

CORRESPONDENCE

Validation of a convolutional neural network that reliably identifies electromyographic compound motor action potentials following train-of-four stimulation: an algorithm development experimental study—Reply to: *Br J Anaesth Open* 2024:100264.



Richard H. Epstein^{1,*}, Olivia F. Perez¹, Ira S. Hofer², J. Ross Renew³, Réka Nemes⁴ and Sorin J. Brull³

¹Department of Anesthesiology, Pain Management and Perioperative Medicine, University of Miami Miller School of Medicine, Miami, FL, USA, ²Department of Anesthesiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA, ³Department of Anesthesiology and Perioperative Medicine, Mayo Clinic College of Medicine and Science, Jacksonville, FL, USA and ⁴Department of Anesthesiology and Intensive Care, University of Debrecen, Debrecen, Hungary

*Corresponding author. E-mail: repstein@med.miami.edu

Keywords: electromyography; machine learning; neural network; neuromuscular block; train-of-four

Editor—we appreciate the careful reading of our recent publication¹ and the insightful comments by Silliman and colleagues.² We inadvertently described the network as ‘convolutional’ when it was actually a fully connected network, which went unnoticed during the peer review process. A corrigendum has been submitted to the *BJA Open*. Nonetheless, this nomenclature error does not affect our results. We presented the pseudocode for the fully connected neural network we implemented in Figure 3 and the accompanying legend.¹ Thus, we think the model we used was adequately described, despite the labelling issue.¹ We also think it should have been obvious from the manuscript text (‘performance was assessed for the 29 subjects by 28 leave-one-out cross-validation runs.’) that our leave-one-out approach to cross-validation testing excluded one subject for testing from each validation cycle, not a single result of the 28 205 waveforms evaluated.¹ We performed 10-fold cross-validation (excluding subjects) during the original analysis but elected to present the leave-one-out results to better identify any subjects in whom the model performed poorly (none of which was found). In preparing our response to the correspondence, we repeated the 10-fold cross-validation using the same criteria as for

external validation to address hyperparameter tuning and found almost no differences in model performance among the range of values tested, other than related to execution times.

This is the first of a planned series of manuscripts exploring the use of fully connected and convolutional neural networks to evaluate physiological waveforms and various sampling and generative strategies to reduce the number of tagged samples needed to train such networks. We are addressing many of the issues raised by Silliman and colleagues² in our ongoing research. To briefly summarise our findings to date, the extremely accurate classification with the tested fully connected network is replicated by a simple convolutional neural network comprising two convolutional layers, two pooling layers, two densely connected layers (each with a dropout), and a softmax output layer with two categories. It is also replicated by a more complex convolutional neural network comprising six convolutional layers, three pooling layers, two densely connected layers (each with a dropout), and a softmax output layer with two categories. A simple time series input of the voltages in the compound motor action potential (cMAP) responses, as suggested by Silliman and colleagues,² only resulted in a balanced accuracy of ~92%. Interpolation of the cMAPs at 0.1 ms resolution does not seem necessary for cMAP classification because the models perform equivalently with the native 1-ms sampling interval. However, we clarify that the inputs for each sample to the fully

DOIs of original article: [10.1016/j.bjao.2023.100236](https://doi.org/10.1016/j.bjao.2023.100236), [10.1016/j.bjao.2024.100264](https://doi.org/10.1016/j.bjao.2024.100264).

connected neural networks are vectors, the same as is the case for input for the model when used to classify handwritten digits.³ The set of normalised cMAP images (order 3 tensors) are reshaped using row-major ordering into a two-dimensional array for input into the fully connected network. We are unsure, yet, if scaling the cMAPs to produce similar-sized waveforms is necessary, although this step executes quickly. However, we have not yet determined if these findings will hold as we reduce the number of samples in the training dataset and use generative methods to reduce the burden of tagging tens of thousands of cMAPs to train the model. There are differences among the various models in terms of the time required to train the model and to classify each waveform, the latter ranging from 0.1 to 1 ms per cMAP on a moderately powered desktop computer without use of a graphical processing unit. Nonetheless, even the slowest executing models would be sufficiently rapid given that each stimulation sequence only involves classifying four waveforms. We used a multiclass output layer because our ongoing work distinguishes between invalid responses that reflect the absence of an EMG response to nerve stimulation and those with superimposed electrical noise that might be masking a real signal. The binary output approach suggested by Silliman and colleagues² would not have been sufficient for such classification.

We recognise that multiple alternative models might perform more efficiently or better than our model, but our goal was not to identify the best-performing model. Given that all performance metrics are close to 100%, we think chasing a few tenths of a percentage of additional projected performance would be of marginal benefit. Our paper was not a Kaggle neural network competition entry⁴ but rather an exploration of the feasibility of the approach we described for the identification of valid cMAPs after electrical stimulation. For manufacturers considering implementation, there would be a benefit from balancing efficiency, accuracy, and cost, but such concerns were not relevant to the paper's hypothesis, namely that a neural network used for digit recognition would have high accuracy when applied to the identification of valid cMAPs.

Silliman and colleagues² raised three specific issues to which we would like to respond.

First, they questioned performance when the compound motor action potential (cMAP) amplitudes are small. We reran our model using the training dataset against the testing dataset and compared accuracy from cMAPs with amplitude >1 mV vs those where the amplitude was ≤ 0.5 mV. Overall accuracy for the 14 825 cMAPs with amplitudes >1 mV was 99.09% (135 misclassified), and for the 6087 cMAPs with amplitudes ≤ 0.5 mV, accuracy was 99.98% (one misclassified). Silliman and colleagues² also questioned class imbalance in the training dataset (22.4% valid cMAPs), suggesting that this would result in a biased model 'that would incorrectly predict valid responses as noise.' Although overrepresenting the majority class can potentially lead to the bias described, we observed the opposite results. Of the 137 misclassified cMAPs in the test dataset (with a prevalence of 81.9% valid cMAPs), 132 were invalid waveforms (noise) reported as valid. These results indicate that the performance of the algorithm was unaffected by small amplitude cMAPs and that the degree of class imbalance present in the training dataset was not sufficient to bias toward misclassifying valid responses as noise. We suspect this finding is related to the very uniform morphology of the valid cMAPs across the range of amplitudes

from 0.05 to 20 mV compared with the artifacts, the shapes of which were much more diverse. Thus, although the valid cMAP prevalence was relatively low in the training dataset compared with the test dataset, the monotonous cMAP morphology may have been sufficient to generate a disproportionately strong signal for the neural network to neutralise the class imbalance.

The second question raised relates to what we think is a misunderstanding of the study's objective, which was to classify cMAPs at the adductor pollicis as a surrogate for diaphragmatic paralysis. Surgeons performing laparoscopic and robotic procedures often request 'complete paralysis' to prevent interference with their magnified view of the surgical field. A response by the anaesthetist that, 'the patient is fully paralysed', based on the absence of twitches from the hand, is generally neither well received nor believed by surgeons when the image on the screen is moving as a result of diaphragmatic activity. Our goal is to provide a more reliable method than arbitrary amplitude thresholds to ensure intense neuromuscular block of the diaphragm, which requires that fewer than one or two post-tetanic twitches are present at the adductor pollicis muscle.⁵ Our suggested approach allows preservation of at least a small amount of neuromuscular function (i.e. a small number of post-tetanic cMAPs), which may be advantageous for reliable antagonism of the neuromuscular block.⁶ The algorithm we describe can detect depolarisation and repolarisation events from the adductor pollicis muscle an order of magnitude below what several commercial EMG devices report.¹ Correlation with visible twitches would be interesting, but not of much practical significance given that the hands are typically not accessible during laparoscopic or robotic surgery, assessment of the presence of twitches is subjective and inconsistently related to quantitative assessment, and that EMG monitors often undercount the number of visible twitches.⁷ Because it is really the diaphragm that is of interest during deep neuromuscular block during laparoscopic and robotic surgery, we think that high sensitivity to detect small degrees of depolarisation of the adductor pollicis muscle in response to ulnar nerve stimulation is both desirable and necessary. It is currently problematic that calculation of the train-of-four count depends on the device used rather than on the fundamental neuromuscular block physiology. We anticipate that anaesthetists would interpret the magnitude of cMAP in mV when assessing the need for any potential clinical intervention. In a recent correspondence, Todd⁸ admonished, 'Clearly a "twitch is not always a twitch" when we are comparing quantitative devices with our own eyes—and we should not expect perfect concordance. This also applies when comparing "weak responses" between different devices.'

Finally, Silliman and colleagues² assert that implementing a neural network would significantly increase the cost of the monitors. We do not concur. The open-source TensorFlow Lite library (Google, Mountain View, CA, USA) used for neural network processing can run on small, inexpensive micro-computer systems such as the Raspberry Pi (currently priced online between £30 and £100 depending on the speed of the central processing unit and the amount of random access memory).⁹ To put this cost in perspective, in the United States, single-use EMG sensors typically are priced around \$20 (£15.74). Although one would not want to train a neural network using such devices, classification using a prebuilt model is rapid because it only involves a single forward pass

through the model (i.e. arithmetic). There are numerous descriptions online for image-processing neural networks implemented on microcomputers such as Raspberry Pi (Google search: raspberry pi neural network project).

We stand by our conclusions that an image-based neural network approach to recognising valid cMAPs at the adductor pollicis muscle after train-of-four stimulation of the ulnar nerve at the wrist is highly accurate, practicable, and would greatly improve the ability to monitor deep levels of neuromuscular block. Further research is needed to validate whether a model developed using EMG waveforms obtained from one device can be used interchangeably with EMG waveforms from other devices and whether the model will apply to groups of patients or circumstances not represented in the current modelling (e.g. in children, where waveforms are from different muscles such as the hallucis brevis muscle, in patients with diseases affecting neuromuscular transmission).

Declaration of interests

SJB has intellectual property assigned to Mayo Clinic (Rochester, MN); is a consultant for Merck & Co., Inc. (Kenilworth, NJ); is a principal, shareholder, and Chief Medical Officer in Sensime AB (Uppsala, Sweden); and an unpaid member of the Scientific/Clinical Advisory Boards for The Doctors Company (Napa, CA); Coala Life Inc. (Irvine, CA); NMD Pharma (Aarhus, Denmark); and Takeda Pharmaceuticals (Cambridge, MA). JRR has ongoing industry-sponsored research (Merck & Co., Inc) with funds to his employer and has served on a Scientific Advisory Board for Sensime AB (Uppsala, Sweden). ISR is the founder and President of Extrico Health, an informatics company that helps hospitals leverage data from their electronic health record for decision-making purposes, receives research support and serves as a consultant for Merck, and is funded, in part, by NIH grant 1K01HL150318. The other authors report no conflicts.

References

1. Epstein RH, Perez OF, Hofer IS, Renew JR, Nemes R, Brull SJ. Validation of a convolutional neural network that reliably identifies electromyographic compound motor action potentials following train-of-four stimulation: an algorithm development experimental study. *BJA Open* 2023; **8**: 100236
2. Silliman W, Wedemeyer Z, Jelacic S, Bowdle A, Michaelsen KE. Validation of a convolutional neural network that reliably identifies electromyographic compound motor action potentials following train-of-four stimulation. Comment on *Br J Anaesth Open*. 2023; **8**: 100236. *Br J Anaesth Open* 2024
3. Radečić D. *Deep learning with R and Keras: build a handwritten digit classifier in 10 minutes*. Available from: <https://appsilon.com/r-keras-mnist/>. [Accessed 27 January 2024]
4. Kaggle. *Digit recognition*. Available from: <https://www.kaggle.com/competitions/digit-recognizer>. [Accessed 22 January 2024]
5. Werba A, Klezl M, Schramm W, et al. The level of neuromuscular block needed to suppress diaphragmatic movement during tracheal suction in patients with raised intracranial pressure: a study with vecuronium and atracurium. *Anaesthesia* 1993; **48**: 301–3
6. Hunter JM, Blobner M. Under-dosing and over-dosing of neuromuscular blocking drugs and reversal agents: beware of the risks. *Br J Anaesth Adv* 2024; **132**: 461–5
7. Bowdle A, Bussey L, Michaelsen K, et al. Counting train-of-four twitch response: comparison of palpation to mechanomyography, acceleromyography, and electromyography. *Br J Anaesth* 2020; **124**: 712–7
8. Todd MM. Agreement of posttitanic count between monitors: comment. *Anesthesiology* 2023; **139**: 910–1
9. TensorFlow for Mobile & Edge. Available from: <https://www.tensorflow.org/lite>. [Accessed 22 January 2024]

doi: 10.1016/j.bjao.2024.100265