Developing a Probabilistic Graphical Model for Heart Failure Prediction

Joshua Freeze and Sakin Kirti
CSDS 491: Probabilistic Graphical Models
Dr. Michael Lewicki

Introduction

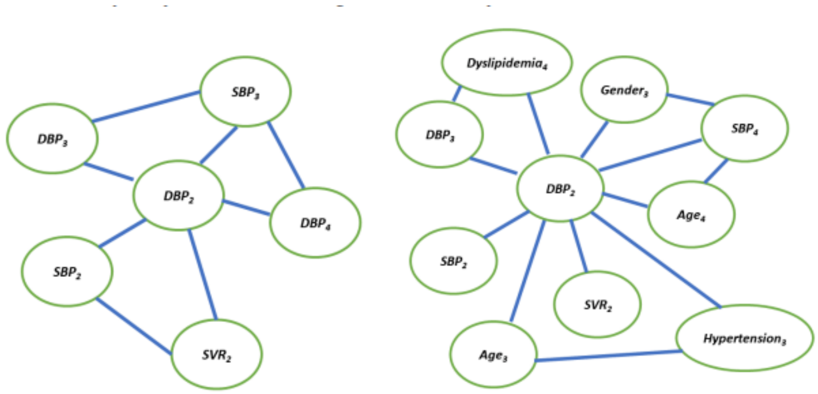
Heart failure is a serious medical condition that affects millions of people worldwide. Early detection of heart failure can help improve patient outcomes and reduce healthcare costs. In this project, we aim to develop a probabilistic graphical model for heart failure prediction using clinical data. Depending on the available time, we hope to generate a Kaplan-Meier curve to demonstrate the risk of heart failure as time progresses.

Objectives

1. Develop a probabilistic graphical model for heart failure prediction.
2. Evaluate the performance of the model on a real-world clinical dataset.
3. Compare the performance of the model with several other machine learning methods.
4. Investigate the impact of different features on the model's performance.

Methodology

1. Data collection: We have access to a dataset through Josh's independent research. We have received permission from Josh's PI to use the dataset to explore the creation of a probabilistic graphical model to predict heart failure. This data comes from patients at UH that received CTA scans and includes data on whether or not they had a Heart Failure event and at what time. The final dataset is composed of 2176 different patients with 66 different features relating to the patient image information and another 24 relating to general risk factors.
2. Data preprocessing: We will preprocess the data to handle missing values, normalize the data, and remove any irrelevant features. Currently, the dataset contains over one hundred features. Reducing the number of features will make the dataset easier to work with, more within the scope of this project, make data cleaning easier, and allow us to create and explore true causal relationships in the data.
3. Model development: We will develop a probabilistic graphical model for heart failure prediction using the preprocessed data [4]. The network will be composed of our clinical data and we will add the relationships between factors to create our graph structure.



(a) Single data source (b) Multi-data source

Figure 1. Graph structure created by [2]

The nodes and connections will initially be chosen based on the relevant literature but we also plan to have the graph learn the connections between nodes as the model trains resulting in patient classification.

4. Model evaluation: We will evaluate the model's performance using standard machine learning metrics such as accuracy, precision, recall, and F1-score.
5. Comparison with other methods: We hope to compare the model performance against a crude probabilistic model, a simple logistic regression model (without final thresholding), and a simple deep neural network [2[. We plan to base this comparison on this literature review though retrained for our dataset [1]. We also plan to make a comparison of the time-to-event results of heart failure compared to ASCVD, the current method for identifying heart failure risk.
6. Feature importance analysis: We will investigate the impact of different features on the model's performance using feature importance analysis techniques such as permutation feature importance, floating window with adaptive size for feature elimination (FWAFE), elastic-nets and SHapley Additive exPlanations (SHAP) [3].

Expected Outcomes

1. A probabilistic graphical model for heart failure prediction that achieves high accuracy and precision above current model indicators created by the models in [1]. Our ideal outcome includes the generation of a Kaplan-Meier Curve which would allow clinicians to assess how the risk of HF increases with time.
2. A comparison of the performance of our model with several other machine learning methods.
3. An analysis of the impact of different features on the model's performance.

Conclusion

In this project, we aim to develop a probabilistic graphical model for heart failure prediction using clinical data. We expect that the proposed model will outperform several other machine learning methods and provide insights into the features that are most important for heart failure prediction. This research has the potential to improve patient outcomes and reduce healthcare costs by enabling early detection and potentially, more accurate care methods for heart failure.

Roles and responsibilities

Josh - Use the relevant literature and physician knowledge to assist in feature selection. Replicating the work of a paper to imitate a graphical network for this application. Add in SMOTE or other techniques to maximize the results of the model

Sakin - Replicate the work of a paper to imitate a graphical network for this application. Create a deep learning reference to ensure that the model can be adequately compared to the current model.

Sources

[1] Jing Wang 2021 *J. Phys.: Conf. Ser.* 2031 012068

[2]Predicting Physiological Response in Heart Failure Management: A Graph Representation Learning Approach using Electronic Health Records, Shaika Chowdhury, Yongbin Chen, Andrew Wen, Xiao Ma, Qiying Dai, Yue Yu, Sunyang Fu, Xiaoqian Jiang, Nansu Zong doi: https://doi.org/10.1101/2023.01.27.23285129

[3] Ashir Javeed, Sanam Shahla Rizvi, Shijie Zhou, Rabia Riaz, Shafqat Ullah Khan, Se Jin Kwon, "Heart Risk Failure Prediction Using a Novel Feature Selection Method for Feature Refinement and Neural Network for Classification", Mobile Information Systems, vol. 2020, Article ID 8843115, 11 pages, 2020. https://doi.org/10.1155/2020/8843115

[4] Bandyopadhyay, Sunayan. “Cardiovascular Risk Prediction from Electronic Health Records Using Probabilistic Graphical Models.” University of Minnesota, June 2016, https://conservancy.umn.edu/bitstream/handle/11299/182297/Bandyopadhyay_umn_0130E_17195.pdf?sequence=1.