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.

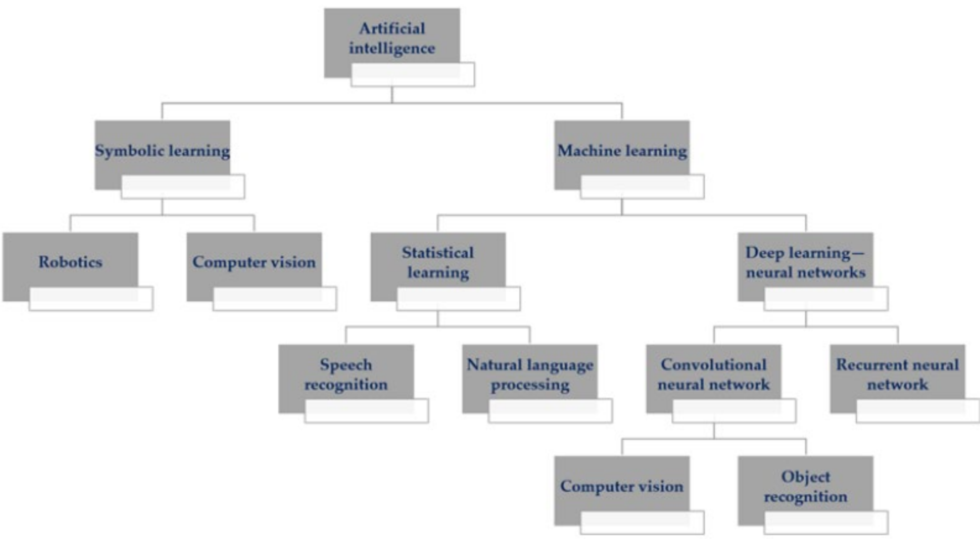

Figure 1 – Overview of the fields of AI(9)

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.

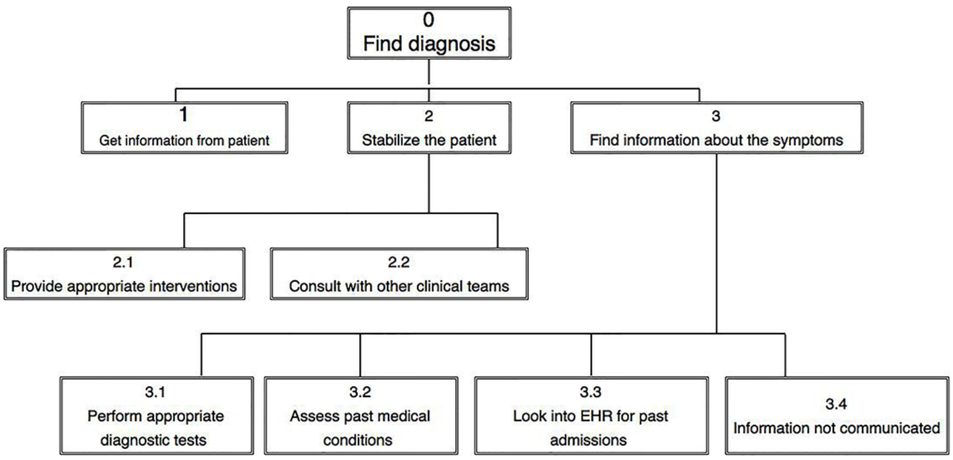

Figure 2 – HTA applied to medical diagnosis(15)

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.

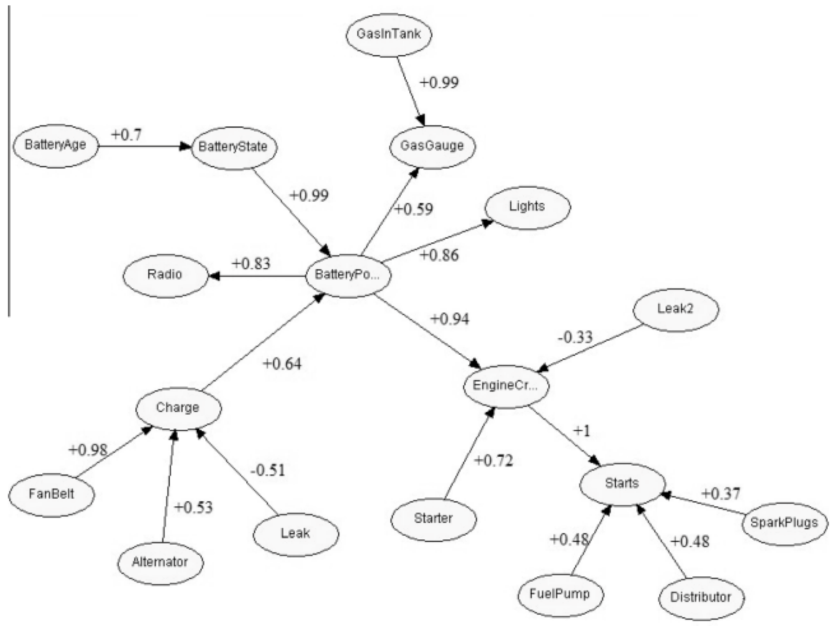

Figure 3 – Fuzzy cognitive map used to diagnose the causes of a stalled car(16)

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.

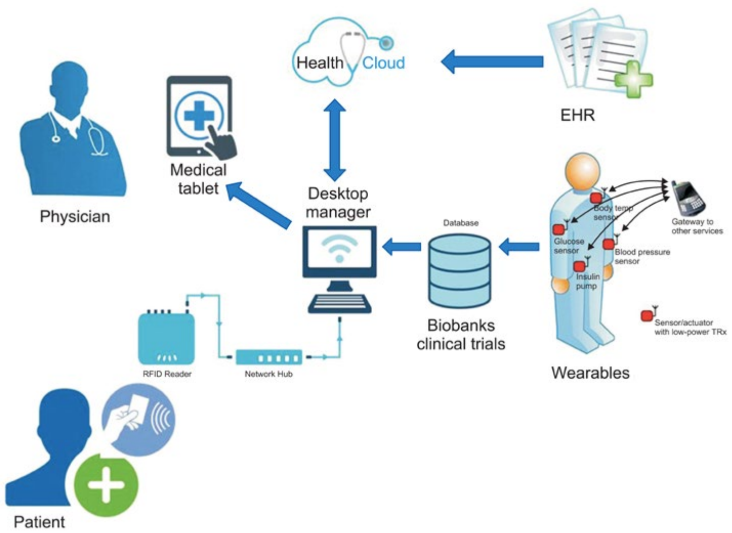

Figure 4 – Simplified map of the IoT paradigm(21)

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

1. Taib IA, McIntosh AS, Caponecchia C, Baysari MT. A review of medical error taxonomies: A human factors perspective. Saf Sci [Internet]. 2011 Jun;49(5):607–15. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0925753510002924
2. Eslami K, Aletayeb F, Aletayeb SMH, Kouti L, Hardani AK. Identifying medication errors in neonatal intensive care units: a two-center study. BMC Pediatr [Internet]. 2019 Dec 22 [cited 2019 Dec 28];19(1):365. Available from: https://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-019-1748-4
3. Kermani A, Mazloumi A, Kazemi Z. Identification of Nurses' Errors in the Emergency Ward, Using SHERPA Technique. Int J Occup Hyg [Internet]. 2016 [cited 2019 Dec 28];8(1). Available from: http://ijoh.tums.ac.ir/index.php/ijoh/article/view/208
4. Christoffersen K, Woods DD. How complex human-machine systems fail: Putting "human error" in context. In: Karwowski W, Marraas WS, editors. The Occupational Ergonomics Handbook. Boca, Raton: CRC Press; 2003.
5. Aljaaf AJ, Al-Jumeily D, Hussain AJ, Fergus P, Al-Jumaily M, Abdel-Aziz K. Toward an optimal use of artificial intelligence techniques within a clinical decision support system. In: 2015 Science and Information Conference (SAI) [Internet]. IEEE; 2015 [cited 2019 Dec 28]. p. 548–54. Available from: http://ieeexplore.ieee.org/document/7237196/
6. Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform [Internet]. 2019 Dec 1 [cited 2019 Dec 30];132:103971. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1386505619302370
7. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol [Internet]. 2017 Dec 1 [cited 2019 Dec 28];2(4):230–43. Available from: http://svn.bmj.com/cgi/doi/10.1136/svn-2017-000101
8. Senders JT, Arnaout O, Karhade A V., Dasenbrock HH, Gormley WB, Broekman ML, et al. Natural and Artificial Intelligence in Neurosurgery: A Systematic Review. Neurosurgery [Internet]. 2018 Aug 1;83(2):181–92. Available from: https://academic.oup.com/neurosurgery/article/83/2/181/4107737
9. Ellahham S, Ellahham N. Use of Artificial Intelligence for Improving Patient Flow and Healthcare Delivery [Internet]. Article in Journal of Computer Science & Systems Biology. 2019 [cited 2019 Dec 28]. Available from: https://www.researchgate.net/publication/336742991
10. Ghasemi M, Khoshakhlagh AH, Mahmudi S, Fesharaki MG. Identification and assessment of medical errors in the triage area of an educational hospital using the SHERPA technique in Iran. Int J Occup Saf Ergon [Internet]. 2015 Jul 3 [cited 2019 Dec 30];21(3):382–90. Available from: http://www.tandfonline.com/doi/full/10.1080/10803548.2015.1073431
11. Habibi E, Gharib S, Mohammadfam I, Rismanchian M. Human error assessment in Isfahan oil refinery's work station operators using systematic human error reduction prediction approach technique. Int J Environ Health Eng [Internet]. 2013 [cited 2019 Dec 30];2(1):25. Available from: http://www.ijehe.org/text.asp?2013/2/1/25/113214
12. Azari DP, Frasier LL, Quamme SRP, Greenberg CC, Pugh CM, Greenberg JA, et al. Modeling Surgical Technical Skill Using Expert Assessment for Automated Computer

Rating. Ann Surg [Internet]. 2019 Mar 1 [cited 2019 Dec 30];269(3):574–81. Available from: http://insights.ovid.com/crossref?an=00000658-201903000-00028
13. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, et al. Computer Vision Analysis of Intraoperative Video. Ann Surg [Internet]. 2019 Sep 1 [cited 2019 Dec 30];270(3):414–21. Available from: http://insights.ovid.com/crossref?an=00000658-201909000-00003
14. Prebay ZJ, Peabody JO, Miller DC, Ghani KR. Video review for measuring and improving skill in urological surgery. Nat Rev Urol [Internet]. 2019 Apr 8 [cited 2019 Dec 30];16(4):261–7. Available from: http://www.nature.com/articles/s41585-018-0138-2
15. Roosan D, Law A V, Karim M, Roosan M. Improving Team-Based Decision Making Using Data Analytics and Informatics: Protocol for a Collaborative Decision Support Design. JMIR Res Protoc [Internet]. 2019 Nov 27 [cited 2019 Dec 30];8(11):e16047. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31774412
16. Wee YY, Cheah WP, Tan SC, Wee K. A method for root cause analysis with a Bayesian belief network and fuzzy cognitive map. Expert Syst Appl [Internet]. 2015 Jan [cited 2019 Dec 30];42(1):468–87. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0957417414003789
17. Wears RL, Janiak B, Moorhead JC, Kellermann AL, Yeh CS, Rice MM, et al. Human error in medicine: Promise and pitfalls, part 1. Ann Emerg Med [Internet]. 2000 Jul;36(1):58–60. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0196064400697123
18. Cambria E, White B. Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. IEEE Comput Intell Mag [Internet]. 2014 May [cited 2019 Dec 30];9(2):48–57. Available from: http://ieeexplore.ieee.org/document/6786458/
19. Hincapie AL, Slack M, Malone DC, MacKinnon NJ, Warholak TL. Relationship between patients' perceptions of care quality and health care errors in 11 countries: A secondary data analysis. Qual Manag Health Care. 2016 Jan 28;25(1):13–21.
20. Daliya P, Gemmill EH, Lobo DN, Parsons SL. A systematic review of patient reported outcome measures (PROMs) and quality of life reporting in patients undergoing laparoscopic cholecystectomy. HepatoBiliary Surg Nutr [Internet]. 2019 Jun [cited 2019 Dec 30];8(3):228–45. Available from: http://hbsn.amegroups.com/article/view/24693/23931
21. Dimitrov D V. Medical Internet of Things and Big Data in Healthcare. Healthc Inform Res [Internet]. 2016 Jul 1 [cited 2019 Dec 30];22(3):156. Available from: http://synapse.koreamed.org/DOIx.php?id=10.4258/hir.2016.22.3.156
22. Michell V. The Internet of Things and Opportunities for Pervasive Safety Monitored Health Environments. In: Handbook of Research on Patient Safety and Quality Care through Health Informatics [Internet]. IGI Global; 2013 [cited 2019 Dec 30]. p. 382–420. Available from: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-4546-2.ch020
23. Shah R, Chircu A. IoT and AI in Healthcare: A Systematic Literature Review. Issues Inf Syst. 2018;19(3):33–41.
24. Goldhahn J, Rampton V, Spinas GA. Could artificial intelligence make doctors obsolete? BMJ [Internet]. 2018 Nov 7 [cited 2019 Dec 30];363:k4563. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.k4563
25. Yong YP, Karangizi A, Banerjea A. A persuasive intervention: improving the

compliance of extended venous thromboembolism prophylaxis following cancer resections in a tertiary colorectal and hepatobiliary unit. BMJ Qual Improv Reports [Internet]. 2014 Oct 29 [cited 2019 Dec 30];3(1):u202594.w2216. Available from: http://qir.bmj.com/lookup/doi/10.1136/bmjquality.u202594.w2216
26. Kumar P, Jordan M, Caesar J, Miller S. Improving the management of sepsis in a district general hospital by implementing the "Sepsis Six" recommendations. BMJ Qual Improv Reports. 2015;4(1):u207871.w4032.