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 (HF) patients with frailty, against traditional survival analysis techniques. HDL uses unsupervised and supervised layers to estimate prediction probabilities based on contextual and 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

HF is a significant cause of morbidity and mortality worldwide, and accurately predicting risk is crucial for improving patient outcomes. Traditional survival analysis techniques have limitations in handling complex clinical data and accounting for dynamic patient characteristics over time. While deep learning (DL) techniques have shown promise in handling EHRs, there is still room for improvement in better explaining the models since interpretability is a critical aspect for model building in the healthcare domain. This study aims to compare Hybrid Deep Learning (HDL) architectures against traditional survival analysis techniques for predicting mortality and hospitalization in heart failure (HF) patients with frailty. This study will attempt to validate the hypothesis that HDL outperforms traditional survival analysis in terms of prediction accuracy using real-world EHR data. The findings may have important implications for clinical practice, healthcare resource allocation, and future research in risk prediction modeling for HF patients with frailty.

II.   RESEARCH CONTEXT

Traditional survival analysis has been used extensively to predict mortality in this patient population. Deep learning methods have also been employed - see [1], [2], [3], [4] and [5]. 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?

The 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 [6], 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 [7].

3.   Literature Review

Several past papers have addressed predictive modeling for heart failure patients. A deep neural network model with learned medical feature embedding is proposed in [8] to address high dimensionality and temporality in electronic health record (EHR) data. Here, a convolutional neural network is used 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.

Personalized predictive modeling is investigated in [9], which 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, the authors proposed a time-fusion CNN framework. A 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 [10] 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 [11]. 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 to enable continuous probabilistic risk prediction for hospitalization and mortality, addressing a critical need in patient care. The use of EHR such as those available in CPRD, allows access to 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 (HMMs) [12]. This project aims to bridge this gap in the literature.

III. AIM & OBJECTIVES

A. Research Questions:

The research questions I am looking to answer are as follows.

1. *To what extent can HDL predict HF as time-to-event?*

   In an attempt to answer this question, I will build a hybrid deep learning architecture which take as input the contextual/ historical features for a patient from EHR. The output sought from this model will be in the structure of an array denoting a persistency metric for the patient over time periods. This is akin to survival curves used in traditional survival analytics.

2. *How transparent/ interpretable can the HDL model be for prediction of HF?*

   SHAP and LIME can be used to add interpretability to the neural network to understand how the various covariates affect the prediction of HF vs non-HF. The extent to which this can unravel the model predictions can estimated with further research.

3. *What are the trade-offs between HMM and HDL with respect to interpretability and performance?*

   Generally, HMMs being natively linear models, can perhaps not capture embedded non-linearities in

real-world data (it is important to note that non-linear extensions have been suggested [13]). However, that is weighed against the fact that the model is highly transparent. Neural networks on the other hand may offer better performance, but their level of explainability suffers.

B. Objectives:

This project aims to develop a scalable predictive model for frail patients with heart failure, combining deep neural networks with statistical machine learning. 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 objective is to compare the hybrid deep learning models with a more mathematically transparent model, Hidden Markov Models. If time permits, I will also assess the trade-off between model performance and interpretability.

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 contaning EHR records of a vast number of patients, consisting of both HF patients and non-HF patients. 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.

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 usually 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 60,000 patients admitted to the Beth Israel Deaconess Medical Center ICU between 2001 and 2012. 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. 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.

In the absence of CPRD Aurum data (if it does not become available within a feasible time period), I will be using MIMIC-III data in a similar fashion to analyse covariates and their interactions to influence heart failure risk. Previous authors [14] have used this data to look at predicting risk amongst patients in intensive care units. ECG readings have been used in another study to detect atrial fibrillation (AF) in a subset of patients [15], which is a good predictor for heart failure risk [16].

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. In the event that simulation is required, R also has 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 [17] and Kaplan-Meier curves [18]. 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 the 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 adverse outcomes in frail 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) [19] and Linear Discriminant Analysis (LDA) [20] 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 even of interest - in this 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 [21].

2. Secondary Aim: HMM

The secondary aim would be to use either the multivariate time series of observed biosignals or their latent vector representation $o_{\{1,\cdots,T\}}$ as observed states in a Hidden Markov Model $\lambda = (A, B, \pi)$ [1]. Algorithms for likeli-

———

[1] A and B are respectively the state transition matrix and the observation (emission) matrix. π denotes initial probabilities.

hood 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 [22]. 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, casting 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.

VI. DATA GOVERNANCE & ETHICS

Although CPRD data is irreversibly anonymized and does not fall under the category of human subjects, ethics approval is still required 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. In case getting access to CPRD Aurum becomes difficult, MIMIC-III data will be used, as described earlier. 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 issues on the University website that is delayed me from getting access to the platform. I currently have gained access and am in the process of filling out my approval form before sending it for review.

VIII. PROJECT PLAN

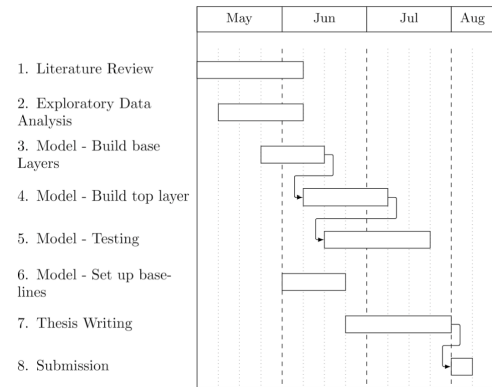The project plan is outlined by the following chart.

FIG. 1. Literature research will be front-loaded within the first month, during which exploratory data analysis will also be conducted. Building the base and upper layers for the model(s) along with simultaneous testing against baselines will continue through most of June and July. Most of July is expected to be devoted to writing the thesis, so that it is ready for submission by early August.

IX. CONCLUSION

As established earlier, timely detection of HF risk can be a useful tool in delivering required medical intervention to vulnerable patients, particularly those suffering from frailty. This project will leverage CPRD's rich data source in an attempt to develop a robust predictive model for hospitalization and mortality in frail HF patients. Traditional survival analytics can present a probabilstic view of time-to-event. This project would aim to supercede their performance, while maintaining explainability of the model(s). If feasible, the ultimate goal woudl be to assess the trade-off between performance and interpretability arising from the two different paradigms of HDL and HMM.

1.[1] Martin Gjoreski, Anton Gradisek, Borut Budna, Matjaz Gams, and Gregor Poglajen. Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. *IEEE Access*, 8:20313–20324, 2020.

2.[2] Jeffrey J Nirschl, Andrew Janowczyk, Eliot G Peyster, Renee Frank, Kenneth B Margulies, Michael D Feldman, and Anant Madabhushi. A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of h&e tissue. *PloS one*, 13(4):e0192726, 2018.

3.[3] Liqin Wang, Long Sha, Joshua R. Lakin, Julie Bynum, David W. Bates, Pengyu Hong, and Li Zhou. 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, 07 2019.

4.[4] Jose Roberto Ayala Solares, Francesca Elisa Diletta Raimondi, Yajie Zhu, Fatemeh Rahimian, Dexter Canoy, Jenny Tran, Ana Catarina Pinho Gomes, Amir H. Payberah, Mariagrazia Zottoli, Milad Nazarzadeh, Nathalie Conrad, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. Deep learning for electronic health records: A comparative review of multiple deep neural architectures. *Journal of Biomedical Informatics*, 101:103337, 2020.

5.[5] Giulia Lorenzoni, Stefano Santo Sabato, Corrado Lanera, Daniele Bottigliengo, Clara Minto, Honoria Ocagli, Paola De Paolis, Dario Gregori, Sabino Iliceto, Franco Pisanò, and et al. Comparison of machine learning techniques for prediction of hospitalization in heart failure patients. *Journal of Clinical Medicine*, 8(9):1298, 2019.

6.[6] Electronic Frailty Index (EFI). `https://www.england.nhs.uk/ourwork/clinical-policy/older-people/frailty/efi/`, 2023. Accessed on April 24, 2023.

7.[7] Stacey Peters, Renee Johnson, Samuel Birch, Dominica Zentner, Ray E Hershberger, and Diane Fatkin. Familial dilated cardiomyopathy. *Heart, Lung and Circulation*, 29(4):566–574, 2020.

8.[8] Zhengping Che, Yu Cheng, Zhaonan Sun, and Yan Liu. Exploiting convolutional neural network for risk prediction with medical feature embedding. *arXiv preprint arXiv:1701.07474*, 2017.

9.[9] Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Aidong Zhang, and Jing Gao. Personalized disease prediction using a cnn-based similarity learning method. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 811–816. IEEE, 2017.

10.[10] Landon Brand, Aditya Patel, Izzatbir Singh, and Clayton Brand. Real time mortality risk prediction: A convolutional neural network approach. In *HEALTHINF*, pages 463–470, 2018.

11.[11] William Caicedo-Torres and Jairo Gutierrez. Iseeu: Visually interpretable deep learning for mortality prediction inside the icu. *Journal of biomedical informatics*, 98:103269, 2019.

12.[12] LAWRENCE R. RABINER. A tutorial on hidden markov models and selected applications in speech recognition. *Readings in Speech Recognition*, page 267–296, 1990.

13.[13] Andrew D Wilson Aaron F Bobick. Nonlinear parametric hidden markov models. ND.

14.[14] Fuhai Li, Hui Xin, Jidong Zhang, Mingqiang Fu, Jingmin Zhou, and Zhexun Lian. Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the mimic-iii database. *BMJ Open*, 11(7), 2021.

15.[15] Syed Khairul Bashar, Eric Ding, Daniella Albuquerque, Michael Winter, Sophia Binici, Allan J. Walkey, David D. McManus, and Ki H. Chon. Atrial fibrillation detection in icu patients: A pilot study on mimic iii data. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 298–301, 2019.

16.[16] Bilato C Bordignon S, Chiara Corti M. Atrial fibrillation associated with heart failure, stroke and mortality. 2012.

17.[17] Dhananjay Kumar and Bengt Klefsjö. Proportional hazards model: a review. *Reliability Engineering & System Safety*, 44(2):177–188, 1994.

18.[18] J Martin Bland and Douglas G Altman. Survival probabilities (the kaplan-meier method). *Bmj*, 317(7172):1572–1580, 1998.

19.[19] Hervé Abdi and Lynne J. Williams. Principal component analysis. *WIREs Computational Statistics*, 2(4):433–459, 2010.

20.[20] Petros Xanthopoulos, Panos M Pardalos, Theodore B Trafalis, Petros Xanthopoulos, Panos M Pardalos, and Theodore B Trafalis. Linear discriminant analysis. *Robust data mining*, pages 27–33, 2013.

21.[21] DY Lin. Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in medicine*, 13(21):2233–2247, 1994.

22.[22] Daniel Jurafsky and James H. Martin. Speech and language processing, 2000.