Enhancing early prediction of heart failure: Hidden Markov model with covariates

Luca Brusa[1], **Fulvia Pennoni**[1], **Francesco Bartolucci**[2], **Laura Pierantozzi**[1]

[1] *University of Milano-Bicocca, Department of Statistics and Quantitative Methods, Italy, luca.brusa@unimib.it, fulvia.pennoni@unimib.it, l.pierantozzi@campus.unimib.it*

[2] *University of Perugia, Department of Economics, Italy, francesco.bartolucci@unipg.it*

Abstract

The clinical assessment of heart failure severity is commonly guided by the New York Heart Association (NYHA) scale, which stratifies patients into four risk classes based on observed symptoms and physician judgment. Despite its widespread adoption and clinical utility, the NYHA scale is inherently empirical, introducing potential variability due to subjective interpretation. Moreover, although heart failure is typically diagnosed in the symptomatic phase, there are earlier asymptomatic stages in which patients may already exhibit cardiac alterations at risk of evolving towards the full-blown disease.

We propose a hidden Markov (HM) model (Bartolucci, Farcomeni, Pennoni, 2013) tailored for the analysis of longitudinal data, assuming that the observed clinical data of each patient are driven by a latent process following a first-order Markov chain. This approach enables the identification of a finite number of unobservable states, corresponding to subpopulations of patients with varying levels of heart failure risk (Pandolfi, Bartolucci, Pennoni, 2023). By modeling the transitions between these latent states over time, the HM model provides a dynamic representation of disease progression, offering critical insights into the temporal evolution of heart failure severity. A key objective of this proposal is to evaluate the classification capability of the HM model related to two primary outcomes of interest, such as hospital readmission and death. Specifically, we aim to determine whether the data-driven clustering of patients provided by the HM model aligns with, complements, or enhances the traditional risk assessment process. To investigate the development of advanced early warning systems for heart failure we also evaluate the predictive performance of HM models estimated with the inclusion of demographic, clinical, and comorbidity covariates. This predictive dimension is fundamental for the early detection of disease onset or worsening, enabling timely and targeted interventions that may mitigate hospitalizations and improve long-term outcomes.

To validate the proposal, we compare the results with recent analyses of a well-known historical dataset frequently cited in the literature, which originates from a comprehensive study conducted at a Chinese hospital (Zhang et al., 2021). The data are publicly available on the PhysioNet platform and have been previously used in several studies; see, among the others, Rahman et al. (2023) and Tong, Zhu, Ling (2023). They comprise 2,008 patients hospitalized for heart failure between December 2016 and June 2019, considering all forms of conditions: acute, chronic, left-sided, right-sided, or mixed. Subsequent hospitalizations and mortality data were obtained during compulsory follow-up evaluations at 28 days, 3 months, and 6 months. Information on medications administered during hospitalization are also collected, as well as 166 covariates capturing a broad spectrum of demographic, clinical, and laboratory measurements.

Keywords

Discrete latent variable models, Early warning systems, Hospital readmission and mortality, Longitudinal data

References

Bartolucci, F., Farcomeni, A. and Pennoni, F. (2013). Latent Markov Models for Longitudinal Data, Chapman and Hall/CRC, Boca Raton.

Ben-Assuli, O., Heart, T., Vest, J.R., Ramon-Gonen, R., Shlomo, N. and Klempfner, R. (2021). Profiling readmissions using hidden Markov model - the case of congestive heart failure. *Information Systems Management*, *38*, 237–249.

Pandolfi, S., Bartolucci, F. and Pennoni, F. (2023). A hidden Markov model for continuous longitudinal data with missing responses and dropout. *Biometrical Journal*, *5*, 1–28.

Rahman, S., Rahman, H.R., Prithula, J., Chowdhury, M.E.H., Ahmed, M.U., Kumar, J., Murugappan, M. and Khan, M.S. (2023). Heart failure emergency readmission prediction using stacking machine learning model. *Diagnostics*, *13*, 1–20.

Tong, T., Zhu, Z., Ling, J. (2023). Comparison of linear and non-linear machine learning models for time-dependent readmission or mortality prediction among hospitalized heart failure patients. *Heliyon*, *9*, 1–11.

Zhang, Z., Cao, L., Chen, R., Zhao, Y., Lv, L., Xu, Z. and Xu, P. (2021). Electronic healthcare records and external outcome data for hospitalized patients with heart failure. *Scientific Data*, *8*, 1–6.