Team H2
Member 1: Jie Feng CHAI
GTID: 903842323
gt email: jchai41

Member 2: Man Kong SIT
GTID: 903760158
gt email: msit6

Summary

Paper1: SurvTRACE: Transformers for Survival Analysis with Competing Events

1) Task: Hazard & survival function approximation.
2) Innovation: Using a transformer model to handle competing events (assumption that events observed in EHR are not independent) and censored data (lack of follow up data).
3) Dis/Adv: Consistently outperforms other models (although metric is not specified in paper). Model includes an interpretable attention mechanism (automatic feature engineering and adds model interpretability).
4) Data Accessibility: Yes (https://www.cbioportal.org/study/summary?id=brca_metabric, https://sahirbhatnagar.com/casebase/reference/support.html, https://seer.cancer.gov/data/)
5) Code Accessibility: https://github.com/RyanWangZf/SurvTRACE

Paper 2: Learning of Cluster-based Feature Importance for Electronic Health Record Time-series

1) Task: Predicting patient outcome (multi-class prediction).
2) Innovation: Mapping patient time-series features into a latent representation and group into clusters representation vectors. The cluster representation is then used for patient outcome prediction.
3) Dis/Adv: Better clinical interpretability and able to handle imbalance outcome distribution.
4) Data Accessibility: Partial (https://physionet.org/content/mimic-iv-ed/2.2/)
5) Code Accessibility: Code is not provided by author

Paper 3: Unified Auto Clinical Scoring (Uni-ACS) with Interpretable ML models

1) Task: Translating ML models into clinical scores
2) Innovation: Using explainable outputs (SHAP) from ML models
3) Dis/Adv: Translated scores keep more of the ML models' predictive performance compared to baseline. Framework is also model agnostic. Yet, it's not clear if it works for other diseases and can't be fully automated.
4) Data Accessibility: Yes (MIMIC-III Clinical Database v1.4 (physionet.org), MIMIC-IV v2.0 (physionet.org))
5) Code Accessibility: https://github.com/llja0112/Uni-ACS

Target Paper Decision

– Which paper in the candidate you will replicate?

Paper 3

– Why did you choose the paper?

The output of this paper, clinical scores translated from ML models, bridges the gap between data science academia and the medical industry. The clinical scores format is something medical professionals are familiar with (Walker & Habboushe, 2022). This makes the insights from ML models more accessible and practical for real-world medical scenarios.

– What are the specific hypotheses from the paper that you plan to verify in your reproduction study?

One of the hypotheses that we plan to verify is whether the clinical score, when translated from ML models, retains a higher level of predictive performance compared to the baseline clinical score, as evidenced by a lower drop in AUROC (2.44% versus 5.79%).

– Briefly state how you are assured that you can obtain appropriate data and computational resources including software and hardware demanded in the paper.

The datasets are publicly accessible. The required Python libraries specified in the code are available. If our machines can't handle the computational demands, we have the option to use cloud resources like Colab GPU or AWS.

References

Wang, Z., & Sun, J. (2022). SurvTRACE: Transformers for Survival Analysis with Competing Events. https://arxiv.org/pdf/2110.00855.pdf

Aguiar, H., Santos, M., Watkinson, P., & Zhu, T.. (2022). Learning of Cluster-based Feature Importance for Electronic Health Record Time-series. https://proceedings.mlr.press/v162/aguiar22a.html

Li, A., Ong, M.L., Oei, C.W., Lian, W., Phua, H.P., Htet, L.H., & Lim, W.Y.. (2022). Unified Auto Clinical Scoring (Uni-ACS) with Interpretable ML models. 182:26-53 Available from https://proceedings.mlr.press/v182/li22a.html

Walker, G., & Habboushe, J. (2022) About us. URL https://www.mdcalc.com/about-us