Machine Learning in Intraoperative Neurophysiological Monitoring

Subjects: Clinical Neurology

Contributor: Dougho Park , Injung Kim

Intraoperative neurophysiological monitoring (IONM) is being applied to a wide range of surgical fields as a diagnostic tool to protect patients from neural injuries that may occur during surgery. However, several contributing factors complicate the interpretation of IONM, and it is labor- and training-intensive. Meanwhile, machine learning (ML)-based medical research has been growing rapidly, and many studies on the clinical application of ML algorithms have been published.

intraoperative neurophysiological monitoring artificial intelligence machine learning deep learning

1. Introduction

Intraoperative neurophysiological monitoring (IONM) is an essential diagnostic tool for the improvement of patient safety via detection of neurological changes during surgery. IONM is currently being applied in various types of surgery, such as open cranial surgery, spinal decompression, head and neck surgery, and peripheral nerve surgery [1][2][3]. The most prominent advantage of IONM is its use to confirm functional integrity in real time during surgery. When a warning sign occurs, an immediate rescue intervention can be performed in the operating room, minimizing neural injuries and enabling rapid postoperative treatment [4].

However, despite its advantages, some limitations exist in interpreting IONM. In particular, several factors complicate the real-time interpretation of multimodal signals during IONM. In interpreting neurophysiological changes related to surgical factors, a multidisciplinary approach between surgeons and physiatrists is essential, and substantial information sharing between them is vital for accurate interpretation [5]. Further, non-surgical factors, such as anesthesia, the general condition of the patient, and mechanical defects, have to be considered simultaneously with surgical factors [6]. Another hurdle in interpreting IONM is that experts may interpret the same results differently [7]. Therefore, the performance and interpretation of IONM require a substantial level of training to minimize inter-rater variability. Lastly, to respond sensitively to neural deterioration that occurs during surgery, the expert must maintain high concentration even during long operations. Consequently, IONM is a complicated, labor-intensive, and expensive process (**Figure 1**) [8].

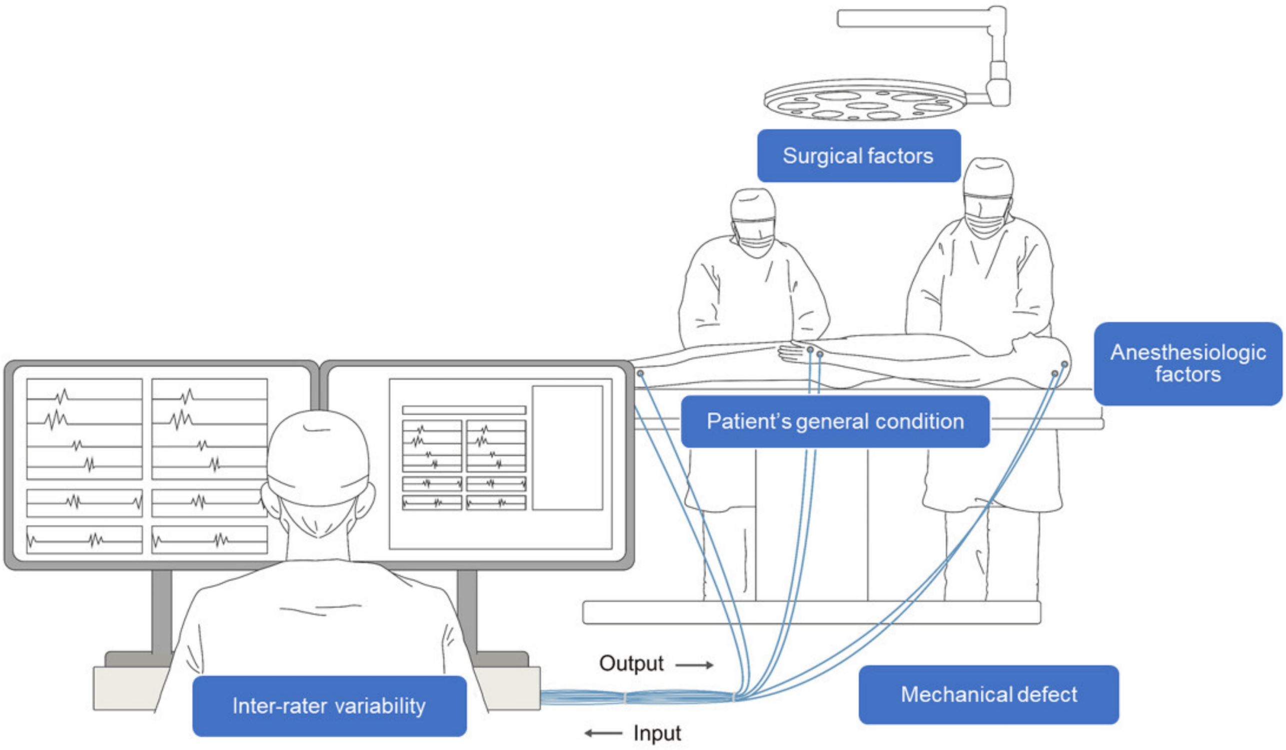

Figure 1. Schematic illustration of intraoperative neurophysiological monitoring (IONM). A multidisciplinary approach between the surgeon, physiatrist, and anesthesiologist is necessary throughout the process. Several confounding factors, such as surgical, anesthesiologic, and mechanical factors, as well as the patient's condition and inter-rater variability complicate the interpretation of IONM.

The use of machine learning (ML) in medical research and clinical practice is rapidly expanding [9]. In particular, ML is increasingly being used for diagnosis and prognosis [10][11], as well as for disease classification, replacing existing methods [10]. Moreover, ML can execute proxy decision-making at the level of medical experts [12] and can readily and efficiently handle a large number of samples and variables [13]. Artificial intelligence (AI) models have the additional benefit of continuously improving themselves by learning from additional data and by applying more advanced techniques [14]. Although their performance depends on data quality and learning algorithms, in general, ML models have yielded promising results in clinical medicine [10][15].

2. Related Studies on the Application of ML Algorithms to IONM

Jamaludin et al. [16] presented k-nearest neighbors (KNN)- and bagged trees-based ML models to predict positive outcomes after lumbar surgeries in 55 patients. The positive outcome was defined as a motor improvement after the surgery. They compared ML-based prediction methods with pre-existing criteria (50% of transcranial motor evoked potential improvement). In their work, the ML-based method showed a relatively higher sensitivity (87.5%) but lower specificity (33.3%), which was inferior to the pre-existing criteria for predicting postoperative motor improvement. Consequently, they suggested that their proposed method had more room to advance with a large-scale study.

Agaronnik et al. [17] developed a deep learning-based automated detection system for neuromonitoring documentation. They first identified operative reports containing neuromonitoring documentation by rule-based natural language processing. Subsequently, they applied a deep learning-based natural language processing model to identify events indicating a change in status, difficulty in establishing baseline signals, and a stable course, from the reports of 993 patients who underwent spinal surgery. For detection of a change in status, they achieved an area under the receiver operating characteristic curve (AUROC) of 1.0 and an F1 score of 0.80 (discussed further below). Their results suggest that deep learning-based natural language processing models can identify medical documentation of IONM from a large volume of reports in a validated and timely manner.

Kortus et al. [18] predicted electromyography (EMG) signal characteristics during thyroid surgery in 34 patients. They utilized a Bayesian convolutional neural network (CNN) approach to simultaneously classify action potentials and assess signal characteristics. The extended model with two hidden layers with sigmoid activation yielded the best predictive value, with an accuracy of 97.6%. By applying a Bayesian deep learning model, they estimated the uncertainty of the model output, which improved the interpretability of the prediction. They demonstrated that the deep learning model was suited for robust interpretation of electrophysiological signals.

Zha et al. [19] applied a deep learning model to free-running EMG for recurrent laryngeal nerve monitoring during thyroid surgery. They classified the EMG according to morphology and presented a hybrid model that combined a CNN approach with a long short-term memory (LSTM) network. Their proposed CNN-LSTM hybrid model yielded an accuracy of 89.54% and a sensitivity of 94.23%. Their results demonstrated the possibility of reducing inter-rater variability in the reading of free-running EMGs by using deep learning models, reducing the interpretive burden on the expert.

Verdonck et al. [20] presented a model for the interpretation of outliers via train-of-four (TOF) measurements during intraoperative acceleromyographic neuromuscular monitoring. They used a cost-sensitive logistic regression model to analyze 533 TOF measurements from 35 patients. In terms of the predictive power of this model, the AUROC was 0.91 (95% confidence interval: 0.72–0.97) and the F1 score was 0.86 (0.48–0.97). Their model proved outstanding for binary classification. Their study is important since it showed that the model could analyze TOF measurements to automatically identify outliers during intraoperative accelero-myographic neuromuscular monitoring.

Qiao et al. [21] conducted visual evoked potential (VEP) monitoring in 76 patients who underwent surgical decompression for sellar region tumors. They presented a model that could classify amplitude changes in VEPs during surgery, by combining CNN and recurrent neural network (RNN) algorithms. The target class was divided into three groups: increased VEP amplitude (>25% increase), decreased VEP amplitude (>25% decrease), and no change in VEP amplitude. In this study, the overall accuracy of multiclass classification was 87.4% (84.2–90.1%). The sensitivities for classification of no change in VEP, increasing VEP, and decreasing VEP were 92.6%, 78.9%, and 83.7%, respectively, and their specificities were 80.5%, 93.3%, and 100.0%, respectively.

Somatosensory evoked potential (SEP) is a modality that acts as the framework of intraoperative spinal surgery monitoring [4]. Fan et al. [22] utilized least squares and multi-support vector regression models on 15 patients undergoing spinal surgery to intraoperatively interpret the SEP results. They defined the warning criteria as an amplitude reduction of ≥50% or a latency delay of ≥10%. Target outcomes were classified as successful, false-positive, or trauma cases. Their intelligent decision system lowered the false warning rate compared with their conventional method and enabled more accurate detection of spinal cord trauma. The multi-support vector regression model performed better than the least squared support vector regression model.

Table 1 summarizes studies on the application of ML algorithms to IONM.

Table 1. The application of machine learning in the field of intraoperative neurophysiological monitoring.

Author (Year)	Samples	IONM Modality	Models	Target Outcome	Summary of Results
Jamaludin et al. [16] (2022)	55 patients who underwent lumbar surgeries	MEP	KNN and Bagged trees	Positive outcome (motor improvement)	The proposed method was inferior to the existing criteria. - Sensitivity: 87.5% - Specificity: 33.3%
Agaronnik et al. [17] (2022)	993 patients who underwent spinal surgery	MEP and SEP	Deep learning-based natural language processing	Change in status	- AUROC: 1.00 - F1 score: 0.80
				Difficulty establishing baseline	- AUROC: 0.97 - F1 score: 0.80
				Stable course	- AUROC: 0.91 - F1 score: 0.93
Kortus et al. [18] (2021)	34 patients who underwent thyroid surgery	EMG	Bayesian CNN	Classification of action potentials	- Accuracy: 97.6% - Precision: 97.7% - Recall: 97.6%

Author (Year)	Samples	IONM Modality	Models	Target Outcome	Summary of Results
Zha et al. [19] (2021)	5 patients who underwent thyroid surgery	Free-running EMG	Hybrid CNN-LSTM model	EMG signal waveforms (quiet, evoked, irritation, burst, injury, and artifact)	The hybrid model could automatically classify the free-running EMG. - Accuracy: 89.54% - Sensitivity: 94.23%
Verdonck et al. [20] (2021)	533 TOF samples from 35 patients	AMG	Cost-sensitive logistic regression	Outlier TOF measurement	AMG-based intraoperative measurements of TOF outliers displayed an increased monitoring consistency. - F1 score: 0.86 - AUROC: 0.91
Qiao et al. [21] (2019)	76 cases with sellar region tumor	VEP	CNN and RNN combination	Increasing, decreasing, or no change of VEP amplitude	- Overall accuracy of CNN/RNN combined vs. traditional method using single VEP images: 87.4% and 83.1%, respectively
Fan et al. [22] (2016)	10 successful surgeries (158 samples) 4 false positives (72 samples) 1 trauma case (14 samples)	SEP	LS-SVR and M-SVR	Successful case: no interruption False positive case: surgery interrupted by an expert without spinal cord injury Trauma case: surgery interrupted by an expert, with spinal cord injury	- False positive rates NBM vs. LS-SVR vs. M-SVR: 0.304, 0.080, and 0.068, respectively. - True warning rate NBM vs. LS-SVR vs. M-SVR: 0.500, 0.714, and 0.714, respectively.

IONM, intraoperative neurophysiological monitoring; MEP, motor evoked potential; KNN, k-nearest neighbors; SEP, somatosensory evoked potential; AUROC, area under the receiver operating characteristic curve; EMG, electromyography; CNN, convolutional neural network; LSTM, long short-term memory; TOF, train-of-four; AMG,

References

1. Stankovic, P.; Wittlinger, J.; Georgiew, R.; Dominas, N.; Hoch, S.; Wilhelm, T. Continuous intraoperative neuromonitoring (cIONM) in head and neck surgery—A review. HNO 2020, 68, 86–92.
2. Shiban, E.; Meyer, B. Intraoperatives Neuromonitoring in der rekonstruktiven Halswirbelsäulenchirurgie. Orthopade 2018, 47, 526–529.

accelerometry, phy.: VEP: visual evoked potential; RNN: recurrent neural network; LS: least squares; ML: multi; SVR: support vector regression; NBM: nominal baseline method.

3. Limitations

Nonsurgical factors are important confounders in the interpretation of IONM. In particular, IONM modalities are very sensitive to changes brought about by anesthesia-related factors [23]. The anesthetic methodology used, the use of neuromuscular blockade, the patient's blood pressure and body temperature, and prolonged operation time, among other factors, can substantially affect IONM signals, even in the absence of a neural insult [24]. Cross-disciplinary collaboration is essential in the construction of a model that can consider these various factors simultaneously. In addition, since many variables need to be processed, it is essential that models are trained and validated on high-quality datasets with large numbers of samples that share the same features.

Another point to consider is that false positives results in ambiguity in the interpretation of IONM signals [25]. When a warning signal occurs during surgery, regardless of its reliability (true or false positive), surgeons and anesthesiologists respond by initiating a rescue intervention process [6]. This may be an important confounding factor in the predictive value of ML algorithms. If postoperative neurological deficit is defined as a dependent variable during the construction of an ML model, there may be disagreements in the input to provide to an ML algorithm when a warning signal is issued. Therefore, as demonstrated by Zha et al. [19], morphological classification may be a more realistic alternative to the direct interpretation of evoked potential. In other words, although the ML algorithm reads the signals, the human expert's intervention remains essential in determining the reliability and cause of a warning sign.

However, inter-rater variability in the interpretation of IONM signals is inevitable when human experts are involved [26]. For example, results will be interpreted differently depending on the definition of the baseline. The presence or absence of a warning sign depends on whether the baseline is static or changes in response to previous waveforms during the surgery [27]. Studies also vary in their definitions of postoperative neurological deficits [28]. This difference in the interpretation of the ground truth can cause high variability between ML algorithms. The interpretation of IONM results may also vary depending on the degree of training of the expert [8].

3. Einarsson, H.B.; Poulsen, F.R.; Derejko, M.; Korshoj, A.R.; Qerama, E.; Pedersen, C.B.; Halle, B.; Nielsen, T.H.; Clausen, A.H.; Korshoj, A.R.; et al. Intraoperative neuromonitoring during brain surgery. Ugeskr Laeger 2021, 183, V09200712.

4. Stecker, M. A review of intraoperative monitoring for spinal surgery. Surg. Neurol. Int. 2012, 3, S174–S187.

5. Tewari, A.; Francis, L.; Samy, R.N.; Kurth, D.C.; Castle, J.; Frye, T.; Mahmoud, M. Intraoperative neurophysiological monitoring team's communique with anesthesia professionals. J. Anaesthesiol. Clin. Pharmacol. 2018, 34, 84–93.

6. Park, D.; Kim, B.H.; Lee, S.-E.; Jeong, E.; Cho, K.; Park, J.K.; Choi, Y.-J.; Jin, S.; Hong, D.; Kim, M.-C. Usefulness of Intraoperative Neurophysiological Monitoring during the Clipping of Unruptured Intracranial Aneurysm: Diagnostic Efficacy and Detailed Protocol. Front. Surg. 2021, 8, 631053.

7. Gruenbaum, B.F.; Gruenbaum, S.E. Neurophysiological monitoring during neurosurgery: Anesthetic considerations based on outcome evidence. Curr. Opin. Anaesthesiol. 2019, 32, 580–584.

8. Wojtczak, B.; Kaliszewski, K.; Sutkowski, K.; Głód, M.; Barczyński, M. The learning curve for intraoperative neuromonitoring of the recurrent laryngeal nerve in thyroid surgery. Langenbeck's Arch. Surg. 2016, 402, 701–708.

9. Toh, C.; Brody, J.P. Applications of Machine Learning in Healthcare. In Smart Manufacturing—When Artificial Intelligence Meets the Internet of Things; IntechOpen: London, UK, 2021.

10. Park, D.; Jeong, E.; Kim, H.; Pyun, H.W.; Kim, H.; Choi, Y.-J.; Kim, Y.; Jin, S.; Hong, D.; Lee, D.W.; et al. Machine Learning-Based Three-Month Outcome Prediction in Acute Ischemic Stroke: A Single Cerebrovascular-Specialty Hospital Study in South Korea. Diagnostics 2021, 11, 1909.

11. Kim, J.O.; Jeong, Y.-S.; Kim, J.H.; Lee, J.-W.; Park, D.; Kim, H.-S. Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database. Diagnostics 2021, 11, 943.

12. Yoo, T.K.; Ryu, I.H.; Choi, H.; Kim, J.K.; Lee, I.S.; Kim, J.S.; Lee, G.; Rim, T.H. Explainable Machine Learning Approach as a Tool to Understand Factors Used to Select the Refractive Surgery Technique on the Expert Level. Transl. Vis. Sci. Technol. 2020, 9, 8.

13. Schinkel, M.; Paranjape, K.; Nannan Panday, R.S.; Skyttberg, N.; Nanayakkara, P.W.B. Clinical applications of artificial intelligence in sepsis: A narrative review. Comput. Biol. Med. 2019, 115, 103488.

14. Telikani, A.; Tahmassebi, A.; Banzhaf, W.; Gandomi, A.H. Evolutionary Machine Learning: A Survey. ACM Comput. Surv. 2022, 54, 1–35.

15. Shin, S.; Austin, P.C.; Ross, H.J.; Abdel-Qadir, H.; Freitas, C.; Tomlinson, G.; Chicco, D.; Mahendiran, M.; Lawler, P.R.; Billia, F.; et al. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. ESC Heart Fail. 2020, 8, 106–115.

16. Jamaludin, M.R.; Lai, K.W.; Chuah, J.H.; Zaki, M.A.; Hasikin, K.; Abd Razak, N.A.; Dhanalakshmi, S.; Saw, L.B.; Wu, X. Machine Learning Application of Transcranial Motor-Evoked Potential to Predict Positive Functional Outcomes of Patients. Comput. Intell. Neurosci. 2022, 2022, 2801663.

17. Agaronnik, N.D.; Kwok, A.; Schoenfeld, A.J.; Lindvall, C. Natural language processing for automated surveillance of intraoperative neuromonitoring in spine surgery. J. Clin. Neurosci. 2022, 97, 121–126.

18. Kortus, T.; Krüger, T.; Gühring, G.; Lente, K. Automated robust interpretation of intraoperative electrophysiological signals—A bayesian deep learning approach. Curr. Dir. Biomed. Eng. 2021, 7, 69–72.

19. Zha, X.; Wehbe, L.; Sclabassi, R.J.; Mace, Z.; Liang, Y.V.; Yu, A.; Leonardo, J.; Cheng, B.C.; Hillman, T.A.; Chen, D.A.; et al. A Deep Learning Model for Automated Classification of Intraoperative Continuous EMG. IEEE Trans. Med. Robot. Bionics 2021, 3, 44–52.

20. Verdonck, M.; Carvalho, H.; Berghmans, J.; Forget, P.; Poelaert, J. Exploratory Outlier Detection for Acceleromyographic Neuromuscular Monitoring: Machine Learning Approach. J. Med. Internet Res. 2021, 23, e25913.

21. Qiao, N.; Song, M.; Ye, Z.; He, W.; Ma, Z.; Wang, Y.; Zhang, Y.; Shou, X. Deep Learning for Automatically Visual Evoked Potential Classification During Surgical Decompression of Sellar Region Tumors. Transl. Vis. Sci. Technol. 2019, 8, 21.

22. Fan, B.; Li, H.-X.; Hu, Y. An Intelligent Decision System for Intraoperative Somatosensory Evoked Potential Monitoring. IEEE Trans. Neural Syst. Rehabil. Eng. 2016, 24, 300–307.

23. Ugawa, R.; Takigawa, T.; Shimomiya, H.; Ohnishi, T.; Kurokawa, Y.; Oda, Y.; Shiozaki, Y.; Misawa, H.; Tanaka, M.; Ozaki, T. An evaluation of anesthetic fade in motor evoked potential monitoring in spinal deformity surgeries. J. Orthop. Surg. Res. 2018, 13, 227.

24. Nunes, R.R.; Bersot, C.D.A.; Garritano, J.G. Intraoperative neurophysiological monitoring in neuroanesthesia. Curr. Opin. Anaesthesiol. 2018, 31, 532–538.

25. Chung, J.; Park, W.; Hong, S.H.; Park, J.C.; Ahn, J.S.; Kwun, B.D.; Lee, S.-A.; Kim, S.-H.; Jeon, J.-Y. Intraoperative use of transcranial motor/sensory evoked potential monitoring in the clipping of intracranial aneurysms: Evaluation of false-positive and false-negative cases. J. Neurosurg. 2019, 130, 936–948.

26. Ney, J.P.; van der Goes, D.N.; Nuwer, M.; Emerson, R.; Minahan, R.; Legatt, A.; Galloway, G.; Lopez, J.; Yamada, T.; Ney, J.P.; et al. Evidence-based guideline update: Intraoperative spinal monitoring with somatosensory and transcranial electrical motor evoked potentials: Report of the

Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and the American Clinical Neurophysiology Society. Neurology 2012, 79, 292–294.

27. Chen, J.H.; Shilian, P.; Cheongsiatmoy, J.; Gonzalez, A.A. Factors Associated With Inadequate Intraoperative Baseline Lower Extremity Somatosensory Evoked Potentials. J. Clin. Neurophysiol. 2018, 35, 426–430.
28. Nasi, D.; Meletti, S.; Tramontano, V.; Pavesi, G. Intraoperative neurophysiological monitoring in aneurysm clipping: Does it make a difference? A systematic review and meta-analysis. Clin. Neurol. Neurosurg. 2020, 196, 105954.