2. Binary classification models were developed using logistic regression, decision tree, random forest and boosted ensemble methods. In- and out-of-sample performance are measured using AUC on the 2007-2011 cohort, along with predictive performance on a later validation cohort.

Methods

This project was carried out in the South London & Maudsley NHS Foundation Trust (SLaM). Electronic records have been in use since 2006 and anonymised patient data were extracted using Clinical Record Interactive Search (CRIS). We selected patients aged 18 years and over following their first presentation to SLaM services

Patient demographics, 90-day risk assessment, service use and symptom scores generated using natural language processing were used as predictor variables to determine high cost care, defined as cases receiving the top decile of cost.

Training data from 2007 – 2011 were used to develop logistic regression, decision tree, random forest and boosted ensemble models using cross validation. Model development was performed using the tidymodels package in R. Predictive models underwent hyperparameter tuning with the best performing specification used for the final model.

The model performance was evaluated using receiver operating characteristic (ROC) curves, calculating the area under the curve (AUC). AUCs were obtained for the training set, a test set from the same year and data from 2012 – 2017 for external validation.

Findings

36,300 patients were used to train, test and validate four predictive binary classification models (Table 1). These models were developed using logistic regression, decision tree, random forest and ensemble boosted supervised machine learning. Model performance was evaluated using AUC (Fig. 1, Table. 2), with similar performance across model types and between test and future validation cohorts.

Conclusion

- We demonstrate that routinely collected patient data held in EHRs can be used to train models to predict high intensity users of mental healthcare with similar performance to previous work.
- Four models were trained to identify high service use demonstrating similar performance across model types.
- Discriminatory performance remains constant between the test and future validation cohorts, suggesting good longitudinal model performance
- Future work is needed to integrate the use of predictive algorithms into clinical workflows

Kadra G et al. Predicting high-cost care in a mental health setting. BJPsych Open. 2020

The time is now: making the case for a UK registry of deployment of radiology artificial intelligence applications

M.E.W.M. Silkens • J. Ross • M. Hall • H. Scarbrough † • A. Rockall † • Show footnotes

DOI: https://doi.org/10.1016/j.crad.2022.09.132 • Check for updates

FESSH BSSH

Impact Factor: 2.206 / 5-Year Impact Factor: 2.490

Available access | Review article | First published online February 9, 2023

Insights and trends review: artificial intelligence in hand surgery

Robert Miller, Simon Farnebo, and Maxim D. Horwitz View all authors and affiliations

Volume 48, Issue 5 | https://doi.org/10.1177/17531934231152592

Artificial intelligence for automated radiomics for non-small cell lung cancer

Dr Jack Ross, Dr Mitch Chen, Prof Sue Copley, Prof Eric Aboagye, Prof Andrea Rockall, Imperial College London

Introduction

Aim: to develop a pipeline for automating objective and quantitative measures of lung cancer imaging for prognosis and prediction

Lung cancer exhibits phenotypic characteristics that can be captured on medical imaging. Some subtle signs that are indiscernible to the human eye can be analysed by machine learning. There is increasing evidence that these radiomic features can be used to help assess clinical outcomes including survival, histology, treatment response and adverse effects. However, these studies have been limited by the time-consuming image segmentation.

This project aimed to develop a pipeline for a fully automated deep learning-based lung cancer segmentation and radiomic feature reproducibility analysis.

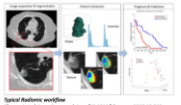
Radiomics

Radiomics is the high-throughput extraction of quantitative features from imaging to objectively and quantitatively describe tissues.

This compares to traditional subjective descriptions of lesions shape & structure ("spiculated, septated, round") as well as texture & intensity ("homogenous, uniformly enhancing"). These quantitative features are hopefully able to better predict patient outcomes such as prognosis and response to treatment. Machine learning techniques can also be used to quantify higher level features such as wavelet decomposition, that are harder for humans to identify but have been shown to correlate with outcomes.

Workflow

1. Patient identification and anonymisation
2. Tumour segmentation
3. Pre-processing, feature extraction and feature selection
4. Model validation

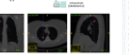


Typical Radiomic workflow
Thawani et al, Lung Cancer - doi.org/10.1016/j.lungcan.2017.10.015

Segmentation

The first step is to segment the lung cancer from the images. We have 957 CT scans across imaging cohorts have been manually segmented by two expert radiologists, as well as external datasets from e.g. The Cancer Imaging Archive.

For automated segmentation, we trained an nnU-Net model, a deep-learning medical image segmentation tool based on a U-Net architecture



We assessed nnU-Net automated performance stratified against a radiologist "difficulty assessment" for each tumour

Feature Extraction & Selection

A huge number of quantitative features can then be extracted using software such as PyRadiomics. This captures a range of features, from simple first order statistics and shape based measures to more complex features based on wavelet or Gaussian filters.

These features are statistically compared to actual patient outcomes to select the most important and representative features, and importantly exclude non-reproducible features.

Conclusion and Next Steps

We have successfully designed a fully automated pipeline to carry out the whole workflow for obtaining lung cancer radiomics.

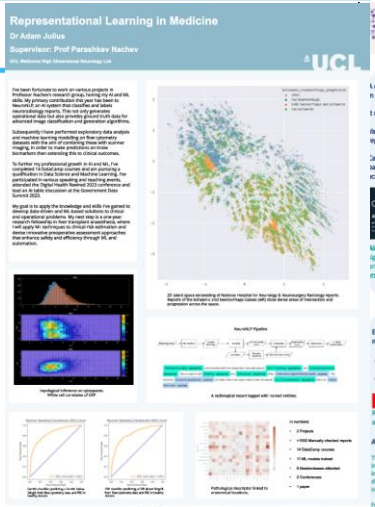
We are continuing to train the model, and validating the radiomic features against local and external data.

As a fully automated pathway, this can now be more readily extended for federated learning across multiple sites and datasets, using platforms such as FLIP (Federated Learning Interoperability Platform) and ADE (AI Deployment Engine).

Imperial College London | Imperial College Healthcare NHS Trust | Guy's and St Thomas' NHS | Digital Academy

Representational Learning in Medicine

Dr Adam Julius
Supervisor: Prof Parashkev Nachev
UCL Wellcome High Dimensional Neurology Lab



Acknowledgements
Dr Parashkev Nachev
Dr Robert Hobbs
Dr Dhiraj Dhagat
Prof Parashkev Nachev

Digital Academy | Guy's and St Thomas' NHS Foundation Trust

ACCEPT-AI

A stepped-wedge, cluster randomised study to assess the clinical effectiveness and acceptability of qER artificial intelligence software to prioritise CT head interpretation.
Dr Kavitha Vimalesvaran, Dr Sarim Ather, Dr Alex Nowak, Mr Haris Shuaib, Prof David Lowe

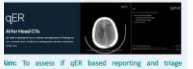
Introduction

I deep learning algorithm for detection of critical findings n head CT scan. CE Class IIa approved medical device.

t can detect nine critical findings on head CT scans.

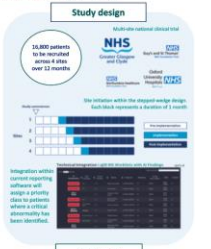
alidated on over 300,000 scans versus clinical radiology eports.

an potentially be used for triaging or notification of atients with critical findings as soon as a head CT is cquired.



Aim: To assess if qER based reporting and triage significantly reduced report turnaround time (TAT) of prioritised non-contrast head CT scans for patients attending the emergency department.

Study design



Rationale

Challenges

Foresight - Generative Pretrained Transformer (GPT) for Modelling of Patient Timelines using EHRs

Zeljko Kraljevic[1,6], Dan Bean[1,6], Anthony Shek[2,3], Rebecca Bendayan[1,6], Harry Hemingway[4,5,7], Joshua Au Yeung[2,3], Alexander Deng[3], Alfred Baston[3], Jack Ross[3], Esther Idowu[3], James T Teo[2,3] and Richard JB Dobson*[1,4,5,6,7]*

Front Digit Health. 2023 Apr 12;5:1161098. doi: 10.3389/fdgth.2023.1161098. eCollection 2023.

AI chatbots not yet ready for clinical use

Joshua Au Yeung [1 2], Zeljko Kraljevic [3], Akish Luintel [1], Alfred Balston [2], Esther Idowu [2], Richard J Dobson [3 4], James T Teo [1 2]

Affiliations + expand

PMID: 37122812 PMCID: PMC10130576 DOI: 10.3389/fdgth.2023.1161098

Free PMC article

Ann Surg. 2023 May 3. doi: 10.1097/SLA.0000000000005896. Online ahead of print.

A Surgical Perspective on Large Language Models

Robert Miller [1 2]

Affiliations + expand

PMID: 37132392 DOI: 10.1097/SLA.0000000000005896

NHS Digital Academy learning programmes

Testimonial: Prof Eric Topol

aicentre.co.uk/fellowships
ai4vbh@kcl.ac.uk
@ai4vbh

Dr Alexander T Deng BMBCh MA MSc MRCP DipRCPath

Programme Lead for *Fellowships in Clinical Artificial Intelligence*

GSTT Clinical Scientific Computing

Email: *alexander.deng@nhs.net*

Twitter: *@alextdeng*

INTELLIGENT HEALTH UK 2023

Breaking down the barriers between tech and healthcare