have neurologic deterioration and 2 (3.6%) underwent further surgical intervention (hematoma evacuation). Clinical and radiological postoperative changes were highly related (p=5.16e-06), and the rate of EPCT being adverse without neurologic deficit, managed surgically, was 1/56 (1.8%). Conclusions: EPCT is routine practice. Given the low rate (1.8%) of repeat surgical intervention in the absence of neurologic deficit despite abnormal EPCT, omitting EPCT in neurologically intact patients may be warranted.

P.109

Machine learning based approach to improving the prediction of neurological deterioration in mild Degenerative Cervical Myelopathy

A Al-Shawwa (Calgary) M Craig (Calgary) K Ost (Calgary) S Tripathy (Calgary) D Cadotte (Calgary)*

doi: 10.1017/cjn.2024.212

Background: Degenerative cervical myelopathy (DCM) is the most common form of atraumatic spinal cord injury globally, yet clinical guidelines remain unclear on surgical recommendations for patients with mild forms of DCM. This is in part due to limitations in current MR imaging interpretation and complex mechanisms of neurological deterioration. Supervised machine learning (ML) models can help to identify clinical and imaging indicators of deterioration within mild DCM patients. Methods: 127 MRI scans (T2w, Diffusion Tensor Imaging, and Magnetization transfer scans) accompanied by a series of clinical tests underwent a semi-automated analysis to derive quantitative metrics. Random forest classifier, Support Vector Machine, and Logistic Regression models were trained and tested to predict 6-month neurological deterioration within patients. Results: The ML models performed, on average, better than previous studies with a balanced accuracy ranging between 70-75%. "Advanced" imaging metrics such as diffusion tensor imaging and magnetization transfer scans played an important role in improving model accuracy but only when used near the maximally compressed disc level, suggesting that limited yet targetted imaging metrics support ML model performance. Conclusions: The inclusion of specific, targeted imaging and clinical metrics support ML model performance in predicting neurological deterioration within mild DCM patients.

Neuroscience Education

P.110

Evaluating AI performance in written neurosurgery exams: a comparative analysis of large language models

E Guo (Calgary) R Sanguinetti (Calgary) R Ramchandani (Ottawa) S Lama (Calgary) GR Sutherland (Calgary)*

doi: 10.1017/cjn.2024.213

Background: The integration of Artificial Intelligence (AI) in medical education is an area of growing importance. While AI models have been evaluated extensively in multiple-choice question formats, their proficiency in written exams remains to be explored. Methods: Four AI models—GPT-4 (OpenAI), Claude-2.1 (Anthropic), Gemini Pro (Google), and Perplexity 70B (Perplexity)—were tested using the Canadian Royal College Sample Neurosurgery Exam. The written exam covered diagnostic reasoning, knowledge of neurosurgical conditions, and understanding of radiographic imaging techniques. Results: GPT-4 and Perplexity 70B both achieved a score of 68.42%, followed by Claude-2.1 with 60.53%, and Gemini Pro with 57.89%. The models showed proficiency in answering questions that required factual knowledge, such as identifying pathogens in spinal epidural abscess. However, they struggled with more complex diagnostic reasoning tasks, particularly in explaining the pathophysiology behind a sudden rise in blood pressure during surgery and interpreting radiographic characteristics of intracranial abscesses on MRI. Conclusions: The findings indicate that while AI models like GPT-4 and Perplexity 70B are adept at handling factual neurosurgical questions, their performance in complex diagnostic reasoning in a written format is less consistent. This underscores the need for more advanced and specialized AI training, particularly in the nuances of medical diagnostics and decision-making.

P.111

The surgeon experience of flow

SA McQueen (Toronto) M Hammond Mobilio (Toronto) A McParland (Vancouver) R Sonnadara (Hamilton) C Moulton (Toronto)*

doi: 10.1017/cjn.2024.214

Background: Cognitive flow has been linked with enhanced performance, career satisfaction, and decreased burnout. However, while elite sport has long trained athletes to enter flow states, the concept has not been adopted strongly in healthcare. Flow has primarily been explored from a unidimensional (cognitive) perspective. The present study sought to understand the experience of flow among surgeons through a multidimensional lens. Methods: Using a constructivist grounded theory methodology, semi-structured interviews were conducted with 19 staff surgeons at the University of Toronto, purposively sampled. Data were coded and analyzed iteratively by three researchers until theoretical saturation was achieved. Results: Although many surgeons had not previously heard of cognitive flow, the phenomenon deeply resonated with most. Participants described different physical, cognitive, emotional, sociocultural, and environmental components that interacted to shape the subjective experience of flow: "I think that there are many different facets of [flow] that don't always come together all at the same time, you may feel different parts of it… depending on what the kind of case is, who your help is, if you recently had a complication." (P4) Conclusions: Understanding flow in clinical practice may lead to new avenues for enhancing career satisfaction and promoting physician wellness.