Case Studies

Scenario 2: AI System for Predicting Heart Disease Risk in the UK

Background: AI system is planned for predicting the risk of developing heart disease for people in the UK. The system will be used to identify high-risk individuals to whom lifestyle changes can be suggested to. Currently, such models are based upon simple characteristics such as diet, age, BMI and lifestyle. The AI system will attempt to exploit a wider range of data, such as ECG data and imaging data, for making more personalised risk assessments. The intention is for the system to automatically segment the left ventricle from cardiac MR images and use the segmentations to compute a range of functional metrics; these metrics will then be linked with outcome data to train a machine learning model to predict the risk of developing heart disease. The developers will also investigate incorporating non-imaging data such as BMI and age as extra input data to the system.

Development and Training: The data to train the AI system will come from a network of 3 London NHS hospitals, with a fourth NHS hospital's data held out for testing the system. Access to electronic health records of all patients is also possible.

Questions to consider:

- What dangers/risks of the use of AI for this problem can you identify at this stage?
- How would you go about addressing these?
- What fairness metric(s) do you think might be appropriate when assessing the AI tool for potential bias?

REMINDER - DEFINITIONS OF FAIRNESS

- *False Negative Rate (FNR): the rate at which positive cases are missed by the classifier*
- *Demographic parity - equal chance of being classified positive for each protected group*
- *Equalised odds - equal true positive rate (TPR) & false positive rate (FPR) for each protected group*
- *Equal opportunity - only equalise either FPR or FNR, not both*

Suggested answers/discussion points:

- What dangers/risks of the use of AI for this problem can you identify at this stage?
    - There is a risk of demographic bias as racial bias has been reported in cardiac MR image segmentation. The training data will come from London – this is quite racially diverse so this could be a good thing, but it depends on where in London these hospitals are as some parts of London are not so diversified.
    - The MR imaging data will presumably come from a range of scanners so the model should be robust to such domain shifts although this should be checked.
    - We should be careful about including any non-imaging data without thinking about it carefully – e.g. BMI is arguably a biased metric since the normal ranges of BMI were calculated from white males (see: https://www.washingtonpost.com/lifestyle/wellness/healthy-bmi-obesity-race-/2021/05/04/655390f0-ad0d-11eb-acd3-24b44a57093a_story.html).
- How would you go about addressing these?
    - We have access to electronic health records which should store race, age and sex information. We should make sure these are recorded for the purpose of bias assessment.
    - Thoroughly test performance of the tool on different demographic groups, e.g. races, sexes.
    - Fairness metrics should be calculated for different protected groups such as those based on race, sex, age and even pathology.
    - Investigate whether the fairness metrics employed generalize to other domains, e.g. data from other held-out test hospitals
- What fairness metric(s) do you think might be appropriate when assessing the AI tool for potential bias?
    - There are known differences in the prevalence of some cardiovascular diseases between sexes and races, so if a classification model is to be developed demographic parity might not be appropriate.
    - False negatives are probably more important to control than false positives – FPs may lead to unnecessary lifestyle adjustments but FNs could lead to more disease – so perhaps Equal Opportunity based on equalizing FNs could be appropriate?

Conclusion: This case study underscores the importance of addressing data imbalance and ensuring fairness in AI systems for medical predictions. By diversifying the training dataset, incorporating socio-economic data, applying bias mitigation techniques, and using appropriate fairness metrics such as demographic parity, equal opportunity, and calibration, the AI tool can be developed to provide accurate and equitable risk predictions for heart disease across the UK. This approach will help improve health outcomes, promote personalized healthcare, and reduce disparities in medical care.