How can AI be used to minimise medical errors?

Exploratory analysis using a proposed AI augmented SHERPA workflow

Ferenc Kohsuke Gutai
University of Dundee
Word count: 1500 words (excluding cover page and references)
Introduction

“I’ve made a mistake”
Four words that no healthcare professional wants to hear, let alone utter. With an approximate global incidence of 10%(1,2), medical errors constitute a non-negligible proportion of adverse medical outcomes and carry a significant burden of cost to healthcare systems and patients(3). Four main factors behind medical errors are: the complex nature of medical procedures and equipment, the multi-professional/multi-component nature of healthcare, the ease at which errors can propagate along the chain and the difficulty of human beings to viably predict downstream consequences within the system(4). The pressing need for reducing medical errors, coupled with the intricate human involvement demands an extra-human solution – enter Artificial Intelligence (AI). Much has been written on the use of AI clinical decision support systems in reducing medical error(5), but its role in error analysis itself has yet to be elucidated clearly. This review will explore how AI could be pivotal in error reduction strategies, using a specific analysis paradigm as a case study.

AI

AI in healthcare is useful in speeding up data collection and analysis(6) and exists in 3 forms: classical machine learning (ML), natural language processing (NLP) and deep learning(7)(Figure 1). ML algorithms include techniques to classify, predict and determine behavior of target phenomena(8) and come in 3 types: supervised, which requires prelabeled data; unsupervised, which does not require any pre-labeling; and reinforcement learning, which allows an agent to alter its behavior to achieve a specific goal(8). NLP comprises text processing and classification, using keywords to extract meaningful information from written language(7). Deep learning includes the field of computer vision (CV), which allows AI to identify objects and patterns in images/video(9). An important principle is the concept of under- and overfitting(8). A simplified model with high bias can become too generalized, failing to capture meaningful patterns in the data (underfit). Overcomplicate things, however, and the model can become too specific, overfitting the training dataset and failing to adapt to new data.

![Figure 1 – Overview of the fields of AI](image-url)
**SHERPA**

Systematic Human Error Reduction and Prediction Approach (SHERPA) is a tool that has been used in both medical and non-medical fields to predict and prevent errors (10,11). Workflow begins with hierarchical task analysis (HTA) followed by a rigorous checklist which queries: (a) what errors can occur, (b) the likelihood of such errors occurring, (c) the severity of the consequences, (d) how one could remedy the situation and (e) future prevention plans (3). The results of SHERPA can then be fed back to influence process design and policy. Crucially, the HTA and error prediction stages encourage professionals to think more deeply about tasks that may be unconsciously ingrained in their practice, thus exposing critical ‘danger points’ in the procedure (3).

**HTA**

Firstly, the target procedure is deconstructed into well-defined tasks (Figure 2). This is done with input from subject matter experts (SMEs). Although most medical procedures are not temporally challenging to fathom, the majority of information within each step is contained visually, thus representing a difficult but potential avenue for CV to step in. Recent studies have demonstrated the efficacy of CV in measuring surgical skill during tying and suturing tasks (12) as well as video segmentation of key steps in laparoscopic procedures (13). This could change the game for HTA – AI ‘critics’ could scour through clinical footage and extract not only task classification but also operator skill, feeding into the next step in SHERPA of error prediction. Such developments are not without their caveats: CV is still dependent on human input through annotation of training data (13), and patient outcomes are unobserved (12), making the output less relevant for consequence analysis. Arguably, the latter is nullified by a longitudinal analytical toolbox such as SHERPA which juxtaposes HTA alongside other methods of error analysis. Lastly, large amounts of video data would be required to train such algorithms, necessitating changes in infrastructure, workflow and data access policies (14). Although the technology is in its early days, the potential benefits of such a system once it is up and running would outweigh the short-term costs.

![Figure 2 – HTA applied to medical diagnosis (15)](image-url)
Error generation and classification

Once the procedure has been dissected into subtasks, SMEs must go through each node and determine if and what potential errors might occur there. The reversed root cause analysis (RCA) that would be demanded in this phase requires an intimate understanding of complex systems and how they could fail, a capability that challenges current AI systems. Interestingly, the use of fuzzy cognitive maps, a specific type of AI modelling framework, has been demonstrated in carrying out RCA for an albeit simplified case(16)(Figure 3). Although this is still the inverse of what is required, such maps could be reverse engineered to link specific tasks with errors. There is still much work to be done in terms of associating causative factors to actual tasks, however these are promising results.

![Figure 3 – Fuzzy cognitive map used to diagnose the causes of a stalled car(16)](image)

Likelihood and severity prediction

Once the errors have been generated, predictions are then made regarding the probabilities that such errors would occur and their severities. This is achieved by one of 2 ways, either through SMEs providing anecdotal experience or by trawling the literature and creating statistical likelihoods and severity indexes – ceteris paribus, the former is much easier to implement but the latter is more reliable. The variance in standards of reporting probabilities and likelihoods makes literature review problematic(17). Database crawling NLP ‘bots’ could be used to generate preliminary results based on sorting keywords for subsequent analysis by researchers, cutting short the review time(6). The shift in NLP away from syntactic and towards narrative-based paradigms(18), would allow for the assessment of contextual information to produce standardised scores e.g. on a scale of 1 to 10, ‘lifelong hemiparesis’ might be judged as a 9 compared to ‘death’ being a 10. Any medical professional, however, would understand the importance that quality of life (QoL) plays in adverse outcomes and how it affects these results(19). Unfortunately, patient-reported outcome measures, good indicators of QoL, are poorly presented in research(20). Some
might argue that it is still better to let algorithms decide relative ratings. Whatever one’s stance on the matter, it is good to have a choice of an alternative option based on AI.

**Remedy steps**

Once errors have been explored, SHERPA then calls for solutions. This could be as simple as documenting cannula insertion if one has forgotten to do so, or as complex as suturing an artery that has been torn intraoperatively. NLP could once again be used to augment the process, searching databases of adverse incidents/events and generating solutions that have been tried before, although this would require the existence of a database in the first place. Such a repository would be feasible to curate as the healthcare system begins to plug into the Internet of Things (IoT) paradigm(21)(Figure 4). Loggers implemented in various medical devices could connect to central information stores and be utilised to piece together an informatic story when an adverse event occurs. IoT is still in its infancy and comes with its own set of issues: increasing complexity of the healthcare system could lead to more errors(22) and its Orwellian nature might hamper adoption, but here it would be argued that clear rules regarding data security and privacy would be enough to garner support for the technology(23). Regardless, the story echoes current issues with regards to human factors analysis of medical errors: more data is needed for AI to play with.

![Figure 4 – Simplified map of the IoT paradigm(21)](image)

**Error prevention**

The last step of SHERPA entails a thorough evaluation as to what strategies could be implemented to prevent the occurrence of the identified errors. This is perhaps the most cognitively demanding part of the process as the researcher is required to think laterally, a task that is still beyond the reach of current AI(24). For example, consider an error in which a doctor forgets to assemble a vital piece of kit in a procedure. Even a simple, low-cost solution that has proven to be effective, such as putting up a reminder poster(25,26), would be beyond current AI capabilities. As such, it is difficult to theorise a situation whereby AI could assist in this phase of SHERPA, although this difficulty might be dispelled in the future with advances in the field.
Conclusion

Through this review, we have illustrated a realistic case study of how AI could be used to minimise medical errors by augmenting and in some places replacing human involvement in carrying out the SHERPA process. Clearly some human help is still necessary, and it is important to remember how this could be a source of error in terms of design and methodology flaws. This should not be taken as a failure, however, but a cause for celebration: reducing demand in laborious tasks such as HTA and likelihood/severity prediction processes can redirect human cognitive energy towards more important issues such as remedy and mitigation strategies. AI would be a powerful behind-the-scenes ally – after all, prevention is better than cure.
References


25. Yong YP, Karangizi A, Banerjea A. A persuasive intervention: improving the