

Neuroréanimation (pronostic, DVE, Sedation)

ID: 574

Evaluation of Artificial Language Model ChatGPT4 to predict 6-month outcome after Traumatic Brain Injury

C. Le barbey*(1), M.Giabicani(2), A.Sar(3), G.Bonnet(4), J.Moyer(5), C.Gakuba(6)

(1) AR, CHU CAEN, Caen, France , (2) AR, Beaujon, Clichy, France , (3) AR, CAEN, Caen, France , (4) MAD TEAM GREYC, Université Caen Normandie, Caen, France , (5) AR, CHU Caen Normandie, Caen, France , (6) Anesthésie Réanimation, CHU Caen Normandie, Caen, France

**Auteur présenté comme orateur*

Position du problème et objectif(s) de l'étude:

ChatGPT has demonstrated remarkable abilities in generating coherent and contextually relevant responses. As the popularity of ChatGPT and open artificial intelligence continues to grow, it is inevitable that patients and their families will seek answers from the AI model regarding disease prognoses. We aim to evaluate ChatGPT4's ability to predict 6-month functional prognosis in traumatic brain injury (TBI) patients admitted to our intensive care unit (ICU).

Matériel et méthodes:

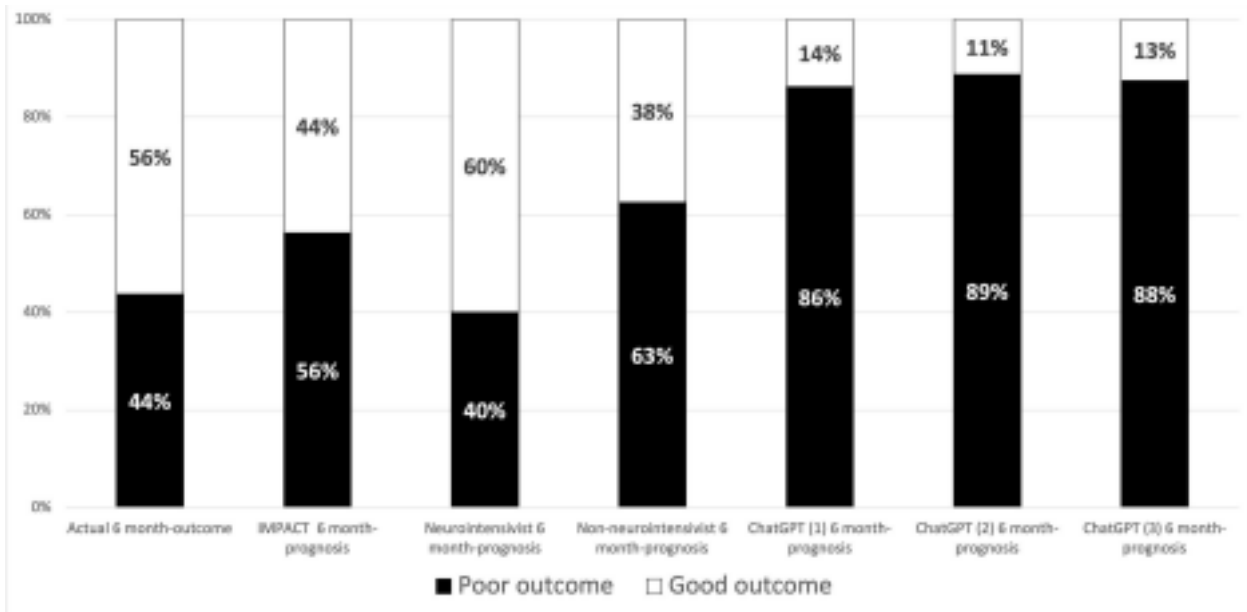
We analyzed the abilities of ChatGPT4's to predict 6-month outcome after TBI. TBI patients were consecutive included patients on a 20-month period in a prospective database. We also compared ChatGPT4's predictions for 6-month patients' outcome to those of the IMPACT score, of one neurointensivist and of one non-neurointensivist. Outcome was dichotomized in good (extend Glasgow outcome scale (GOSE 5-6-7-8) or poor (GOSE 1-2-3-4). Model performance was assessed using standard metrics such as sensitivity, specificity, and area under the curve (AUC-ROC). ChatGPT was run three times but only the first run for evaluation of the main outcome. Other secondary outcomes were assessed : probability of patients benefiting from an intracranial pressure transducer, external ventricular shunt, or craniectomy, the likelihood of pneumonia, and the likelihood of ethical discussions concerning treatment continuation.

Résultats & Discussion:

80 patients' prognosis admitted in ICU were analyzed. The AUC ROC for 6-month prognosis were : 0.622, 0.813, 0.703 and 0.706 respectively for ChatGPT, IMPACT, the neurointensivist, and the non-neurointensivist. To provide such prediction, the non-neurointensivist predicted more patients with poor outcome whereas the neurointensivist predicted more patients with good outcome than the real outcome. ChatGPT tended to predict more patient with poor prognosis than the reality (Figure).

Conclusion:

Our study highlights that ChatGPT4 prediction of 6-month outcome after TBI is less performant than the IMPACT model and that physicians. While the AI model performed on large medical database seems to be efficient to increase prediction, ChatGPT showed inferior accuracy in estimating functional outcomes. ChatGPT4 is a rising open access AI used for various purposes but it should not be used for TBI prediction. Considering the widespread use of open access AI, the limitations of AI-generated predictions including potential biases and ethical considerations warrant further explorations.



Les auteurs déclarent ne pas avoir toute relation financière impliquant l'auteur ou ses proches (salaires, honoraires, soutien financier éducationnel) et susceptible d'affecter l'impartialité de la présentation.