Unstructured data

- Performance of multimodal prediction models for intracerebral hemorrhage outcomes using real-world data
- Using Generative AI to Extract Structured Information from Free Text Pathology Reports
- Generative AI Models in Time-Varying Biomedical Data: Scoping Review
- Improving rupture status prediction for intracranial aneurysms using wall shear stress informatics
- Systematic Review of Experimental Deep Brain Stimulation in Rodent Models of Epilepsy
- An Accurate and Rapidly Calibrating Speech Neuroprosthesis
- Artificial intelligence and headache
- Portrait of mental health identified by people with the post-covid syndrome

---

Unstructured data (or unstructured information) is information that either does not have a pre-defined data model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or annotated (semantically tagged) in documents.

Accurate prediction of outcomes among patients in intensive care units (ICUs) is important for clinical research and monitoring care quality. Most existing prediction models do not take full advantage of the electronic health record, using only the single worst value of laboratory tests and vital signs and largely ignoring information present in free-text notes. Whether capturing more of the available data and applying machine learning and natural language processing (NLP) can improve and automate the prediction of outcomes among patients in the ICU remains unknown.

Marafino et al., evaluated the change in power for a mortality prediction model among patients in the ICU achieved by incorporating measures of clinical trajectory together with NLP of clinical text and asessed the generalizability of this approach.

This retrospective cohort study included 101 196 patients with a first-time admission to the ICU and a length of stay of at least 4 hours. Twenty ICUs at 2 academic medical centers (University of California, San Francisco [UCSF], and Beth Israel Deaconess Medical Center [BIDMC], Boston, Massachusetts) and 1 community hospital (Mills-Peninsula Medical Center [MPMC], Burlingame, California) contributed data from January 1, 2001, through June 1, 2017. Data were analyzed from July 1, 2017, through August 1, 2018.

In-hospital mortality and model discrimination as assessed by the area under the receiver operating characteristic curve (AUC) and model calibration as assessed by the modified Hosmer-Lemeshow statistic.

Among 101 196 patients included in the analysis, 51.3% (n = 51 899) were male, with a mean (SD) age of 61.3 (17.1) years; their in-hospital mortality rate was 10.4% (n = 10 505). A baseline model using only the highest and lowest observed values for each laboratory test result or vital sign achieved a cross-validated AUC of 0.831 (95% CI, 0.830-0.832). In contrast, that model augmented with measures of clinical trajectory achieved an AUC of 0.899 (95% CI, 0.896-0.902; P < .001 for AUC difference). Further augmenting this model with NLP-derived terms associated with mortality further increased the AUC to 0.922 (95% CI, 0.916-0.924; P < .001). These NLP-derived terms were associated with improved model performance even when applied across sites (AUC difference for UCSF: 0.077 to 0.021; AUC difference for MPMC: 0.071 to 0.051; AUC difference for BIDMC: 0.035 to 0.043; P < .001) when augmenting with NLP at each site.

Intensive care unit mortality prediction models incorporating measures of clinical trajectory and NLP-derived terms yielded excellent predictive performance and generalized well in this sample of hospitals. The role of these automated algorithms, particularly those using unstructured data from notes and other sources, in clinical research and quality improvement seems to merit additional investigation [1).].

[1)]
Marafino BJ, Park M, Davies JM, Thombley R, Luft HS, Sing DC, Kazi DS, DeJong C, Boscardin WJ, Dean ML, Dudley RA. Validation of Prediction Models for Critical Care Outcomes Using Natural Language Processing of Electronic Health Record Data. JAMA Netw Open. 2018 Dec 7;1(8):e185097. doi: 10.1001/jamanetworkopen.2018.5097. PubMed PMID: 30646310.