Development of a Machine Learning Tool to Improve Intraoperative Neurophysiological Monitoring: Proof of Concept

Varun Arvind, Omar Taha, Matthew Weintraub, Mehdi M Elfilali, Anil Mendiratta, Erik Williams, Miles Bomback, Michael G Vitale

INTRODUCTION:

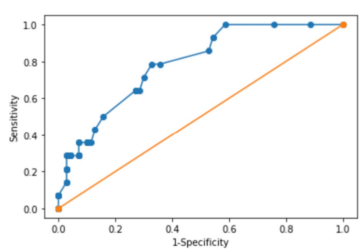
Intraoperative neuromonitoring (IONM) improves safety during pediatric spinal deformity surgery by providing real time neurophysiological assessment. However, current IONM interpretation is subject to variability and relies on human expertise. This study evaluates a machine learning (ML) algorithm designed to identify subtle changes in motor-evoked potentials (MEPs) that may precede neurological injury.

METHODS: IONM data from 84 pediatric spine surgeries were retrospectively analyzed. Fourteen patients experienced intraoperative MEP signal loss, with seven developing postoperative deficits. A ML model was trained on baseline MEPs from each patient to identify signal changes post-instrumentation. A “red-flag” alert was triggered if signal deviation exceeded 10% per minute. Model performance was assessed using sensitivity, specificity, predictive values, and accuracy, and compared against standard clinical alerts.

RESULTS: The model accurately identified cases as positive or negative in 58 out of 84 patients (accuracy: 69%). Among patients with IONM loss, red flag warnings were identified in 11 out of 14 patients. Five of six patients who had a postoperative neurological deficit were correctly identified by the ML model. ML-generated alerts preceded clinical IONM alerts by an average of 23.3 minutes. The ML model signaled an alert earlier than the IONM team in 71% (10/14) of the patients with IONM changes, at an average of 23.3 ± 23.5 minutes before the direct detectable change by the IONM team ($p = 0.0312$). Overall, there were 11 true positive, 47 true negative, 23 false positive, and 3 false negative cases. The sensitivity of the model was calculated as 78.6%, specificity 67.1%, positive predictive value (PPV) 32.4%, and negative predictive value (NPV) as 94.0%. The overall accuracy of the model was 69.0% with an area under the receiver operating curve (AUC) of 0.78

DISCUSSION AND CONCLUSION: This ML-based tool demonstrated promising early detection of intraoperative neuromonitoring changes, often preceding clinical alerts. Its high sensitivity and negative predictive value suggest potential for real-time support during pediatric spine surgery. Further validation in larger, prospective cohorts is warranted.

		Control (n = 70)	Positive (n = 14)	p-value
Age		15.1 ± 3.7	13.5 ± 3.2	0.112
Sex	Male	31	2	0.072
	Female	39	12	
Major Cobb		59.1 ± 13.8	76.8 ± 11.3	**<0.001**
Etiology	Idiopathic	40	3	0.096
	Neuromuscular	7	4	
	Congenital	6	4	
	Syndromic	13	3	
	Scheuermann’s Kyphosis	2	0	
	Spondylolisthesis	1	0	
	Other	1	0	
Surgery Type	Fusion	65	14	0.785
	MCGR Insertion	2	0	
	Shilla	2	0	
	TGR lengthening	1	0	



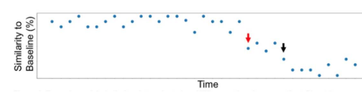


Figure 1. Example model similarity plot evaluated on MEPs over time for one patient. Blue dots correspond to the average accuracy of the model to identify muscle groups to baseline MEPs. The red arrow mark is the first abnormal signal detected by the ML model. The black arrow is the earliest change noted by the attending neuromonitoring team.