Approaches to Address Missingness in Electronic Health Records

The solution is a machine learning model that accurately simulates recurrence in patients where frequent testing data is not available in their electronic health records.

What is the Problem?

Electronic health records (EHRs) contain important patient history information, such as biomarkers, that can reveal the relationship between biomarkers and future outcomes. EHRs can be used to facilitate predicting patients' future health risks, but the lack of frequent biomarker testing hinders the use of EHRs for risk prediction. Intermittent patient follow-up leads to sporadically measured patient data and thus delayed outcome ascertainment. As a result, there is a need to develop approaches to address the missingness in EHRs.

What is the Solution?

The solution is a machine learning model that accurately simulates recurrence in patients where frequent testing data is not available in their electronic health records. The model uses a two-stage statistical approach for pre-processing patient data to create a complete dataset that contains biomarker measurements and outcome ascertainment every three months. The technology recreates data on a frequent testing schedule that could be used for risk prediction by imputing both the specific recurrence outcome and the longitudinal biomarker values using computational models.

What is the Competitive Advantage?

The competitive advantage of this technology lies in its ability to create synthetic data to augment real data, improving the development of risk prediction algorithms. The use of this machine learning model can help address the missing patient data due to lack of frequent surveillance testing in EHRs. This technology is novel in that it couples the imputation of longitudinal biomarker data and the recurrence outcome to preserve the relationship between predictors and outcomes.

References

1. Rodriguez, P. J., Heagerty, P. J., Clark, S., Khor, S., Chen, Y., Haupt, E., Hahn, E. E., Shankaran, V., & Bansal, A.(2023) , https://ascopubs.org/doi/full/10.1200/CCI.23.00066, https://ascopubs.org/journal/cci, 7, e2300066

Technology ID

BDP 8778

Category

Selection of Available Technologies

Diagnostic

Authors

Aasthaa Bansal

Learn more