F1000Research

Check for updates

CORRESPONDENCE

Unintended consequences of machine learning in medicine? [version 1; referees: 2 approved]

Laura McDonald[1], Sreeram V. Ramagopalan [iD][1], Andrew P. Cox[2], Mustafa Oguz[2]

[1]Centre for Observational Research and Data Sciences, Bristol-Myers Squibb, Uxbridge, UB8 1DH, UK
[2]Evidera, London, W6 8DL, UK

Abstract

Machine learning (ML) has the potential to significantly aid medical practice. However, a recent article highlighted some negative consequences that may arise from using ML decision support in medicine. We argue here that whilst the concerns raised by the authors may be appropriate, they are not specific to ML, and thus the article may lead to an adverse perception about this technique in particular. Whilst ML is not without its limitations like any methodology, a balanced view is needed in order to not hamper its use in potentially enabling better patient care.

Open Peer Review

Referee Status: ✔ ✔

	Invited Referees	
	1	**2**
version 1 published 19 Sep 2017	✔ report	✔ report

1 **Arturo Gonzalez-Izquierdo**, University College London, UK

Maria Pikoula, University College London, UK

Spiros Denaxas [iD], University College London, UK

2 **Hugo Schnack** [iD], University Medical Center Utrecht, Netherlands

Zimbo Boudewijns, University Medical Center Utrecht, Netherlands

Discuss this article

Comments (0)

Corresponding author: Laura McDonald (laura.mcdonald@bms.com)

Author roles: McDonald L: Conceptualization, Writing – Original Draft Preparation; **Ramagopalan SV**: Supervision, Writing – Review & Editing; **Cox AP**: Supervision, Writing – Review & Editing; **Oguz M**: Investigation, Writing – Review & Editing

How to cite this article: McDonald L, Ramagopalan SV, Cox AP and Oguz M. **Unintended consequences of machine learning in medicine? [version 1; referees: 2 approved]** *F1000Research* 2017, **6**:1707 (doi: 10.12688/f1000research.12693.1)

First published: 19 Sep 2017, **6**:1707 (doi: 10.12688/f1000research.12693.1)

There is significant interest in the use of machine learning (ML) in medicine. ML techniques can 'learn' from the vast amount of healthcare data currently available, in order to assist clinical decision making. However, a recent article[1] highlighted a number of consequences that may occur with increased ML use in healthcare, including physician deskilling, and that the approach is a 'black box' and unable to use contextual information during analysis.

Whilst we agree that Cabitza *et al*'s concerns are justified[1], we believe that a more balanced discussion could have been provided with regards to ML-based decision support systems (ML-DSS). As it stands, an impression is given that ML is flawed, rather than the issue being the way in which it is applied. The concerns raised are generally applicable to many analytical approaches, and reflect poor study design and/or a lack of analytical rigour than the particular technique being used.

The authors cite two examples to claim that ML-DSS could potentially reduce physician diagnostic accuracy. The mammogram example[2] shows reduction in sensitivity for *6* of the most discriminating of 50 radiologists. However, the mammogram ML-DSS referred to is old[2], and it is not clear how the underlying model was trained and evaluated. The model may perform well for some types of cancer, but not as well for others as a result of the training data. Indeed updates have been shown to increase detection sensitivity[3]. ML models can be refined by providing more data and results need to be critically appraised in this context. Additionally, no mention is made of the possible benefits of ML-DSS for less experienced staff. In the mammogram example, an improvement in sensitivity for 44 out of 50 radiologists was seen for easier to detect cancers. There was also an increased overall diagnostic accuracy when using ML-DSS in the electrocardiogram study[4]. Accuracy loss for experienced readers when using ML-DSS

is valid, but more reflective of training needed and not an outcome specific to ML-DSS. A knowledgeable doctor may have no need for an ML-DSS, but the tool could greatly assist less experienced staff.

Cabitza *et al.* also argue that the confounding caused by asthma in the outcome of patients with pneumonia would have not been observed in a neural network model. There are, however, methods to obtain the feature importance and the direction of the relationship between predictor variables and outcome in neural networks[5]. Further, some ML approaches, such as random forest, are more transparent than others and ML can easily be coupled with clinical expertise to develop risk models that have their benefits over traditional statistical modelling[6].

The issues highlighted by Cabitza *et al.* are more concerned with the studies themselves rather than an intrinsic flaw in ML methodology. To fully leverage ML or any other approach, users must have a good understanding of the caveats. In summary, we agree that ML-based approaches are not without their limitations, but the growing application of ML in healthcare has the potential to significantly aid physicians, especially in increasingly resource constrained environments. Informed, appropriate use of ML-DSS could, therefore, enable better patient care.

References

1. Cabitza F, Rasoini R, Gensini GF: **Unintended Consequences of Machine Learning in Medicine.** *JAMA.* 2017; **318**(6): 517–518.
   PubMed Abstract | **Publisher Full Text**
2. Povyakalo AA, Alberdi E, Strigini L, *et al.*: **How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography.** *Med Decis Making.* 2013; **33**(1): 98–107.
   PubMed Abstract | **Publisher Full Text**
3. Kim SJ, Moon WK, Kim SY, *et al.*: **Comparison of two software versions of a commercially available computer-aided detection (CAD) system for detecting breast cancer.** *Acta Radiol.* 2010; **51**(5): 482–490.
   PubMed Abstract | **Publisher Full Text**
4. Tsai TL, Fridsma DB, Gatti G: **Computer decision support as a source of**

interpretation error: the case of electrocardiograms. *J Am Med Inform Assoc.* 2003; **10**(5): 478–483.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**
5. Olden J, Jackson DA: **Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks.** *Ecol Model.* 2002; **154**(1–2): 135–150.
   Publisher Full Text
6. Ayer T, Chhatwal J, Alagoz O, *et al.*: **Informatics in radiology: comparison of logistic regression and artificial neural network models in breast cancer risk estimation.** *Radiographics.* 2010; **30**(1): 13–22.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Referee Status: ✓ ✓

✓   **Hugo Schnack** ⓘD , **Zimbo Boudewijns**

Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands

Machine learning (ML) methods are currently being applied in a wide range of fields. Theoretically, the ability to extract meaningful relations from large datasets holds a great promise for health care and could potentially offer new, unexpected insights into disease and recovery.

However, more critical voices have emerged warning of potential issues surrounding the use of ML and a number of points were addressed in the original paper by Cabitza *et al*. The authors of the current F1000 paper (McDonald *et al*) indicate they consider the issues raised as justified but call for a more balanced discussion regarding ML use in medicine. Indeed, Cabitza *et al*'s view is mainly negative. McDonald and colleagues correctly argue that it is often bad study design that leads to unreliable or unwanted outcomes and not the technique that is used. While a valid point, given how common bad study design or misuse of more common statistical techniques is, it is not unlikely that such issues will also arise from use of ML techniques in health care. Given the far reaching implications of unreliable AI-based decision support systems in health care careful scrutiny of the techniques is necessary.

In their discussion of the reduced sensitivity of radiologists evaluating mammograms, McDonald and colleagues bring to the defense of ML techniques that the model that led to unwanted outcomes was old and that it might work better or worse depending on which type of cancer is being studied. In addition they state that it was not clear how the model was trained and evaluated and that having appropriate feedback loops in place can improve the accuracy of a model considerably. These are obviously important points that should always be taken into account both when evaluating a specific model as well as for the evaluation of the use of ML in medicine in general. The quality of the algorithm but also the quality of available data will ultimately decide how useful a ML approach is for a given medical problem. This only goes to show that ML models should be applied with care and does not negate the original concerns put forward by Cabitza *et al.*

McDonald and colleagues also propose the use of more transparent ML techniques to prevent potential confounding variables that would not show in a black-box model. However, the predictive power of ML models often increases with their complexity, making transparency either very difficult or even impossible to obtain. As stated by the authors, users of ML models should have a good understanding of the limitations of the techniques. More often than not, ML models will be implemented in a collaboration between technical experts without medical knowledge and medical experts with limited technical expertise. Ensuring an optimal level of transparency of the models combined with the right amount of clinical expertise is therefore vital and while the authors mention this, there are no specific

recommendations proposed to actually achieve this. With the increased application of ML models in health care there is a need for guidelines on how to optimally make use of the most technically advanced techniques while making sure bias in the data and confounding variables are accounted for. One recommendation the authors could have made is the need to bring the two worlds (medical and ML-technical) together, bridging the gap by educating people to become professionals with a (bio)medical *and* technical expertise, as well as training medical doctors in critically working with computerized predictions.

Another point that could have been addressed is how to deal with the differences in experience between medical doctors. Superior performance of the most-experienced doctors could be used to improve the computerized models, while less-experienced doctors may likely gain expertise by reviewing the computerized predictions –provided the ML tool allows for a clear interpretation *how* it did come to its diagnosis. This property of ML algorithms should be further explored and developed, because it can and, in fact, should lead to further insight in the underlying mechanisms of diseases.

Is the rationale for commenting on the previous publication clearly described?
Yes

Are any opinions stated well-argued, clear and cogent?
Partly

Are arguments sufficiently supported by evidence from the published literature or by new data and results?
Yes

Is the conclusion balanced and justified on the basis of the presented arguments?
Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: (bio)medical data analysis

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 30 October 2017

✔ **Arturo Gonzalez-Izquierdo** , **Maria Pikoula** , **Spiros Denaxas** iD
Institute of Health Informatics, University College London, London, UK

The publication of this letter is both important and timely given the increased interest that statistical learning approaches applied to healthcare data are receiving.

The original article emphasized a negative perception of potential adverse consequences of machine learning (ML). It did not fully highlight the current benefits of using large amount of information for clinical decision making and the potential for methodological improvement with regards to statistical learning approaches. The very field of ML is rapidly evolving as illustrated by the rapid growth of deep learning

over the past years.

Minor comments
- Content could potentially be enhanced with a discussion on the notion that, precisely due to the outlined potential misuses and consequences, a systematic and strategic use of ML approaches must be developed and used to facilitate the robust application of such methods to healthcare data.

- The authors of the letter rightly point out that ML has similar advantages and drawbacks as any other analytical approach applied in a clinical setting. However, they fail to explain how clinician deskilling, which can be a real consequence of automation, could be averted or indeed whether it is an acceptable outcome, given overall positive effects. This is separate issue to that of algorithm predictive accuracy.

- "Accuracy loss for experienced readers when using ML-DSS is valid, but more reflective of training needed" - Is this training to improve interpretation of ML-DSS output in cases where it hinders correct reading? This could be relevant in the context of a study, however, it leaves open the possibility that in busy clinical settings, readers would be more likely to rely on computer aided detection.

- The authors could also highlight the potential for upskilling clinicians in the understanding of ML methods, which can enhance their interpretation of DSS and other data-driven processes. Clinicians could be invaluable in spotting when algorithms go wrong, even (and perhaps especially) for cases where they've been shown to overall outperform humans.

- Finally, the authors would make their case stronger by citing examples that demonstrate "proof of clinically important improvements in relevant outcomes compared with usual care, along with the satisfaction of patients and physicians", as a concrete counter-balance to the negative or inconclusive examples of the original article.

Is the rationale for commenting on the previous publication clearly described?
Yes

Are any opinions stated well-argued, clear and cogent?
Partly

Are arguments sufficiently supported by evidence from the published literature or by new data and results?
Partly

Is the conclusion balanced and justified on the basis of the presented arguments?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: biomedical informatics

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com