Optimising a real-world queueing problem: triaging MRI brain scans using imperfect artificial intelligence models

Zezheng Lyu[1], Siddharth Agarwal[1], Yiran Wei[1], Marc Modat[1], Thomas Booth[1,2]

1. School of Biomedical Engineering & Imaging Sciences, King's College London
2. Department of Neuroradiology, King's College Hospital NHS Foundation Trust

Introduction

Motivation: Despite the introduction of AI tools in the NHS, most hospitals still use a **first-come first-served** queue for radiology lists. However, AI classifier models can optimise this queue, enabling abnormal scans to be interpreted sooner[1]. The challenge lies in the imperfect nature of these AI models; a naive implementation risks causing unacceptable delays in reporting false negatives.

Aim: To develop an adjustable queuing system for **outpatient MRI brain scans** that allows clinicians to set their preferred balance between AI-driven prioritisation and traditional queue order. This system will accommodate varying tolerance levels for potential delays in reporting false negatives.

Methods

Simulation of the data

- Prevalence of abnormality
- AI classifier performance (sensitivity, specificity)

Simulation → Ratios of TP, FP, TN, FN

- Acquisition duration
- Mean reporting time for normal/abnormal scans

Simulation → Random acquisition order

We developed three triage strategies and conducted simulation experiments. Each strategy allows for adjustment of tolerance levels for delays in false negative reporting through tuneable hyperparameters.

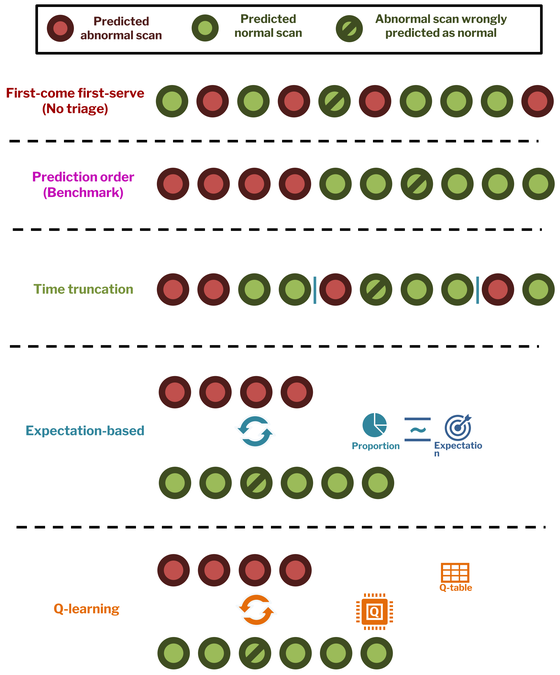
- **Prediction order (benchmark):** Always prioritises predicted abnormal scans

1. **Time truncation:** Only begins reading new scans after all scans acquired within a preset period are reported. Prioritises predicted abnormal scans within this period.
   Tuneable hyperparameter: truncation time
2. **Expectation-based:** Separates the queue into predicted abnormal and normal queues. Arranges the radiologist to switch between queues to meet preset expectations for the proportion of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) in reported scans. Requires the ground truth (normal/abnormal) to be recorded per scan.
   Tuneable hyperparameter: coefficient of expectation threshold

Methods

3. **Q-learning:** Builds upon **2** –Train Q-table with customisable reward function based on simulated data. Applicable for both binary and multi-class queuing. Requires the ground truth (normal/abnormal) to be recorded per scan.
   State: Proportion of actual-*i*, predicted-*j* in reported scans; Duration of leaving queue *i;* Reported fraction in queue *i.*
   Actions: Move to queue *i.*
   Tuneable hyperparameter: a parameter in preset reward function / new reward function

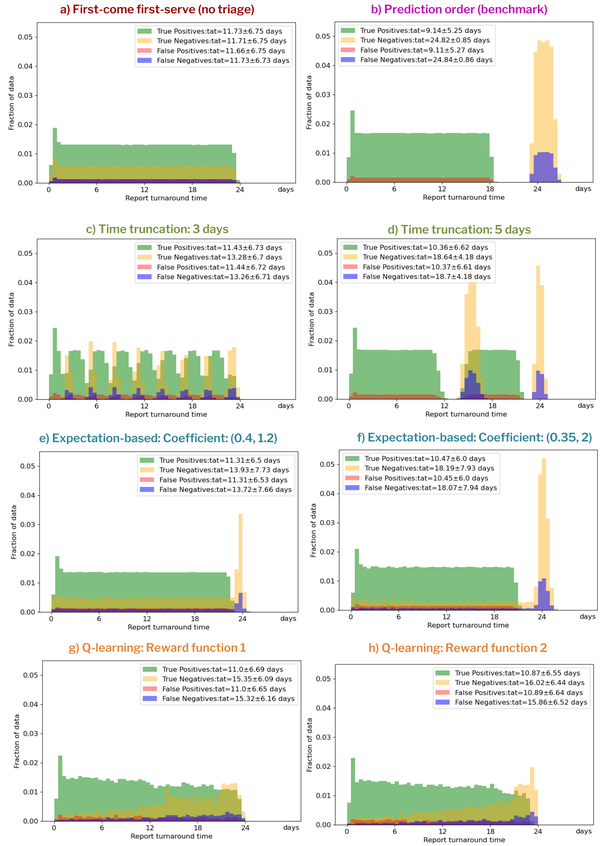
Figure 1. Schematic of triage strategy for binary class



Results

An example for binary class triage - the following table and graphs are the average result of 1,000 simulations using a prevalence of abnormality: 63%, AI sensitivity: 91%, AI specificity: 83%.[1]

Figure 2. TAT for each queueing system for binary class triage example.



Results

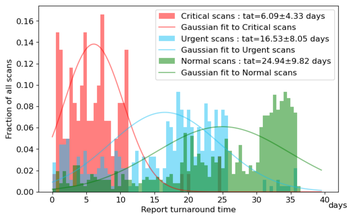
Table 1. Average TAT for each queueing system for binary class triage example.

Report turnaround time (TAT)		True positive (days)	False negative (days)
First-come first-serve (no triage)		11.73 ± 6.75	11.73± 6.73
Prediction order (benchmark)		9.14 ± 5.25	24.84 ± 0.86
Time truncation	Truncation time:3 days	11.43± 6.73	13.26 ± 6.71
	Truncation time:5 days	10.36 ± 6.62	18.70 ± 4.18
Expectation -based	Coefficient: (0.40, 1.2)	11.31 ± 6.50	13.72 ± 7.66
	Coefficient: (0.35, 2.0)	10.47 ± 6.00	18.07 ± 7.94
Reinforcement learning	Reward function 1	11.00 ± 6.69	15.32 ± 6.16
	Reward function 2	10.87 ± 6.55	15.86 ± 6.52

An example for three-class triage using Q-learning approach

Urgency level:
Critical > Urgent > Normal

Figure 3. An example for TAT for three-class triage



Conclusion

Using imperfect AI classifiers for MRI brain scan triage balances faster reporting of true positives against potential delays for false negatives. We developed three adjustable methods to address varying risk tolerances. While these methods converge to similar results, they offer different implementation trade-offs.

References

1. D. A. Wood et al., 'Deep learning models for triaging hospital head MRI examinations', Med. Image Anal., vol. 78, p. 102391, May 2022, doi: 10.1016/j.media.2022.102391.