Personalising education for radiologists using AI: a breast imaging case study

Ziba Gandomkar, Mo'ayyad Suleiman, Mary Rickard, Catherine Jones, Patrick C Brennan

Discipline of Medical Imaging Sciences, Faculty of Medicine and Health, University of Sydney, NSW 2006, Australia.
DetectedX, Merewether Building, The University of Sydney, NSW 2006, Australia.

Introduction

- ✓ Case-based learning strategies which involve extensive exposure to normal and abnormal image appearances as well as explicit performance feedback on image interpretation or technical evaluation abilities, are an effective approach to acquire and improve fundamental skills a radiologist.
- ✓ Current platforms for case-based learning do not consider the well-proven fact that individual learners make individual-specific errors and simply provide the same solution for all personnel at a particular level of training.
- ✓ With this generalised one-size-fits-all approach, some clinicians can demonstrate improvements of 50% after using case-based learning platforms, while for some, improvement levels remain below 10%.
- ✓ To tackle this problem, Artificial Intelligence (AI) might help.
- ✓ The disruptive educational impact of AI is clearly advancing teacher-student interactions, however tailoring educational delivery for each individual, particularly in health and medicine, has not progressed.
- ✓ Some attempts have been made to predict radiologists' errors to retrieve difficult cases for radiologists [1-3].
- ✓ A common approach: computer-extracted features from mammograms were fed into a personalised AI model to categorise images as false-positives, true-negatives or false-negative or true-positives [3].

Purpose

Breast self-assessment modules (test sets) have been used to educate radiologists for decades, however the one-size-fits all approach commonly used does not confer benefits uniformly to all users. This work looks at developing artificial intelligence (AI) engines around personalizing test sets in three situations:

- ✓ New users who have never engaged
- ✓ New material that no-one has engaged with
- ✓ Existing users who have been engaging with the platform

Materials and methods

The first engine: AI for predicting performance of the new users who have never engaged

- The first engine focused on new users
- who have not completed any test sets.
- 905 individuals completed a questionnaire around demographics, reading volume and years of experience.
- A Pearson correlation coefficient between the predicted values by the model and the actual users' AUCs was performed.

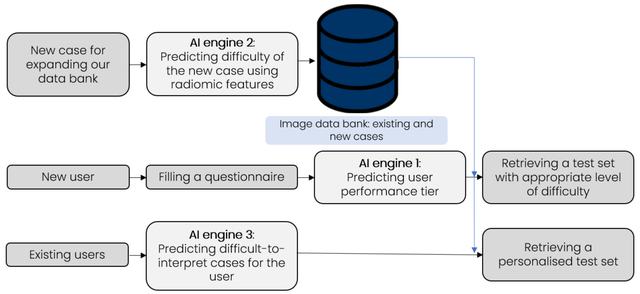


Figure 1- The proposed AI engines

The second engine: AI for categorising new materials

- The second engine focused on predicting the difficulty of cases, newly uploaded on to the platform and relied on global and local radiomic features to estimate difficulty.
- An independent dataset of 480 mammographic examinations that were eventually read by 46 to 352 radiologists was used
- The AUC calculated for discriminating the easy-to-interpret from the difficult-to-interpret cases.

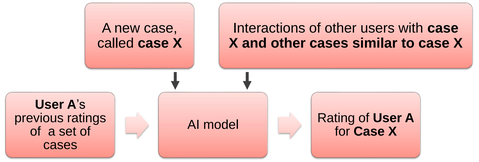


Figure 2- The third AI engine for predicting difficult-to-interpret cases for existing users

The third engine: AI engine for predicting difficult-to-interpret cases for existing users

- The third engine using matrix factorization to build a model to predict the rating provided by existing reader to a case based on the reader's previous interactions with other cases as well as other readers' interaction data.
- 260 radiologists, 12 breast physicians, and 207 radiology trainees were involved.

Results

The first engine: AI for predicting performance of the new users who have never engaged

- The Pearson correlation coefficient between the predicted AUC by the first engine and actual AUC was 0.60 ($p < 0.001$).
- The AI model's ability to differentiate first and fourth quartile readers based on AUC values was 0.86 (95% Confidence Interval of 0.83–0.89).

The second engine: AI for categorising new materials

The AUCs produced by the second model for predicting difficult-to-interpret cancer and normal cases for new material were 0.87 ± 0.21 and 0.70 ± 0.29, respectively.

The third engine: AI engine for predicting difficult-to-interpret cases for existing users

The model for predicting difficult from easy cases for existing users showed an AUC value of 0.81 (95% CI: 0.74-0.88).

Conclusions

This work focused on breast imaging shows how tailored AI approaches can be used to optimize radiology education in different situations. Whilst the focus here is mammography, the approaches employed should have value for other areas of medical imaging.

For further information: Please contact ziba.gandomkar@sydney.edu.au; patrick.brennan@sydney.edu.au

References

[1] Grimm et. al, Academic radiology, 2014.
[2] Mazurowski et. al, Academic radiology, 2012.
[3] Zhang et. al, J biomedical informatics, 2015.