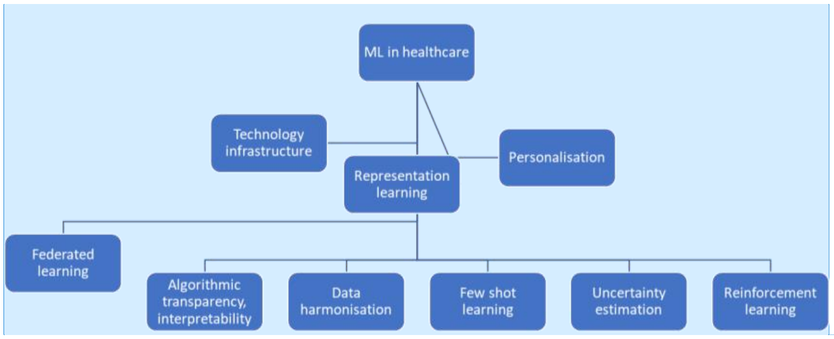
2.1 ML for healthcare: Where are we on the Pathway to Personalisation
Dr Dianelle Belgrave, DeepMind

- Heterogeneity
    - May affect individual's responses
- Axes of impact in AI for healthcare
    - Research
        - Scientific discovery
        - ML and healthcare research
        - ML to solve particular problem
    - Practice
        - Human-centered solutions
        - Human interaction to ML
    - Innovation
        - Real-World impact
        - Interdisciplinary
    - Grand challenge:
        - Translating research insights into actionable impacts
- Overview
    - ML for Healthcare at DeepMind
    - Discovery: Understanding heterogeneity as a pathway towards personalisation
    - Practice: Decision Support Tools for real-world impact
    - Innovation: Moving forward on the path towards personalisation
        - Fairness
        - Should be embedded in ML/AI healthcare application
- DeepMind
    - Solving intelligence to advance science and benefit humanity
    - Health @ DeepMind:
        - Part of DL team
        - Health portfolio
        - Two aspects:
            - Advancing science
            - Benefiting humanity
- Machine learning for Healthcare
    - Data types:
        - EHR
        - Imaging
        - Omics
        - Multimodal data
    - Generalize heterogeneity and provide fair AI models
    - Aim:
        - Scientific discovery: Understanding disease mechanisms for early detection and prevention
        - Disease treatment: predict health outcomes, and enable preventative, personalized care

        - Reinventing the future of health: Creating fairer AI for healthcare
- Machine Learning in Healthcare
    - ML techniques



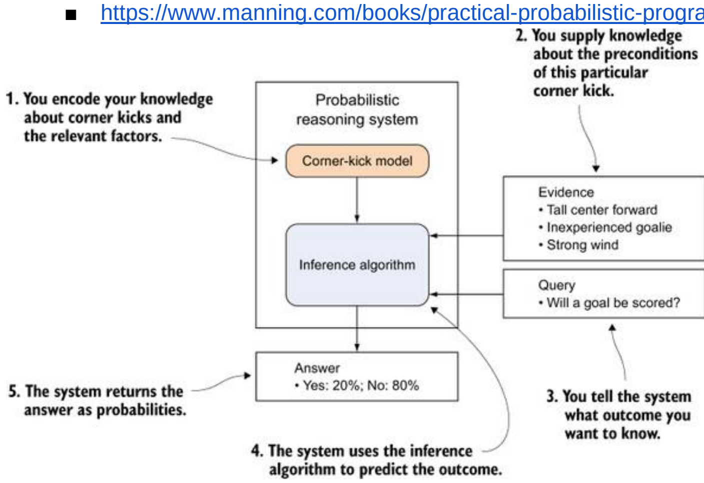
    - Healthcare modality
        - Imaging: MRI, Ultrasound
        - EHR: lab test results, clinical notes, ICU measurement
        - Omics: transcriptomics, metabolomics, genomics
        - Multimodal data: wearables, sensor data

Discovery: Understanding heterogeneity as a pathway towards personalisation

- Combine domain expertise + causality (heuristic approach) + ML (imitation, DL,) -> personalisation
- Not one size fits all:
    - Generalisability:
        - Disaggregate complex phenotype understanding causal mechanisms
        - Leads to better personalisation strategies
    - A deep understanding of patient heterogeneity can lead to better generalisability
    - Important step toward creating fairer ML for healthcare models and systems
- Healthcare grand challenge
    - Heterogeneity over treatment response and conditions manifest at time
    - Trade-off between risk reward benefits
    - I.e.
        - Patients: same diagnosis and same prescription – potential outcomes;
            - Drug not toxic and beneficial
            - Drug toxic but NOT beneficial
            - Drug toxic but beneficial
            - Drug NOT toxic and NOT beneficial
- Identifying disease endotypes
    - Endotypes: probabilistic latent variable frameworks
        - Natural history of condition sof time based on phenotypes/ different symptoms
        - Can we learn parsimonious descriptions of data in data

Commented [1]: Note the interactions captured are not accurate.

- Aim: to obtain parsimonious description of data (endotype) inferred from what is observed (phenotypes)
- Probabilistic programming
    - Latent variable: inferred
        - Commonly used in psychiatry
        - Based on some clinical features/ questions
        - https://www.manning.com/books/practical-probabilistic-programming



    - E.g. Causal mechanisms of asthma and allergy
        - https://www.jacionline.org/article/S0091-6749(10)01858-0/fulltext
        - Understanding the underlying mechanisms, better explaining clinical manifestation
        - Heterogeneity in response to treatment of asthma, eczema, and rhinitis
            - Progression of allergy: eczema -> asthma -> rhinitis
            - 12,000 children at different locations
        - 3 HMM:
            - https://research-information.bris.ac.uk/ws/portalfiles/portal/103820045/1_s2.0_S0091674916313458_main.pdf
            - Independent conditions
            - Allergic march
            - Independent conditions across time
                - Best model
                - Structural independence over time and over condition
        - Combination of symptoms and not mutually exclusive
            - Development profiles are heterogenous
            - Allergic March

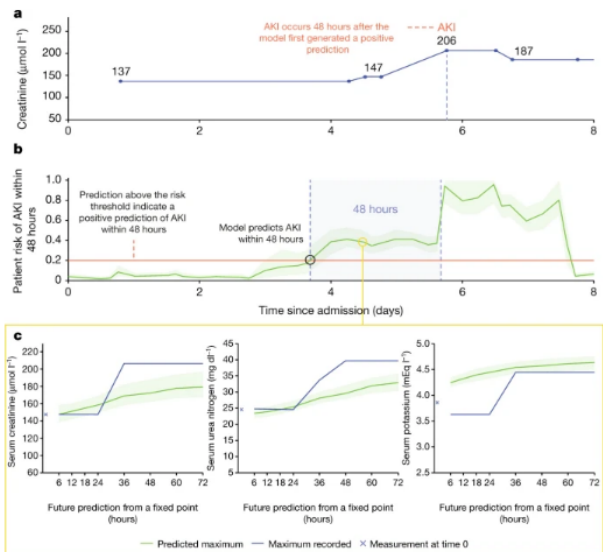
Practice:

- Decision Tools on the Path to Personalisation

    - Clinical decision support tools can act as early-warning systems for more optimal personalised interventions
    - Personalised predictive models can help to define individual risk of AE
    - Important step towards creating fairer ML for Healthcare Models and systems
- EHR and AKI
- AKI: sudden drop in kidney function and creatinine in blood
    - Associated with poor outcomes
    - AKI problems before 2009:
        - https://www.ncepod.org.uk/2009aki.html
        - Delays in recognition
        - Poor management once recognized
        - Delays in access to specialist care
    - National Patient Safety Alert;
        - https://www.england.nhs.uk/akiprogramme/aki-algorithm/
        - Diagnostic AKI algorithm
        - Flagging AKI at clinical records
        - Population of national registry of AKI patients
- Continuous prediction of acute kidney injury (Nature 2019)
    - https://www.nature.com/articles/s41586-019-1390-1
    - RNN for continuous prediction of AKI up to 48 hrs using EHR data
    - End-to-end training with 7 auxiliary tasks (electrolyte lab values)
    - State-of-art performance and calibration
        - Prediction of 55.8% of all inpatient episodes of AKI
        - 90.2% of all AKIs that required subsequent dialysis
        - Ratio of 2 FPs for every true alert
    - Generalisability experiments showing model fairly robust to site/time
    - Limitation:
        - Sig. no of false positive
        - Demographically skewed
        - Ground truth label was not perfect
        - Not yet prospectively validate
- Feature representation
    - No imputation of missing values
    - 460k presence features encode missingness
    - 160k numerical feature normalized to unit range after capping at 1st and 99th percentiles
    - Age and time of day provided to model
    - Provide median yearly Cr baseline + min 48h baseline
    - Diagnosis pushed to end of admission to avoid information leakage
    - Aggregations of recent values included for baseline models
- Regularisation
    - L1 regularization on first layer of embedding module
    - Multi-task
        - Multiple windows for AKI

- Multiple windows for each lab auxiliary task
    - Different adjustment
        - Dropout
        - Autoencoders/ VAEs
        - L2 throughout
    - RNN outperforms non-RNN models
        - By ROC AUC
        - Predict AKI 24hrs before admission
    - LImitation:
        - False positive reads for 1 to 2
            - Acceptable by clinical experts
- With different stages of AKI, model still with outperforms than actual task
- Example success case:

Fig. 1: Illustrative example of risk prediction, uncertainty and predicted future laboratory values.



The first 8 days of admission for a male patient aged 65 with a history of chronic obstructive pulmonary disease. **a**, Patient creatinine measurements during admission. Creatinine measurements, showing AKI occurring on day 5. **b**, Model predictions for any AKI within 48 h. Continuous risk predictions: the model predicted increased AKI risk 48 h before it was observed. A risk above 0.2 (corresponding to 33% precision) was the threshold above which AKI was predicted. Lighter green borders on the risk curve indicate uncertainty, taken as the range of 100 ensemble predictions (after these were trimmed for the highest and lowest 5 values). **c**, Laboratory value predictions 4.5 days into admission. Predictions of the maximum future observed values of creatinine, urea and potassium.

- False positive rates:
    - 49%

- AKIs that go on requiring dialysis: 90%
    - Impact on workflow
    - False positive: most had events later on
    - Maybe a good intuition of later deterioration of kidney function
- Personalisation and Generalisability
    - **Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records**
        - https://www.nature.com/articles/s41596-021-00513-5
        - Utility of the pipeline across a number of different endpoints
        - Utility of pipeline on additional/different clinical data
        - Original models generalized well across time and hospital
        - Inconsistencies in feature representations meant it wasn't possible to generalize across EHR system boundaries
    - **Multitask prediction of organ dysfunction in the intensive care unit using sequential subnetwork routing**
        - https://academic.oup.com/jamia/article/28/9/1936/6307184
        - MTL has shown promise in improving model performance and training efficiency
            - Still suffers from negative transfer
        - Sequential subnetwork routing (SNR)
            - Soft parameter sharing to find related tasks and encouraging cross-learning
        - Improvements in label efficiency

Innovation: Moving forward on path to personalisation

- Personalisation: part of a pipeline
    - Centered to causal modeling frameworks to determine risk
        - Data-driven approach to stratifying interventions
        - Data-driven approach to stratifying underlying causal mechanism
    - Hypothesis generating:
        - Modeling frameworks to disentangle heterogeneity
        - Robust study design
        - Domain expertise
    - Hypothesis testing
        - Basic science to understand biological mechanisms and plausibility
        - Develop mechanism-based actionable interventions

- Creating fair systems for personalisation
    - https://informatics.bmj.com/content/27/3/e100237.long

Table 1 Areas of emphasis for ensuring machine learning for healthcare (MLHC) works for all	
Area of emphasis	**Recommendations**
Ensure MLHC is equitable by design	▶ Develop pipelines for the promotion of diverse teams in all aspects of MLHC ▶ Ensure the inclusion of data from a broad range of groups, in a broad range of contexts ▶ Incorporate global partners to ensure health data science promotes global health equity.
Encourage public and open MLHC research	▶ Fund both direct MLHC research and research into ethical aspects of MLHC ▶ Harmonise ethical oversight between public and private research domains
Ensure adequate access to health information technology (IT) infrastructure	▶ Ensure all are included in the datasets by funding health data gathering infrastructure in underserved communities ▶ Develop MLHC products with an awareness of the broad range of health IT contexts for deployment
Ensure MLHC is clinically effective and impactful	▶ Ensure the presence of multidisciplinary teams that represent both clinical and data science perspectives ▶ Promote pathways for interdisciplinary training ▶ Hold MLHC innovations to the same standards as other healthcare interventions, including requirements for prospective validation and clear demonstration of impact
Audit MLHC on ethical metrics	▶ Mandate assessments of the performance of novel MLHC technology for impacts on marginalised and intersectional groups. ▶ Record the data necessary to perform these audits in an ongoing fashion
Mandate transparency in data collection, analysis and usage	▶ Build patient trust by ensuring that protocols for the collection, analysis and usage of data are transparent and open
Promote inclusive and interoperable data policy	▶ Ensure the existence of clear and ethical methods for ensuring the sharing of data between different sources while protecting patient rights and privacy ▶ Improve the standardisation of medical data generation and labelling across contexts ▶ Ensure that global partners are included, so that interoperability barriers do not hinder inclusive global collaboration

- Context matters
    - Solutions specific to problems
    - Merging different schools of thought for bigger picture
    - Data driven approach + Domain Knowledge = Problem-led approach with patient at the center