Modelling Techniques to Assess Surgeons' Cognitive Workload: A Systematic Review

Yifan Chen*, Arnab Majumdar, Niki Boyce and Jose Escribano

Abstract

Background: Excessive intraoperative cognitive workload can impair surgical performance and threaten patient safety. Machine learning (ML) offers potential for real-time objective monitoring, yet existing studies vary widely in design, input data, and modelling strategies. This systematic review synthesises current evidence on ML-based approaches to estimate surgeon intraoperative workload.

Methods: Searches of nine databases (inception-March 2025) followed PRISMA 2020. Eligible studies applied ML to model cognitive workload in surgical or simulated settings. Data were extracted on participants, input modalities, labelling strategies, preprocessing, modelling approaches, and performance metrics. Quality was assessed using the Mixed Methods Appraisal Tool (MMAT).

Results: Fifteen studies were included. Electroencephalography (EEG, 60%), electrocardiography (40%), and eye-tracking (40%) were used most frequently, with 60% adopting multimodal configurations. Workload was labelled using questionnaires (40%), task difficulty (40%), or hybrid approaches (13%). Most studies (87%) addressed classification tasks, with reported accuracy ranging from 54% to 99.9%. Both classical algorithms (e.g., support vector machines, random forests) and deep learning architecture (e.g., CNNs, LSTMs) achieved competitive results. However, all selected studies relied on small cohorts (fewer than 30 participants) in simulated environments, with inconsistent reporting of preprocessing framework.

Conclusion: ML-based cognitive workload modelling is feasible and shows strong performance. Adoption in clinical practice will require validated labelling frameworks, larger and more diverse cohorts, standardised preprocessing pipelines, and ecologically valid operating room datasets. Sensor practicality, comfort, and acceptability remain critical considerations for real-world deployment.

Keywords: Cognitive workload, machine learning, surgery, intraoperative monitoring

(ANNSURG 2025; 12: 1-3)

Department of Civil and Environmental Engineering, Imperial College London, London, UK

*Corresponding author: Yifan Chen, Department of Civil and Environmental Engineering, Imperial College London, London, UK. E-mail: yc5816@ic.ac.uk

Published Online: NA

DOI: 10.31487/j.ANN.2025.03.10

Appendix

TABLE 1: MMAT quality appraisal for included studies.

Study	Rep.	Meas. App.	Comp. Data	Conf. Ctrl	Int. Fid.
S1	Yes	Yes	Yes	No	Yes
S2	No	Yes	Yes	Yes	Yes
S3	Yes	Yes	Yes	No	Yes
S4	Yes	Yes	Yes	Yes	Yes
S5	Yes	Yes	Yes	No	Yes
S6	No	Yes	Yes	No	Yes
S7	Yes	Yes	Yes	Yes	Yes
S8	Yes	Yes	Yes	No	Yes
S9	No	Yes	Yes	No	Yes
S10	Yes	Yes	Yes	No	Yes
S11	No	Yes	No	No	Yes
S12	No	Yes	Yes	No	Yes
S13	No	Yes	Yes	No	Yes
S14	No	Yes	Yes	No	Yes
S15	No	Yes	Yes	No	Yes

Rep: Representativeness; Meas. App.: Measurement Appropriateness; Comp. Data: Complete Outcome Data; Conf. Ctrl: Confounder Control, Int. Fid.: Intervention Fidelity.

TABLE 2: Summary of sensor modalities used across the included studies.

Modality	Study IDs	Count	%
EEG	S2, S5, S6, S8, S9, S11, S12, S13, S14	9	60%
fNIRS	S3, S12, S13, S14	4	27%
ECG	S1, S9, S10, S13, S14, S15	6	40%
Eye-tracking	S2, S6, S7, S8, S12, S13	6	40%
EMG	S9, S13	2	13%
EDA	S9	1	7%
Posture	S10	1	7%
Hand Movement	S10	1	7%
Force sensors	S4	1	7%
Video	S4	1	7%
Motion	S9	1	7%

TABLE 3: Summary of cognitive workload labelling strategies used across the included studies.

Study ID	**Labelling Strategy**
S1	Continuous physiological measure (ECG)
S2	Task difficulty
S3	NASA-TLX

S4	SURG-TLX
S5	NASA-TLX
S6	SURG-TLX
S7	NASA-TLX
S8	Task difficulty
S9	NASA-TLX + Task difficulty + Average workload index
S10	SURG-TLX
S11	Task difficulty (NASA-TLX collected but not used for labelling)
S12	Task difficulty (SURG-TLX collected but not used for labelling)
S13	Task difficulty + SURG-TLX
S14	Task difficulty
S15	Task difficulty

TABLE 4: Summary of key performance metrics, modelling techniques, and modalities used across included studies.

Study ID	Performance Metric	Score	Model	Modality
S1	Correlation (ECG)	0.49	Autoregressive	ECG
S2	Accuracy	80.92%	Random Forest	EEG + Eye-tracking
S3	Accuracy	90%	SVM	fNIRS
S4	Accuracy	54.00%	FCNN	Motion, Force
S5	Accuracy	91%	CNN	EEG
S6	R-squared	0.82 ± 0.05	XGBoost	EEG + Eye-tracking
S7	Accuracy	84.70%	Na¨ıve Bayes	Eye-tracking
S8	Accuracy	77.90%	Neural Network	EEG + Eye-tracking
S9	Accuracy	83.20%	SVM	EEG + EDA + ECG + Motion + EMG
S10	Accuracy	76.50%	DT	ECG + Posture + Motion
S11	Accuracy	87%	LDA	EEG
S12	Accuracy	93.70%	1D-CNN	fNIRS + EEG + Eye-tracking
S13	Accuracy	97%	CNN-BiLSTM	fNIRS + EEG + EMG + ECG + Eye-tracking
S14	Accuracy	99.90%	LSTM	EEG + fNIRS + ECG
S15	Accuracy	91.26%	1D-CNN	ECG