

M.Sc. Project Proposal

Souradeep Sen
Department of Computer Science,
University of Exeter
(Dated: April, 2023)

The study aims to compare the performance of a Hybrid Deep Learning (HDL) architecture for predicting mortality and hospitalization in heart failure patients with frailty, against traditional survival analysis techniques. HDL uses unsupervised and supervised layers to estimate prediction probabilities based on low dimensional and contextual/historic features from clinical data. By leveraging longitudinal patient data made available in the form of Electronic Health Records (EHR), a new perspective is sought on the performance of machine learning risk prediction models compared to conventional survival analysis in HF patients.

I. INTRODUCTION

Heart failure (HF) is a complex clinical syndrome that affects millions of individuals worldwide, particularly those with frailty, a common condition in older adults. Predictive models using machine learning (ML) based approaches have shown promise in identifying patients at high risk of adverse outcomes. In this project, we aim to develop and validate predictive models to identify HF patients with frailty who are at increased risk of mortality and hospitalization using EHR data.

II. RESEARCH CONTEXT

Traditional survival analysis has been used extensively to predict mortality in this patient population. Deep learning methods have also been employed in (Gjoreski et al. 2020), (Nirschl et al. 2018), (Wang et al. 2019), (Ayala Solares et al. 2020) and (Lorenzoni et al. 2019). However, limited work has been done in predicting mortality and hospitalization in HF patients with frailty, especially using electronic health records (EHR) data.

1. Who is a ‘frail’ patient?

Our target pool of patients being those suffering from frailty, one first needs to understand what the term ‘frail’ refers to. Although its meaning may be nebulous to some extent, the eFI (Electronic Frailty Index) is a good clinical approximation of the condition. The eFI is based on the cumulative deficit model of frailty (eng 2023), which proposes that frailty is the result of a build-up of health deficits or impairments. It is calculated using a set of clinical indicators, such as the presence of chronic conditions, cognitive impairment, and mobility problems, that are commonly documented in EHRs.

2. Investigating Heart Failure

HF can be influenced by lifestyle risk factors, such as diet, exercise, smoking, and alcohol consumption. Clinical factors such as diabetes, hypertension, BMI, cholesterol, and sugar consumption were also identified as potential risk factors for HF. Familial history may not be pertinent for all cases of HF, but may only be relevant for some cases - familial dilated cardiomyopathy (DCM) is a type of heart failure that can have a strong genetic component (Peters et al. 2020).

3. Past Work

Several past papers have addressed predictive modeling for heart failure patients. (Che et al. 2017) proposes a deep neural network model with learned medical feature embedding to address high dimensionality and temporality in electronic health record (EHR) data. They use a convolutional neural network to capture non-linear longitudinal evolution of EHRs and local temporal dependency for risk prediction, and embed medical features to account for high dimensionality. Experiments show promising results in predicting risks for congestive heart failure. (Suo et al. 2017) focuses on personalized predictive modeling and aims to build specific models for individual patients using similar patient cohorts to capture their specific characteristics. According to this study, although CNNs have shown promise on measuring patient similarity, one disadvantage is that they could not utilize temporal and contextual information of EHRs. To measure patient similarity using EHRs, a time-fusion CNN framework was proposed. Within this framework, the vector representation was generated for each patient, which was then utilized for measuring patient similarity and personalized disease prediction. Dynamic updates to a CNN model are explored in (Brand et al. 2018) as more data is gathered over time - this architecture lends itself well to real-time mortality risk prediction. Maintaining interpretability across deep learning models is explored in (Caicedo-Torres and Gutierrez 2019). Many previous studies using machine learning for modeling the risk of

HF in patients have focused on discretized outputs. This study aims to consider incidences as time-to-event. The use of EHR such as those available in CPRD, offers an opportunity to access comprehensive longitudinal data, which captures the entire cycle of a patient’s diagnosis and treatment. Limited work has been done on predictive modeling for HF patients with frailty, particularly using hybrid deep learning and Hidden Markov Models. This project aims to bridge this gap in the literature.

III. AIMS & OBJECTIVES

A. Research Questions:

1. To what extent can HDL predict HF as time-to-event?
2. How transparent/ interpretable can the HDL model be for prediction of HF?
3. What are the trade-offs between HMM and HDL with respect to interpretability and performance?

B. Aims:

This project aims to develop a scalable predictive model for frail patients with heart failure, combining deep neural networks with statistical machine learning. The model can enable continuous risk prediction for hospitalization and mortality, addressing a critical need in patient care. The results of this project could improve patient outcomes by identifying those at high risk and facilitating targeted interventions. The ML approach relies on hybrid deep learning, specifically an architecture that combines statistical learning in the lower unsupervised layers and convolutional neural networks on the top layer. A secondary aim is to compare the hybrid deep learning models with a more mathematically transparent model, Hidden Markov Models (HMMs). If time permits, I will also assess the trade-off between model performance and interpretability to determine the most suitable model for this task.

IV. DATA & RESOURCES

CPRD (Clinical Research Practice Datalink) data is likely to be used for this project. From a collaboration with the medical school at the University of Exeter, I am being granted access to a set of data containing EHR records of a vast number of patients, consisting of both HF patients and non-HF patients.

A. Data Sources

1. Primary Data Source: CPRD Aurum

The Clinical Practice Research Datalink (CPRD) is a large, longitudinal UK primary care database containing anonymized electronic health records. It includes information on patient demographics, diagnoses, symptoms, prescriptions, and referrals, among other clinical data. The database is widely used for observational research, including epidemiological studies, drug safety monitoring, and healthcare utilization analysis. CPRD is maintained by the Medicines and Healthcare products Regulatory Agency (MHRA) and is available to researchers and industry partners who meet certain criteria and obtain appropriate approvals.

The Clinical Practice Research Datalink (CPRD) Aurum is a database provided by EMIS Health. It includes data for approximately 7 million active patients, and covers consenting practices in England, with consenting practices from Northern Ireland available from 2019. CPRD Aurum is updated monthly, and it is representative of the broader English population in terms of geographical spread, age, and gender. However, it does not natively record data outside of primary care. It is important to be aware of this limitation. To bypass it, the CPRD Aurum database can be linked to datasets such as Hospital Episode Statistics (HES), Death Registration, Cancer data, Mental Health Services Dataset, and Small Area-Level Data.

2. Contingent Data Source: MIMIC-III

The MIMIC-III (Medical Information Mart for Intensive Care III) dataset is a large, freely available critical care database that contains de-identified health data from over 40,000 patients admitted to the Beth Israel Deaconess Medical Center ICU between 2001 and 2012. It includes over 2 million ICU admissions and is one of the most comprehensive ICU databases available. The data is diverse, including demographic information, vital signs, laboratory tests, medications, and more. The dataset also includes information about procedures, diagnoses, and outcomes.

The MIMIC-III dataset has been widely used for research in various fields, such as machine learning, clinical decision support, and epidemiology. It is particularly useful for the development and testing of predictive models and algorithms, as well as for investigating clinical phenomena and interventions. The dataset has contributed to numerous studies, and many of its findings have led to improved patient care and clinical decision-making.

The database has been made publicly available to researchers around the world, but access is restricted and requires a formal application process. This is to ensure the privacy and security of patient data, as well as to maintain the integrity of the dataset. Researchers who wish to use the dataset must complete a certification course on data privacy and ethics before gaining access.

B. Coding

There are several possible tools of choice for the project. The ones I plan to use are Python for exploratory analysis, coding the base unsupervised layers and upper supervised layers as well as encoding contextual and historical data. This is because it has many useful libraries and frameworks at its disposal for numerical computing and neural networks, namely numpy, scipy, Pytorch and Tensorflow. For the more statistically inclined approach of Hidden Markov Models, I plan to use R as it has native functionalities for building Markov chains as well as packages like `brms` and `rstan` for Bayesian simulation methods like Markov Chain Monte Carlo.

V. METHODS & EXPERIMENT DESIGN

A. Model Building

To evaluate the performance of the hybrid deep learning models and Hidden Markov Models developed in this project, I will compare them to traditional survival analysis models such as Cox proportional hazards and Kaplan-Meier curves. These models are widely used in predicting outcomes in heart failure patients, providing a benchmark for ML-based models. I will assess the performance of our models based on metrics such as accuracy, sensitivity, and specificity. The comparison will allow us to determine whether the ML-based models outperform traditional models, providing insights into the potential of these models for predicting outcomes in frail heart failure patients.

1. Primary Aim: CNN

The idea is to parse through time-series data of individual patient histories and identify signals that can discriminate between potential HF and non-HF patients. These signals can then be encoded from their high-dimensional space into a latent vector representation in lower dimensions - possible techniques for dimensionality reduction such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) may be used for this purpose. These latent representations can then be fed into supervised layers of a Convolutional Neural Network. The output from the network should be an array depicting the persistency of a patient till an event of interest - in

our case, time-to-hospitalization or time-to-death. Separate hold-out sets can be reserved for testing purposes. This will be more favourable against a discretized or aggregate output as it will allow more direct comparison with traditional survival analysis like multivariate Cox regression (Lin 1994).

2. Secondary Aim: HMM

The secondary aim would be to use either the multivariate time series of observed biosignals or their latent vector representation $\{o_1, \dots, T\}$ as observed states in a Hidden Markov Model $\lambda = (A, B, \pi)$ ¹. Algorithms for likelihood estimation (Forward Algorithm), decoding (Viterbi Algorithm) and learning (Baum-Welch) will be used to derive ideal parameters for the model and then extract most probable sequence of hidden states given the observations (Jurafsky and Martin 2000). The hidden states here can correspond to either a binary set $S = \{1, 2\}$ corresponding to `alive` and `deceased` with their individual probabilities at time t allowing us to understand the onset of HF and how it may change over a sequence of observed events.

B. Explainability & Interpretability

In the medical domain, model interpretability is crucial to ensure that the predictions made by the model can be trusted and applied in a clinical setting. Although ML-based models may provide higher accuracy, their black-box nature can make them less interpretable, and therefore cast reasonable doubt on their applicability. Therefore, it is important to assess the trade-off between model performance and model interpretability. One approach to assessing this trade-off is to use model-agnostic methods such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), which provide insights into how the model arrived at its prediction. I plan to use these methods to explain the predictions made by the models and compare the level of interpretability with traditional survival analysis models. I can also assess the performance of the models against the level of interpretability, enabling me to evaluate the trade-off between model performance and interpretability. The results will provide guidance on the most suitable model for predicting outcomes in frail heart failure patients in a clinical setting.

¹ A and B are respectively the state transition matrix and the observation (emission) matrix

VI. DATA GOVERNANCE & ETHICS

Ethics approval is an essential requirement for this research project as it involves patient data. Patient data is considered sensitive information and approval ensures that this data is handled with care and confidentiality. For the purpose of this project, the medical school at the University of Exeter has consented to share CPRD Aurum data that they have procured for their own research. Ethical approval is being sought from University of Exeter's management system, WorkTribe. Access to the contingent data source, MIMIC-III is restricted and requires a formal application process to maintain data privacy and integrity. To comply with this, I have undergone the CITI training program for Data or Specimens Only Research.

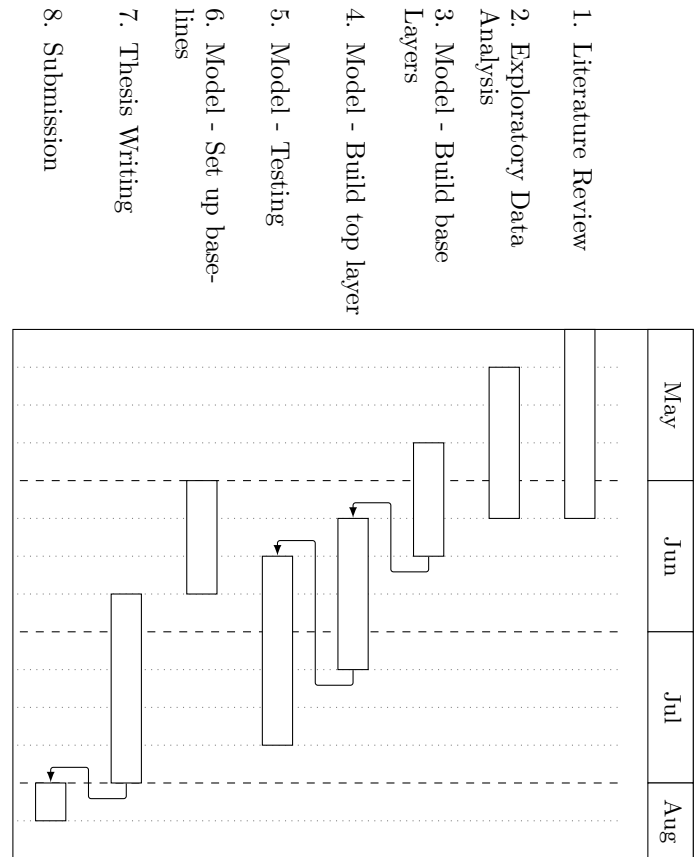
VII. RISK ASSESSMENT

There are a few potential risks to consider before embarking on this project. Some mitigation strategies have already been put into place.

1. Because of the sensitivity of the CPRD data, it involves some degree of navigating through bureaucratic processes of data access and retrieval. There is a chance that the data does not become available within a favourable time period. To mitigate this, access to the open-source dataset MIMIC-III has already been secured.
2. As stated earlier, this work deals with patient data (albeit anonymized) and hence, ethics approval is necessary. As of writing this proposal however, this is still pending due to unforeseen technical glitches on the University website that is barring me from getting access to the platform. However, talks are ongoing with the University IT department so that this may be resolved at the earliest.

VIII. PROJECT PLAN

The project plan is outlined by the following chart.



IX. CONCLUSION

This project will leverage CPRD's rich data source to develop a robust predictive model for hospitalization and mortality in frail heart failure patients, with the ultimate goal of assessing the trade-off between performance and interpretability arising from the two different paradigms of HDL and HMM.

1. (2023). Electronic Frailty Index (EFI). <https://www.england.nhs.uk/ourwork/clinical-policy/older-people/frailty/efi/>. Accessed on April 24, 2023.
2. Ayala Solares, J. R., Diletta Raimondi, F. E., Zhu, Y., Rahimian, F., Canoy, D., Tran, J., Pinho Gomes, A. C., Payberah, A. H., Zottoli, M., Nazarzadeh, M., Conrad, N., Rahimi, K., and Salimi-Khorshidi, G. (2020). Deep learning for electronic health records: A comparative review of multiple deep neural architectures. *Journal of Biomedical Informatics*, 101:103337.
3. Brand, L., Patel, A., Singh, I., and Brand, C. (2018). Real time mortality risk prediction: A convolutional neural network approach. In *HEALTHINF*, pages 463–470.
4. Caicedo-Torres, W. and Gutierrez, J. (2019). Iseeu: Visually interpretable deep learning for mortality prediction inside the icu. *Journal of biomedical informatics*, 98:103269.
5. Che, Z., Cheng, Y., Sun, Z., and Liu, Y. (2017). Exploiting convolutional neural network for risk prediction with medical feature embedding. *arXiv preprint arXiv:1701.07474*.
6. Gjoreski, M., Gradišek, A., Budna, B., Gams, M., and Poglajen, G. (2020). Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. *IEEE Access*, 8:20313–20324.

7. Jurafsky, D. and Martin, J. H. (2000). Speech and language processing.
8. Lin, D. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*, 13(21):2233–2247.
9. Lorenzoni, G., Sabato, S. S., Lanera, C., Bottigliengo, D., Minto, C., Ocagli, H., De Paolis, P., Gregori, D., Iliceto, S., Pisanò, F., and et al. (2019). Comparison of machine learning techniques for prediction of hospitalization in heart failure patients. *Journal of Clinical Medicine*, 8(9):1298.
10. Nirschl, J. J., Janowczyk, A., Peyster, E. G., Frank, R., Margulies, K. B., Feldman, M. D., and Madabhushi, A. (2018). A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of h&e tissue. *PloS one*, 13(4):e0192726.
11. Peters, S., Johnson, R., Birch, S., Zentner, D., Hersherberger, R. E., and Fatkin, D. (2020). Familial dilated cardiomyopathy. *Heart, Lung and Circulation*, 29(4):566–574.
12. Suo, Q., Ma, F., Yuan, Y., Huai, M., Zhong, W., Zhang, A., and Gao, J. (2017). Personalized disease prediction using a cnn-based similarity learning method. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 811–816. IEEE.
13. Wang, L., Sha, L., Lakin, J. R., Bynum, J., Bates, D. W., Hong, P., and Zhou, L. (2019). Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients With Dementia for Earlier Palliative Care Interventions. *JAMA Network Open*, 2(7):e196972–e196972.