Available online at www.sciencedirect.com

Resuscitation

journal homepage: www.elsevier.com/locate/resuscitation

EUROPEAN RESUSCITATION COUNCIL

Letter to the Editor

Comments on: “Machine learning as a supportive tool to recognize cardiac arrest in emergency calls”

To the Editor,

We have with great interest read the recent article in Resuscitation entitled “Machine learning as a supportive tool to recognize cardiac arrest in emergency calls” by Blomberg et al. accepted on the 11th of January 2019.[1] While the study was interesting and did highlight one possible application for machine learning, we believe the way in which it has been reported is partially misleading and fails to address some concerns, which we would like to comment upon.

The study investigated if a machine learning framework could outperform dispatchers in the diagnosis of out hospital cardiac arrest (OHCA).[1] Voice recordings of 108.607 calls were retrospectively analyzed by the machine learning framework. There were a total of 918 OHCA and 107.689 non-OHCA.

1) Results showed that the machine learning framework compared to the dispatcher had significantly higher sensitivity (84.1% vs 72.5%, $p < 0.001$) but significantly lower specificity (97.3% vs 98.8%%, $p < 0.001$).

   However the authors write twice that the “machine learning framework [. . .] showed a significantly higher sensitivity and similar specificity”. We think this is misleading and it should consistently be stated that the sensitivity was significantly higher, but the specificity significantly lower.
2) The overall accuracy of the machine learning algorithm was not listed in the article. Based on the listed number of calls, and the stated sensitivity and specificity we calculated (see Table 1) that the overall accuracy of the machine learning framework was significantly lower than the performance of the dispatcher 97.2% (95% CI 97.1%–97.3%) vs 98.6% (95% CI 98.5%–98.7%). Confidence intervals were calculated using the Clopper–Pearson interval[2].

   The study does therefor not support the statement that the machine learning framework “performed better” which is listed in the abstract of the article.
3) Reducing the specificity of a test by changing a cut-off value at which a condition is diagnosed does increase the sensitivity.[3] A significant higher sensitivity with corresponding significant lower specificity therefore does not necessarily indicate a better test. Either the accuracy or the receiver operating characteristics of the test could instead be used to judge if the test was in fact overall superior.
4) The authors insightfully reflect on the current machine learning model’s value at the end of the discussion. “As such, machine learning should not be used as a stand-alone tool that can independently dispatch ambulances but could act as a supplement to dispatchers’ . . . ” This reflection is unfortunately lacking in both the abstract and the conclusion of the paper. Machine learning has huge potential but the current level of layman hype must be tempered with realistic expectations of current machine learning performance.
5) The machine learning framework approach if directly applied would lead to an additional 1616 False Positive diagnosis of OHCA. This could put a significant strain on the emergency system, including ambulance services. The direct economic consequence of implementing such a system should therefore be carefully considered and compared to other interventions

Table 1 – Positive and negative rates for machine learning vs dispatcher in the diagnosis of OHCA.

	Machine learning framework	Dispatcher
True positive	918 × 84.1% = 772	918 × 72.5% = 666
False positive	107.689 × (100%–97.3%) = 2.908	107.689 × (100%– 98.8%) = 1.292
True negative	107.689 × 97.3% = 104.781	107.689 × 98.8% = 106.397
False negative	918 × (100%–84.1%) = 146	918 × (100%–72.5%) = 252
#True cases	105.553	107.063
Accuracy	97.2% (95% CI 97.1%–97.3%)	98.6% (95% CI 98.5%– 98.7%)

Legend: Results were calculated using published numbers by the authors. Confidence intervals calculated using Clopper–Pearson intervals.

(e.g. instructing dispatchers to considered OHCA more often in unconscious patients).

Funding

No funding was provided for the writing of this letter.

Conflict of interest statement

Authors have no conflicts of interest relating to the present work.

REFERENCES

1. Blomberg SN, Folke F, Ersboll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation 2019;138:322–9.
2. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934;26:404–13.
3. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.

Felix C. Müller*
Mathias W. Brejnebøl
Department of Radiology, Herlev and Gentofte Hospital, University Hospital at the University of Copenhagen, Denmark

* Corresponding author.
E-mail address: Christoph.felix.mueller@regionh.dk (F. Müller).

Received 16 May 2019

http://dx.doi.org/10.1016/j.resuscitation.2019.05.042