Towards Optimizing Left Ventricular Assist Device (LVAD) Therapy for Patients with Advanced Heart Failure: Exploring Machine Learning Applications in Pre- and Post-LVAD Therapy

by

Faezeh Movahedi

M.S. in Health and Physical Activity,

University of Pittsburgh, 2014

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Faezeh Movahedi

It was defended on

July 28, 2023

and approved by

Natasa Miskov-Zivanov, PhD, Assistant Professor, Electrical and Computer Engineering

Amro A. El-Jaroudi, PhD, Associate Professor, Electrical and Computer Engineering

Zhi-Hong Mao, PhD, Full Professor, Electrical and Computer Engineering

Ahmed Dallal, Associate Professor, Electrical and Computer Engineering

Rema Padman, Trustees Professor, Management Science and Healthcare Informatics,

Carnegie Mellon University

James F Antaki, Professor, Meinig School of Biomedical Engineering, Cornell University

Copyright © by Faezeh Movahedi
2023

Towards Optimizing Left Ventricular Assist Device (LVAD) Therapy for Patients with Advanced Heart Failure: Exploring Machine Learning Applications in Pre- and Post-LVAD Therapy

Faezeh Movahedi, PhD

University of Pittsburgh, 2023

Heart failure is a significant global public health concern affecting millions of individuals. Implantable Left Ventricular Assist Devices (LVADs) have proven beneficial for patients with advanced heart failure who are unresponsive to conventional treatments. However, despite the improved survival rates associated with LVADs, this therapy carries a high risk of recurrent and severe adverse events (AEs) that lead to increased morbidity and mortality. Many clinical studies have examined the AE profiles of LVAD patients, but these studies often focus on statistical analysis and treat each AE as a separate event, overlooking potential relationships and interactions among AEs. This thesis aims to overcome these limitations by exploring three machine learning applications integrated with a National registry dataset during both pre- and post-implantation time frames. The first module involves post-LVAD sequential AE pattern mining and clustering, providing a comprehensive view of the AE landscape in the LVAD population. By identifying critical time points and subgroups with distinct AE patterns that significantly impact therapy outcomes, this module may serve to inform personalized care strategies, resource allocation, and follow-up scheduling. The second module developed a pre-implant predictive risk model that assesses AE risk based on patients' preoperative clinical profiles. This model may serve to enhance patient selection and management of high-risk patients prior to the procedure. The third module created a post-LVAD risk model predicting future AEs based on pre- and post-LVAD clinical profiles supplemented with the patients' AE history in first critical weeks. This risk may help clinicians to anticipate subsequent AEs and employ preventive measures. These three modules collectively represent a crucial step towards optimizing LVAD therapy and supporting human decision-making by integrating comprehensive clinical information, treatment history, and temporal events. Future translation of these modules into clinical application software

will contribute to a clinical decision support system through the integration of various data mining tasks linked with temporal clinical milestones.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

1.1 Motivation

1.1.1 Advanced Heart Failure

Heart failure (HF) is characterized as an abnormality in the structure or function of the heart, with typical symptoms including breathlessness, persistent coughing, swelling in ankles, and fatigue [1]. HF is a chronic condition and may gradually progress to advanced HF. There is no globally uniform definition for advanced HF. Some clinicians believe advanced HF should be defined based on high-risk factors of mortality, while others insist on factors like poor quality of life and severe symptoms [2]. Multiple classification systems exist to characterize patients with HF. Based on the 4-level classification by the New York Heart Association, patients with HF who are unable to carry out any physical activity without discomfort and show symptoms of HF at rest are categorized as stage 4 (advanced HF) [3]. The American Heart Association also introduced an A-to-D staging system, in which stage D (advanced HF) is defined as having refractory heart failure despite optimal medical therapy [4]. Recurrent HF-related hospitalizations, intolerance of neurohormonal antagonists, increased diuretic use, organ failure, malnutrition, and refractory arrhythmias are characteristics of patients with stage D of HF [5].

HF is a major public health issue that affects approximately 5.7 million people in the United States and about 26 million people worldwide [6–8]. There are about 670,000 newly diagnosed cases of HF per year in the United States [7,8]. It is estimated that the prevalence of HF will increase by 46% from 2012 to 2030, resulting in over 8 million people aged 18 years or older with HF [6]. HF is one of the leading causes of death and is associated with significant morbidity and poor quality of life [6, 9, 10]. HF is the primary cause of over one million hospitalizations and the secondary cause of over three million hospitalizations in the United States [2]. The annual cost of HF in the United States is $39 billion (estimated in 2010), which accounts for approximately 4%of total healthcare costs [2, 11, 12]. A disproportionate

portion of the annual costs of HF is incurred by patients with advanced HF [11]. There are no accurate statistics about the prevalence of advanced HF due to disagreements among organizations regarding its definition [2, 4]. Nevertheless, it is estimated that advanced HF currently affects approximately 250,000 to 500,000 people in the United States [2, 4]. Advanced HF is a chronic disease, and it is possible that all patients with HF might eventually progress to advanced HF. Therefore, it is impossible to define the incidence of advanced HF.

1.1.2 The Rise of Left Ventricular Assist Devices (LVADs)

Advanced HF poses significant challenges in terms of treatment and management. Traditional therapies, such as medical treatment and symptom management, often fall short in providing satisfactory outcomes for patients with advanced HF [13]. In such cases, alternative therapies become crucial in improving patient outcomes and quality of life. While heart transplant is considered the gold standard treatment for eligible patients, the limited availability of donor organs means that only a minority of patients can undergo heart transplants [13]. In the United States, the number of donors per year is approximately 2,000-2,500, which is about 10 times less than the number of patients in need of heart transplants [14,15]. This shortage of donor organs has necessitated the development of innovative therapies, with left ventricular assist devices (LVADs) emerging as a groundbreaking solution for advanced HF.

LVADs are surgically implantable mechanical circulatory support devices designed to augment the pumping function of the left ventricle of the heart. The latest technology of LVAD, HeartMate 3, is depicted in Fig. 1. These devices serve as blood pumps, facilitating the flow of blood from the weakened left ventricle to the aorta, ensuring adequate blood supply to the rest of the body. The history of LVADs over the past two decades is extensive, marked by numerous significant errors, but also remarkable achievements. LVAD technology has witnessed remarkable advancements, leading to the development of more compact, less invasive, and increasingly efficient devices [16,17]. See 2.1.1 to learn about the generation of LVADs. Over the years, LVADs have evolved from pulsatile-flow pumps to more advanced continuous-flow pumps. These newer generations of LVADs, such as the HeartMate II,

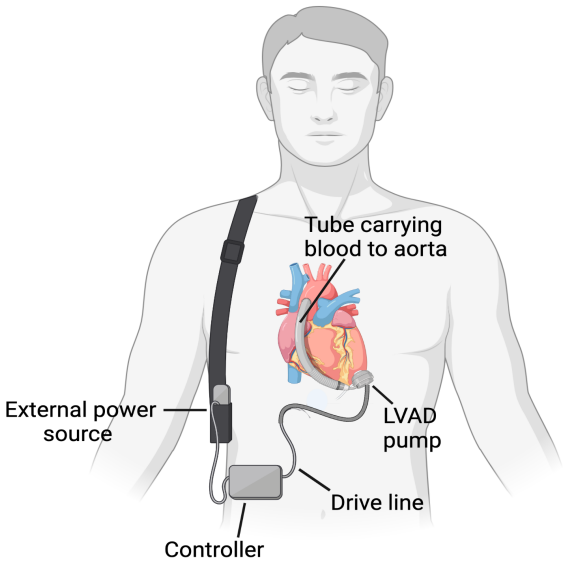


Figure 1: llustration of HeartMate 3 LVAD.

HeartWare Ventricular Assist Device (HVAD), and HeartMate 3, offer improved durability, reduced risks of complications, and enhanced patient outcomes. The miniaturization of devices and the implementation of magnetically-powered and levitating technologies have further contributed to the advancements in LVAD therapy.

The rise of LVADs can be attributed to their significant clinical benefits. Patients with advanced HF who do not respond well to traditional therapies or are ineligible for heart transplants can find a lifeline in LVAD therapy. These devices offer a bridge to transplant (BTT) for patients awaiting heart transplants, providing temporary support until a suitable donor organ becomes available. Additionally, LVADs offer destination therapy (DT) for patients who are not eligible for heart transplants but require long-term mechanical support as a permanent therapy. When analyzing overall survival based on device indication, it becomes evident that patients in BTT group demonstrate superior 1- and 5-year survival rates compared to those in the DT group [18].

In recent years, there has been a notable surge in the number of LVAD implants, establishing these devices as a standard treatment for advanced HF. Based on the Society of

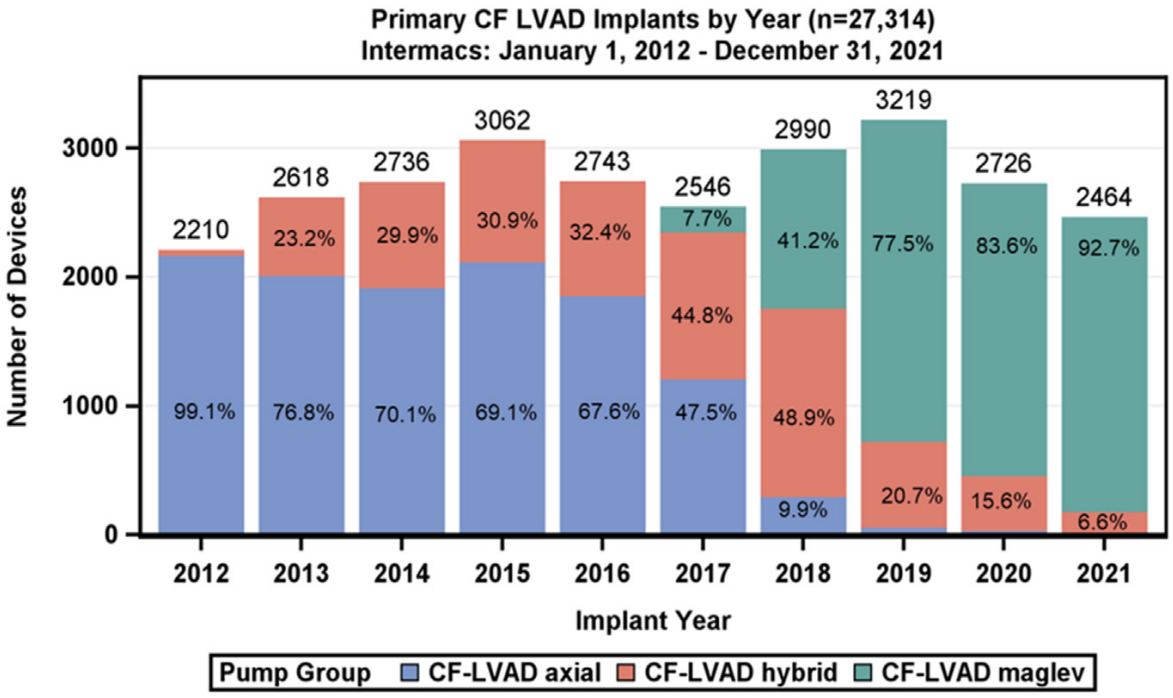


Figure 2: The annual number of CF-LVAD implants categorized by flow configuration. The devices include axial-flow pumps (HeartMate II; Abbott Laboratories), centrifugal hybrid levitation (HeartWare Ventricular Assist Device [HVAD]; Medtronic, Inc), and centrifugal full magnetic levitation (HeartMate 3 [HM3]; Abbott Laboratories) [18].

Thoracic Surgeons (STS) 2022 annual report, from January 2012 to December 2021, 107 received a pulsatile-flow durable LVAD and the majority, 28,597, underwent implantation of a durable continuous-flow LVAD (CF-LVAD) during the initial LVAD operation [18]. The annual number of CF-LVAD implants categorized by flow configuration is presented in Fig. 2 [18]. Since the FDA's approval of the HeartMate 3 (HM3) for long-term support indications such as BTT and DT in 2018 and 2019, there has been a significant surge in the volume of HeartMate 3 (Multicenter Study of MagLev) implants, accounting for a remarkable 92.7% of all CF-LVADs implants in 2021.

The continuous refinement of LVAD technology, improvements in surgical techniques, and the collaborative efforts of healthcare professionals have contributed to the success of these therapies. As research and innovation in the field continue, it is expected that the numbers of LVAD implants will continue to rise, offering hope and improved outcomes for an increasing number of patients with advanced heart failure [19].

1.1.3 Improved Outcomes of in LVAD Therapy

Recent advancements in device technology, surgical implant techniques, and clinical management have led to improved patient outcomes and survival in the past decade. Overall survival rates for patients undergoing LVAD implant between 2017 and 2021 showed significant improvement, with 1- and 5-year survival rates of 83.0% vs 81.2% and 51.9% vs 43.0%, respectively ($P < .0001$). This achievement is particularly noteworthy as it marks the first time that the 5-year survival rate for patients with LVAD exceeds 50% [18, 20].

However, the success of LVAD therapy is not just about survival; LVADs have also shown remarkable benefits in enhancing the quality of life for patients with advanced heart failure [21–26]. Approximately 80% of patients with LVAD were satisfied with their decision to have VAD therapy during the first 2 years [20]. By improving cardiac output and ensuring adequate blood flow, LVADs alleviate symptoms such as fatigue, shortness of breath, and exercise intolerance that significantly impact patients' daily lives [21, 22, 24–26]. With the device providing mechanical support to the heart, patients regain their ability to perform routine activities, engage in physical exercise, and enjoy a more active lifestyle [21,22,24,25].

Additionally, LVADs offer the potential for a more comprehensive and coordinated approach to patient care [27–29]. LVAD recipients are closely monitored by a multidisciplinary team, including cardiologists, cardiac surgeons, nurses, and other healthcare professionals. This collaborative care model ensures that patients receive comprehensive medical management, personalized support, and education to manage their device and maintain their overall well-being.

In summary, LVAD therapy goes beyond simply prolonging survival. It provides a substantial improvement in the quality of life for patients with advanced HF. By restoring functional capacity, reducing symptoms, and offering a BTT, LVADs have transformed the lives of many patients and opened new avenues for managing this challenging condition. With ongoing advancements in LVAD technology and patient care, the future holds even greater promise for enhancing patient outcomes and further improving their quality of life.

1.1.4 Challenges in LVAD Therapy: Adverse Events (AEs)

Despite the improved outcomes observed in patients with LVADs, these devices are not without their challenges. One of the primary concerns is the occurrence of a wide range of adverse events (AEs). Major post-LVAD AEs include bleeding, infection, cardiac arrhythmia, stroke, device malfunction, right heart failure, renal dysfunction, and respiratory failure [20]. These AEs can have a significant impact on the overall outcome of LVAD therapy and, in some cases, may even lead to patient mortality [18, 20, 30–33]. While bleeding and infection are among the most common adverse events associated with LVAD therapy [18], it's worth noting that certain adverse events such as stroke and respiratory failure, although less frequent, can have a significant impact on patient mortality [18, 20, 30–33]. The subsequent survival of LVAD patients is influenced by the occurrence of major AEs within the first three months following implantation [20]. These AEs, such as neurological dysfunction, multiple system organ failure (MSOF), right heart failure (RHF), infection, and device malfunction, contribute significantly to mortality in post-LVAD patients, as depicted in Fig. 3 [20]. MSOF and stroke are identified as the primary causes of death during the early months after LVAD implantation, while stroke remains the leading cause of death between 6 months and 6

years [20].

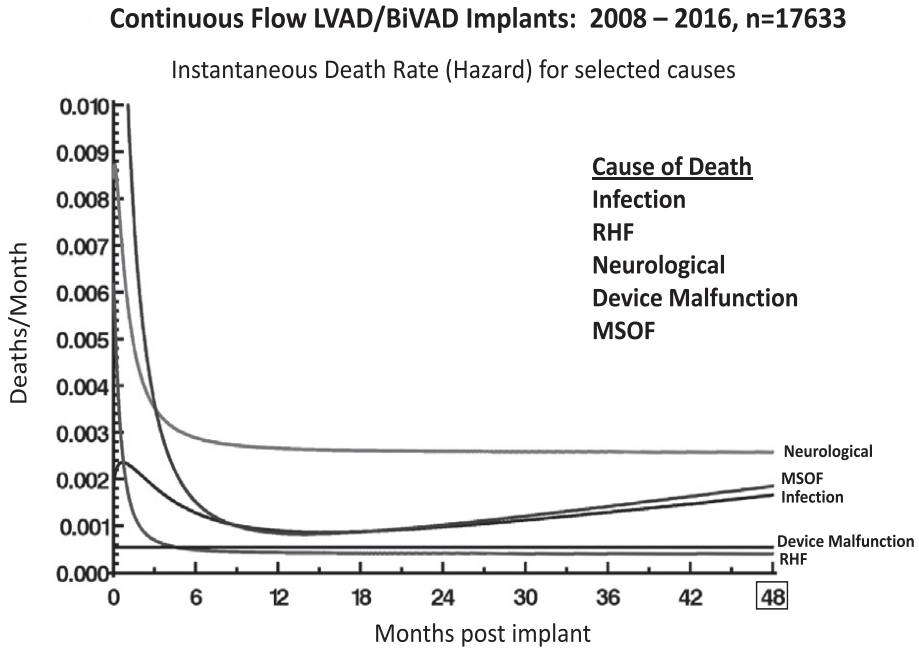


Figure 3: The hazard function curves illustrate the risk of death over time for the primary causes of mortality [20]. RHF (right heart failure) and MSOF (multiple system organ failure).

The occurrence of AEs can also lead to re-hospitalization, further impacting the patient's well-being and incurring additional healthcare costs [34, 35]. Re-hospitalization is often required for the management and treatment of complications related to Gastrointestinal bleeding and LVAD-related infections [34, 35]. These readmissions can disrupt the patient's recovery process, delay their return to normal activities, and add financial burden to the healthcare system.

The timing of adverse events (AEs) in LVAD patients is a crucial consideration, with studies indicating that the majority of these events tend to occur within the early period following LVAD implantation, typically within the first 90 days post-implantation. This early phase is associated with a higher risk of complications [18, 36]. During this initial period, patients are still adjusting to the presence of the LVAD, while the surgical site is healing. Table 1 provides AE rates for different types of AEs in the early (< 90 days) and late (> 90 days) periods, based on data from 13,945 patients who received an LVAD between 2017 and

2021 [18]. The most common AEs during the early and late periods after LVAD implantation are infections and bleeding. Certain AEs, such as renal dysfunction and respiratory failure, are more prevalent in the early period and rare in the late period.

In summary, AEs are a common occurrence in patients undergoing LVAD therapy and can significantly impact treatment outcomes. These events, such as bleeding, infection, stroke, and device malfunction, pose challenges that can lead to increased morbidity, mortality, and decreased quality of life. Investigating AEs in LVAD patients is of paramount importance to understand their underlying mechanisms, risk factors, early detection, and potential interventions. By conducting thorough research, healthcare professionals can gain valuable insights into the prevention, early detection, and management of these events, ultimately minimizing their negative effects on patient outcomes. This knowledge can inform improvements in LVAD technology, surgical techniques, patient selection, and post-implant care, enhancing the safety and effectiveness of LVAD therapy for individuals with advanced heart failure.

1.1.5 Adverse Events and the Cost of LVAD

The costs of LVAD therapy can be broken into LVAD implantation and post-LVAD care. The average cost of the LVAD implantation is estimated at \$141,287 $\pm$ 18,513 [12]. LVAD implantation costs includes the LVAD device, professional payment, hospitalization, operating room, diagnostics, laboratory tests, blood products, drugs, miscellaneous, and rehabilitation. The largest costs are related to the LVAD device itself (48%:), professional payment (17%), and stay in ICU (10%) [12]. The post-LVAD costs include cost of follow-up visits such as professional payment, readmission costs (rate of 2.64 per person), device replacement (rate of 0.06 per person), and outpatient costs including professional services, laboratory tests, and drugs [11]. The post-LVAD costs, especially readmission costs, are significantly affected by post-LVAD AEs [37,38]. Infection, bleeding, and respiratory failure are the major factors that decisively affect LVAD costs [37–39]. In addition, AEs affect LVAD cost efficiency by increasing the mortality and reducing the quality of life of post-LVAD patients, therefore decreasing the cost effectiveness ratio for LVAD therapy [11,37]. Improvements in LVAD de-

Table 1: The Society of Thoracic Surgeons (STS) 2022 annual report of Adverse Events (AEs) in 13,945 Continuous-Flow LVAD Patients (January 1, 2017 - December 31, 2021) [18] provides information on AE rates, which indicate the number of events per patient-year in the early (< 90 days) and late (> 90 days) periods.

AE	Period	AE Count (No.)	AE Rate	Patients (No.)	Patients (%)
Major bleeding	Early	3928	1.22	788	20.0
	Late	5007	0.25	2639	18.9
Cardiac arrhythmia	Early	2873	0.89	2266	16.2
	Late	2093	0.11	1384	9.9
Device malfunction	Early	463	0.14	243	3.0
	Late	1644	0.08	12539	39.0
Major infection	Early	4181	1.30	123	22.4
	Late	8420	0.43	386	31.5
Hepatic dysfunction	Early	273	0.08	267	1.9
	Late	248	0.01	2291	1.6
Myocardial infarction	Early	26	0.01	260	0.2
	Late	36	0.00	235	0.3
Neurologic dysfunction	Early	1556	0.48	1393	10.0
	Late	2161	0.11	1621	11.6
Renal dysfunction	Early	1506	0.47	1404	10.1
	Late	806	0.041	657	4.7
Respiratory failure	Early	2213	0.69	1830	13.1
	Late	773	0.039	646	4.6
Venous thromboembolism	Early	217	0.067	206	1.5
	Late	71	0.004	67	0.5
Other serious AEs	Early	3180	0.99	2106	15.1
	Late	2806	0.14	1717	12.3

vice technology, operation techniques, and post-LVAD management have reduced the LVAD cost over time and improved the quality of life and LVAD final outcome for patients. Yet, the cost effectiveness of LVAD therapy for advanced HF has not yet accomplished the accepted range as an effective new technology. Further reduction in long-term AEs and readmission, and improvement in post-implant care strategies and LVAD devices would help this therapy to be cost-effective.

1.2 Direction and Goals

Considering the rapidly increasing rate of LVAD implantation in recent years, accompanied by the high rate of post-LVAD AEs and their impact on both mortality and quality of life [18], it is evident that early recognition and diagnosis of AEs are of utmost importance. LVAD patients experience a wide range of AEs, each requiring a unique clinical approach to diagnosis and care management [40]. Additionally, detecting post-LVAD AEs is inherently complex, as it involves continuous monitoring of various clinical and device parameters, and requires collaboration among clinical professionals, nurses, perfusionists, and/or bioengineers involved in patient and device management. Consequently, this can significantly contribute to the financial burden associated with an already costly therapy [12]. On the other hand, deficiencies in post-LVAD follow-up visits may lead to delayed AE detection, resulting in additional complications and potentially death.

Considering the limitations in early AE diagnosis, the development of a supplementary prognostic AE tool would be beneficial in alerting healthcare providers and patients about impending risks, thereby enabling them to take preventive actions. Moreover, analyzing LVAD AE trajectories across a broad spectrum of LVAD patients would provide valuable insights to clinicians, educating and informing them about potential treatment approaches.

1.3 Dissertation Scope

To develop a prognostic model for early diagnosis of AE after LVAD, we must first fully characterize LVAD patients based on their post-LVAD AE trajectories. While the risks of developing post-LVAD AEs are clinically well-recognized, the early diagnosis and management of AEs are not well-characterized due to the complexity of patient population and variable symptoms and signs [40, 41]. Many studies have investigated the AE profile of patients after LVAD by focusing on incidence of (or freedom from) various types of AE at different points of time after LVAD, as well as the associated pre and post- LVAD risk factors [30,31,42–50]. However, to date there has been no study to investigate temporal relationships of multiple types of AEs in patients with LVAD. A first step to address this deficit is to identify and model patterns of transitions between AEs and their related outcomes. A descriptive temporal learning approach, like sequential pattern mining and clustering, can extract knowledge about the heterogeneity of temporal AEs patterns after LVAD and can provide comprehensive and clinically meaningful profiles within post-LVAD patients.

Another crucial area that requires further research is the utilization of an AE prognostic tool to screen LVAD candidates for receiving LVAD therapy. Developing a reliable risk stratification model based on patients' pre-LVAD profiles would greatly improve decision-making for LVAD therapy. Previous studies have developed risk stratification models based on patients' pre-LVAD characteristics to predict the final outcome of LVAD (death versus survival) and major AEs such as bleeding and right heart failure [51–54, 54–56]. However, the accuracy of these scores at the patient level is limited due to poor-to-modest discrimination power [56, 57]. To address this limitation, the first step in this analysis is to investigate the improvement in the performance of risk models by addressing issues like systematic overfitting. The next step in this analysis is to determine if patients' pre-implant clinical profile could be associated with patients' post-LVAD trajectories. This analysis could transfer the pre-LVAD risk stratification for single AE or outcome into levels of risk for various post-LVAD trajectories (patterns).

Risk prediction models for LVAD patients developed thus far have relied on data collected at the time of implantation [53, 54, 57–60]. While these models offer valuable insights into

overall mortality risk for patient selection before implantation, their drawback lies in their inability to capture changes in patient health status after LVAD implantation. Thus, there is an opportunity to develop post-LVAD risk models by incorporating follow-up data and patient AE history to enable real-time prediction of future risk. Considering the potential significance of various pre- and post-LVAD factors, ranging from demographic information and lab values to medical therapies and AE history, as predictors of mortality, it is crucial to investigate these relationships further, particularly by exploring their combinations [61–70].

1.4 Specific Aims

Three aims have been designed for this research, focusing on characterizing the patterns of transitions between adverse events (AEs) in LVAD patients and improving risk stratification models for LVAD therapy, with the goal of enhancing the understanding of AE journeys, improving decision-making for LVAD therapy, and enabling accurate prediction of patient outcomes.

Aim 1: Characterize and model patterns of transitions between AEs in patients with LVAD.

- Investigate the temporal aspect and similarity of AE journeys among LVAD patients by utilizing frequent sequential pattern mining on the multi-center National database of longitudinal data for post-LVAD AEs.
- Cluster patients into groups with similar AE sequences.
- Model sequential AE pattern of each group resulting from clustering.
- Characterize patients in each group, exploring various factors, to develop comprehensive patient profiles.

Aim 2: Develop a reliable risk stratification model for LVAD therapy based on patients' pre-LVAD profiles.

- Develop a risk stratification model for post-LVAD mortality.
- Develop a risk stratification model for post-LVAD right heart failure.

- Address the imbalance issue in developing predictive models in the LVAD field and raise awareness about the limitations of ROC as a common metric for evaluating model performance.

Aim 3: Develop post-LVAD risk prediction models by incorporating follow-up data and patient AE history.

- Develop a classifier that predicts risk of early mortality after implant based on the patient's AE history, as well as pre- and post-LVAD clinical profiles.

The outcome of Aim 1 can inform personalized care strategies by providing guidance on critical time points following LVAD implantation and identifying subgroups of patients whose specific patterns of AEs significantly impact the outcome of LVAD therapy. These subgroups require attention and assistance in allocating resources and scheduling follow-up appointments. The outcome of Aim 2 enhances the accuracy of patient selection for LVAD implantation and enables practitioners to stabilize high-risk patients before the procedure, ensuring a more favorable outcome. Finally, the outcome of Aim 3 facilitates the anticipation of AEs in the early weeks following LVAD implantation. This allows for the timely implementation of preventive measures or adjustments to treatment plans, promoting better patient outcomes. It is hoped that the combination of these three aims will dramatically optimize the outcomes of LVAD implantation for patients with advanced heart failure.

1.5 Dissertation Organization

Chapter 2 explores the background of the evolutionary progress of LVAD technology, the post-LVAD care management, LVAD-related AE definitions, and advancements and challenges of applications of machine learning in pre- and post-LVAD therapy. Chapter 3 describes the main topics addressed for the final dissertation defense and the main approaches. Chapter 4 provides insights into the analysis of AEs, exploring their timelines and patterns of similarity using sequential pattern mining. Chapter 5 explores approaches to enhance the discrimination power and generalizability of risk models for mortality and right heart failure

after LVAD implantation. Chapter 6 provides a holistic view of the AE landscape in the LVAD population through clustering analysis, identifying subgroups with specific patterns of AEs. Chapter 7 integrates pre- and post-implantation data with the history of AEs to develop models that predict mortality during critical early weeks following LVAD implantation. Chapter 8 summarizes the conclusions of the research presented in this dissertation and the possible future directions to be investigated as the next steps for this research.

2.0 Background

2.1 Device

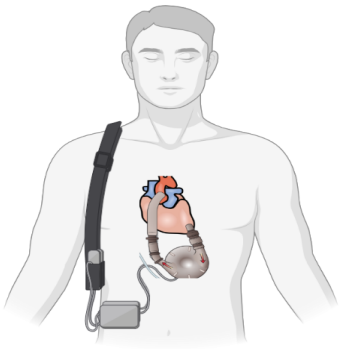
2.1.1 LVAD Generations: Evolutionary Progress

LVADs have evolved over the years, with different generations representing advancements in technology and design. The first generation of LVADs introduced pulsatile-flow pumps, such as the HeartMate I (Thoratec, now Abbott Laboratories) as depicted in Fig. 4a, which aimed to mimic the pulsatile function of the heart. These devices, while effective, had limitations such as large external pumps, limited mobility for patients, and increased risk of bleeding and infection [71, 72].

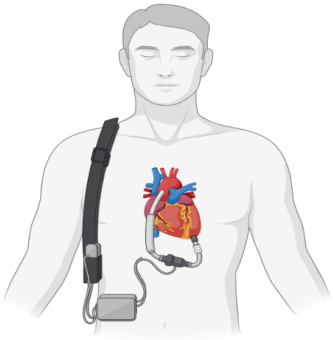
The second generation of LVADs brought about a shift to continuous-flow pumps, such as the HeartMate II (Abbott Laboratories), as illustrated in Fig. 4b, and the HeartWare Ventricular Assist Device (HVAD) (Medtronic, Inc). These devices featured a smaller size, simpler implantation technique, and fewer moving parts. These improvements led to longer durability and reduced risks of thromboembolism, infection, and malfunction [71, 72, 74]. Continuous-flow LVADs became the new standard, offering enhanced efficiency and longer support.

The third generation of LVADs further refined the technology by incorporating magnetically-powered continuous-flow pumps. Notable brands in this generation include the HeartMate 3 (Abbott Laboratories) as depicted in Fig. 4c, which utilizes axial or centrifugal motors and employs noncontact bearings to minimize prothrombotic sites. The HeartMate 3 has shown improvements in efficiency, durability, and patient outcomes [71, 72].

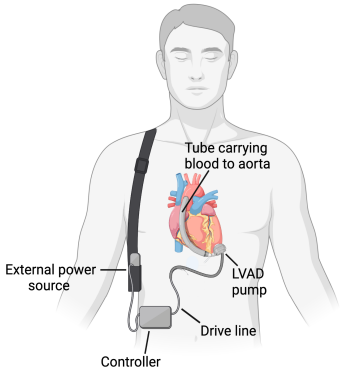
Overall, the successive generations of LVADs, including brands like HeartMate I, HeartMate II, HeartWare HVAD, and HeartMate 3 have demonstrated significant progress in terms of size, invasiveness, hemocompatibility, and clinical outcomes. These advancements have expanded the options for patients with advanced heart failure who require mechanical circulatory support, providing them with improved quality of life and extended survival.



(a) First generation of LVADs: pulsatile-flow pumps. Sketch of PVAD from Thoratec Corporation [73].



(b) Second generation of LVADs: axial continuous-flow pumps. Sketch of HeartMate II LVAD from Thoratec Corporation [73].



(c) Third generation of LVADs: centrifugal continuous-flow pumps. Sketch of HeartMate3 LVAD from Thoratec Corporation [73].

Figure 4: The three generations of LVAD

2.2 Clinical

2.2.1 Post-LVAD care management

After LVAD implantation, recovery and follow-up care play crucial roles in optimizing patient outcomes. The post-LVAD period requires a comprehensive and multidisciplinary approach involving specialized healthcare professionals such as cardiologists, cardiac surgeons, nurses, and rehabilitation specialists. The complexity of post-LVAD management and the need for coordinated care have been topics of extensive research, leading to significant improvements over time [27–29, 75]. These advancements aim to enhance patient recovery, reduce complications, and improve overall quality of life for individuals with LVADs.

The recovery period following LVAD surgery typically involves a hospital stay for a few weeks, during which the medical team closely monitors the patient's progress and adjusts medications and therapies as needed. During this initial recovery phase, patients receive specialized care in the intensive care unit (ICU) or a dedicated cardiac care unit. This allows for close monitoring of vital signs, device function, and potential complications. The medical team focuses on managing pain, optimizing fluid balance, and promoting wound healing.

Once the patient's condition stabilizes, they are transitioned to a step-down unit or a specialized cardiac rehabilitation facility. In these settings, patients undergo a structured rehabilitation program that includes physical therapy, exercise training, and education on LVAD management, lifestyle modifications, and self-care. The goal is to gradually improve strength, mobility, and overall functional capacity.

The follow-up care after LVAD implantation is crucial for ongoing monitoring and adjustment of the device settings. Patients typically have frequent outpatient visits with their LVAD care team, which includes cardiologists, LVAD coordinators, nurses, and other healthcare professionals. During these visits, the team evaluates device function, assesses the patient's overall health, and addresses any concerns or complications that may arise. Regular laboratory tests, imaging studies, and device checks are performed to ensure optimal device performance and detect any potential issues early on. Medication management, including

anticoagulation therapy and immunosuppressive medications (in the case of transplant candidates), is closely monitored and adjusted as needed.

Furthermore, patient education plays a crucial role in post-implantation care. Patients and their caregivers receive extensive training on LVAD management, including device care, driveline maintenance, infection prevention, and emergency procedures. This education empowers patients to actively participate in their care and recognize signs of potential complications.

The follow-up care after LVAD implantation extends beyond the medical aspects. It includes psychosocial support to address emotional and psychological well-being, as the transition to life with an LVAD can be challenging for patients and their families. Support groups, counseling services, and resources for coping with the psychosocial impact of living with an LVAD are often provided.

Overall, recovery and follow-up care after LVAD implantation involve a multidisciplinary approach, focusing on physical rehabilitation, device management, monitoring of health parameters, medication optimization, patient education, and psychosocial support. This comprehensive care aims to maximize patient outcomes, enhance quality of life, and ensure long-term success with the LVAD therapy.

2.2.2 Adverse Events Definitions

Historically, the definition of AEs for LVADs in device clinical trials was determined through agreements between the FDA and device manufacturers. However, there were inconsistencies in criteria for patient selection, study endpoints, and AE definitions among different trials and devices, making it challenging to compare and interpret the data. To address this issue, initiatives were taken to standardize AE definitions for evaluating outcomes in the field.

One significant step towards standardization was the development of registries, such as the mechanical circulatory support (MCS) database of the International Society for Heart and Lung Transplantation (ISHLT), which was first published in 2003 [76]. The experience gained from the ISHLT registry led to the establishment of the Interagency Registry for

Mechanically Assisted Circulatory (INTERMACS) in 2005, supported by the National Heart, Lung, and Blood Institute [77]. INTERMACS, now maintained by The Society of Thoracic Surgeons (STS), became a pivotal resource for longitudinal studies on MCS in the United States and introduced standardized AE definitions [78].

However, recent clinical trials sponsored by device manufacturers have deviated from the established INTERMACS AE definitions [79]. This deviation highlighted the need for a comprehensive update in AE definitions to achieve international standardization and harmonization. To address this issue, collaboration between clinicians, the FDA, device manufacturers, and the Academic Research Consortium (ARC) [80] was initiated. The MCS-ARC took place between November 2017 and December 2018 and was specifically formed to develop and update AE definitions for various specific areas, involving experts from relevant disciplines, including surgeons, HF cardiologists, infectious disease specialists, nurses, engineers, and others [79]. The goal was to update AE definitions to reflect the evolving landscape of LVAD therapy and to improve the accuracy, ease of reporting, and data entry burden associated with AE reporting [79].

In summary, the definition of AEs for LVADs has evolved over time to address to meet the changing landscape of LVAD therapy and improve patient outcomes. In this section, the definition of *bleeding* and *infection* are presented to provide representative examples of AE definitions in INTERMACS. For a comprehensive list of definitions for all types of AEs, please refer to INTERMACS Appendices.

2.2.2.1 Bleeding Adverse Event

The association of the bleeding event should be classified as follows (1) Patient-related: (e.g., coagulopathy unrelated to surgical technique such as non-adherence with anticoagulation medication resulting in an inappropriately high level of anti-coagulation, hepatic failure).(2) Management-related: (e.g., related to surgical technique; hypertension; bleeding in the setting of inappropriate levels of anticoagulation) or to mismanagement of anticoagulants.(3) Pump related: (e.g., bleeding from the outflow graft, apical connector, or other internal components). Below is the definition of types of bleeding AE.

- Type 1

  Bleeding that is not actionable and does not cause the patient to seek unscheduled performance of studies, hospitalization, or treatment by a healthcare professional; may include episodes leading to self-discontinuation of medical therapy by the patient without consulting a healthcare professional. This type is not relevant during a hospitalization.

- Type 2

  Any overt, actionable sign of hemorrhage (e.g., more bleeding than would be expected for a clinical circumstance, including bleeding found by imaging alone) that does not fit the criteria for Type 3, 4, or 5 but does meet at least one of the following criteria:

  - Requiring non-surgical, medical intervention by a healthcare professional
  - Leading to hospitalization or increased level of care
  - Prompting evaluation

- Type 3
  - Type 3a

    Overt bleeding accompanied by hemoglobin drop of 3 to $< 5 g/dl$

    OR

    Any transfusion with overt bleeding
  - Type 3b

    Overt bleeding plus hemoglobin drop $5 g/dl$

    OR

    Cardiac tamponade

    OR

    Bleeding requiring surgical intervention for control (excluding dental/nasal/skin)

    OR

    Bleeding requiring intravenous vasoactive agents

- Type 4

  VAD implantation-related bleeding (includes concomitant cardiac or non-cardiac surgical procedures)

  - Reoperation after the closure of incision or incisions used to implant the VAD to control bleeding

- $\geq$ 50 kg: $\geq 4U$ packed red blood cells (PRBC) within any 48 hours during the first 7 days post-implant.
- $<$ 50 kg: $\leq$ 20 cm3/kg PRBC within any 24 hours during the first 7 days post-implant.
- Chest tube output $>$ 2 liters within 24 hours.

- Type 5

  Fatal bleeding

  - Type 5a

    Probable fatal bleeding; no autopsy or imaging confirmation but clinically suspicious
  - Type 5b

    Definite fatal bleeding; overt bleeding or autopsy or imaging confirmation

2.2.2.2 Infection Adverse Event

The association of the infection event should be classified as (1) Patient-related: (e.g., non-adherence or poor management of driveline exit site or indwelling catheters, IV drug abuse, aspiration).(2) Management-related: (e.g., improper tunneling, contamination of the intraoperative site, prolonged intubation).(3) Device-related: (e.g., Device endocarditis diagnosed by radiological examination or detection of pannus within the conduits or device).

MCS Related infections

- Percutaneous lead site infections
  - Superficial percutaneous lead infection

    A positive culture from the skin surrounding the percutaneous lead when there is clinical evidence of infection such as pain, fever, drainage, erythema, or leukocytosis coupled with the need to treat with anti-microbial therapy. The percutaneous lead exit site may have drainage and/or the surrounding skin may have erythema. The epithelialization of the percutaneous lead exit site is pre- served. The gram stain of the skin specimen at the driveline exit site will contain white blood cells (i.e., positive sign for inflammation).

- Deep percutaneous lead infection

  A positive culture from the driveline exit site deep to the epithelium, when there is clinical evidence of infection such as pain, fever, drainage, erythema, or leukocytosis coupled with the need to treat with anti-microbial therapy. The epithelialization of the percutaneous lead exit site is disrupted and no longer preserved or intact, or there is radiographic evidence of findings consistent with infection along the path of the percutaneous lead outside the mediastinum.

- Infection of external surfaces of an implantable component

  A positive culture from the tissue surrounding the external housing of a pump or one of its components implanted within the body (including device components such as controllers, batteries, etc.), when there is clinical evidence of infection such as pain, fever, drainage, erythema, or leukocytosis coupled with the need to treat with anti-microbial therapy.

- Infection of blood-contacting surfaces of an implantable component (device endocarditis)

  Infection of blood-contacting internal surfaces of the MCS device including inflow/outflow grafts: documented by positive blood cultures or radiographic or echocardiographic evidence of vegetation in blood flow path of the pump coupled with the need to treat with anti-microbial therapy.

Non MCS related infections

- Infective Endocarditis

  Non MCS related Positive blood cultures and echocardiography findings for mass or vegetation only on native valves, ICD, or pacemaker leads.

- Bloodstream Infection.

  Positive blood cultures with no other source identified

  Bloodstream infection: non VAD site or central venous catheter-related

  (definition from the Centers for Disease Control/National Healthcare Safety Network). Should be coupled with the need to treat with anti-microbial therapy.

- Mediastinitis

  - Procedure related mediastinitis

  - Non MCS related mediastinitis
  - Superficial mediastinal or thoracotomy wound infection

- Sepsis
  - Life threatening organ dysfunction caused by a dysregulated host response to infection with (1) Evidence of systemic involvement by infection, manifested by need to treat with anti microbial therapy.(2) Positive blood cultures.

- Localized non-MCS device infection
  Infection localized to a site not involving the MCS device or components (e.g., pneumonia, urinary tract infection, cholecystitis, diverticulitis, dental abscess) coupled with the need to treat with anti-microbial therapy.

2.3 National Registry Data: INTERMACS

INTERMACS, the Interagency Registry for Mechanically Assisted Circulatory Support, is a significant initiative in the field of mechanical circulatory support for heart failure patients. Established in 2005 with the support of the National Heart, Lung, and Blood Institute (NHLBI), INTERMACS collects comprehensive data on patients receiving mechanical circulatory support devices like LVADs to enhance patient care and outcomes [77]. Initially maintained by the Interagency Registry for Mechanically Assisted Circulatory Support and later transitioned to management by The Society of Thoracic Surgeons (STS) [78], this registry ensures continuous data collection and analysis to drive advancements in the field.

The latest report from STS INTERMACS presents outcomes for a large cohort of patients who underwent continuous-flow (CF) durable LVAD implantation from 2012 to 2021 [18]. Including data from 27,314 individuals, this report offers valuable insights into clinical outcomes and trends associated with LVAD therapy over the past decade. With an extensive network of 170 active centers across the United States and Canada participating in the INTERMACS registry, a diverse and robust dataset is available for research purposes, enhancing our understanding of LVAD therapy.

The INTERMACS registry collects comprehensive data on patients undergoing mechanical circulatory support, encompassing medical history, demographic information, clinical measurements, adverse events, and device-related details. Data are primarily obtained from medical records and routine post-implant clinical visits adhering to standard care practices. Scheduled clinical visits occur at specific intervals, such as 1 week, 1 month, 3 months, 6 months, and every 6 months post-implantation. These visits include interviews, physical examinations, functional capacity evaluations, neurocognitive assessments, and Quality of Life (QoL) assessments using the Kansas City Cardiomyopathy Questionnaire (KCCQ). INTERMACS centers diligently follow all patients with implanted mechanical circulatory support devices (MCSD). For patients with the MCSD removed without transplantation, data collection continues for 1 year post-explantation. In cases of transplantation, data collection is ceased, and the patient's follow-up is transferred to a transplant database. In the event of a patient transitioning to another INTERMACS center, seamless continuity of data collection and patient management is ensured through deactivation and reactivation procedures.

2.4 Machine Learning Applications in Pre- and Post-LVAD Therapy

In the field of LVADs, machine learning applications have demonstrated significant potential in improving patient outcomes and decision-making [81–83]. By leveraging the rich and valuable data available from sources such as the INTERMACS registry, a national database for patients receiving mechanical circulatory support (Refer to 2.3 for details about INTERMACS), machine learning applications can extract insights and generate predictive models to enhance LVAD management. Machine learning applications can support clinicians in various aspects of LVAD care. They can assist in the selection of suitable candidates for LVAD implantation by considering patient characteristics, disease severity, and potential risks and benefits [30, 51–54, 56, 81–88]. Additionally, machine learning applications enable continuous monitoring of patient data collected from LVADs, including pump performance, hemodynamic parameters, and patient-reported outcomes [81–83, 89–91]. By analyzing this data in real-time, machine learning applications can detect anomalies, predict AEs, and

provide timely alerts to healthcare providers, facilitating proactive interventions and optimizing patient safety and outcomes. In this section, each potential application of machine learning in the LVAD field is discussed, including pre-implant risk scores for patient selection, identification of post-implant AE patterns, and post-implant risk scores for patient management.

2.4.1 Patient Selection: Pre-implant Risk Scores

LVADs have emerged as a valuable therapeutic option for patients with advanced heart failure, necessitating careful patient selection, evaluation, and timing to achieve optimal outcomes. Factors considered in the selection process include heart failure severity, etiology, comorbidities, age, and patient preferences [92, 93]. Thorough evaluation involves clinical assessment, imaging studies (such as echocardiography and cardiac MRI), laboratory investigations, and psychological evaluation, with a particular focus on assessing right ventricular function to predict the success of LVAD therapy [92, 93]. Timing for referral of patients for LVAD implantation is a complex decision [57, 93]. It is often determined based on several factors, including disease progression, severity of symptoms, frequent hospitalizations, inadequate response to medical therapy, and deteriorating quality of life [93]. Collaboration between heart failure specialists, cardiologists, and cardiothoracic surgeons is essential for making timely and informed decisions regarding LVAD referral. Many patients referred for LVAD are in advanced stages of heart failure, rendering them ineligible for life-prolonging therapy due to declining end-organ function and frailty [93, 94]. Nevertheless, achieving a delicate balance between selecting suitable candidates for LVAD implantation and avoiding futile interventions remains a challenge [57, 93].

Due to the evolution of LVAD technology and the crucial role of patient selection in determining outcomes, risk scores and predictive models have been developed to provide a structured approach to medical decision-making. These tools aim to improve the accuracy of patient selection and aid in the informed consent and shared decision-making process for patients, families, and the healthcare team. The development of risk scores and predictive models is an ongoing process driven by the growing clinical experience, technological ad-

vancements, and the identification of further favorable clinical characteristics. As the field of heart failure evolves, it becomes imperative for the heart failure community to continually refine and enhance these instruments.

Many pre-implant risk scores and predictive models have been developed to aid in decision-making for LVAD implantation by assessing the risk of mortality after LVAD implantation [51–54] and AEs such as bleeding [84–86], infection [30,87,88], stroke [33,50], and right heart failure (RHF) [56]. These tools enable a more objective assessment of a patient's likelihood of success with LVAD therapy, taking into account various clinical, demographic, and laboratory variables. By incorporating these models into the decision-making process, healthcare providers can engage in more informed discussions with patients and their families, facilitating shared decision-making and improving patient satisfaction.

Examples of predictive models for LVAD outcomes include risk scores such as the HeartMate II Risk Score (HMRS) [51], Destination Therapy Risk Score (DTRS) [52], and the Penn-Columbia Risk Score [53]. These models commonly rely on a limited set of clinical variables. However, more recent predictive models, such as the Cardiac Outcome Risk Assessment (CORA), consider a wide range of features [54]. These models vary in their utilization of specific medication doses, types of devices, extrapolated effects of particular devices, exact laboratory and cardiopulmonary hemodynamic data, and even outdated LVADs that are no longer in use. For instance, DTRS assesses the risk of 90-day in-hospital mortality by considering nine clinical variables, including platelet count, serum albumin, international normalized ratio (INR), vasodilator therapy, mean pulmonary artery pressures, aspartate aminotransferase, hematocrit, blood urea nitrogen, and intravenous inotropes [52]. On the other hand, HMRS assesses the risk of 90-day mortality based on factors such as age, creatinine, INR, albumin, and the volume of the center [51]. Additionally, many models have been developed to assess the risk of AEs after LVAD implantation, such as post-implant RHF. Over the last two decades, twenty RHF risk prediction models have been derived and validated in diverse cohorts of heart failure patients undergoing LVAD therapy [56]. These models consider various input variables, including demographic factors (e.g., age and obesity), pre-operative status (e.g., pre-operative inotrope dependency and cardiac surgery), hemodynamic parameters (e.g., pulmonary vascular resistance and heart rate), echocardio-

graphic findings (e.g., severe right ventricular dysfunction), and laboratory test results (e.g., INR and blood urea nitrogen) [56]. The benefits of these risk scores and predictive models lie in their ability to assist healthcare professionals in evaluating the potential benefits and risks of LVAD therapy for individual patients.

It is essential to recognize the limitations and potential pitfalls associated with these models. Many of these risk scores are generated based on data from a single center, thus raising questions about the generalizability of these risk scores. Furthermore, the selected population for some of these studies limits their usability. For instance, heart transplant candidates were not considered in the development of DTRS for predicting mortality after LVAD, which restricts its application specifically in the modern VAD era [52, 57]. Additionally, there is a lack of a global definition for some AEs, such as RHF [56]. As a result, the diagnostic criteria for AEs were heterogeneous among studies, limiting the ability to compare risk scores and potentially causing confusion among clinicians when choosing the appropriate risk scores for their patients. Moreover, many of these studies suffer from a systematic overfitting issue due to the absence of appropriate train and test data splitting [54, 56], which limits the generalizability of the risk scores. Lastly, in the context of LVAD, an important consideration that is often, if not always, overlooked is the imbalanced distribution of outcomes. When developing a classifier for desired outcomes such as mortality and various types of AEs such as RHF, bleeding, and infection, there is a dearth of training data for the minority class. For instance, 90-day mortality after LVAD is less than 90%. Therefore, all risk models should have considered appropriate techniques (such as data sampling) to compensate for the data imbalance [95–97], as well as the choice of an appropriate evaluation metric (such as precision) that emphasizes the performance of the model in predicting the minority class. Although all of these studies used ROC to evaluate the performance of their models, studies have shown that ROC can portray an overly optimistic performance of the model when there is an imbalance issue in the data [98–101]. Therefore, the lack of an accurate evaluation of these models also contributes to the limitation of the practical usage of these risk scores in clinical settings.

In conclusion, risk scores and predictive models have emerged as valuable tools in patient selection and decision-making for LVAD therapy. They play a crucial role in the informed

consent and shared decision-making process. However, it is important to recognize that these models have limitations, such as reliance on single-center data, limited generalizability, and imbalanced outcome distributions. Therefore, ongoing research and refinement are necessary to address these limitations and ensure the continued relevance and utility of these models in clinical practice.

2.4.2 Insights into Adverse Events: Post-LVAD Pattern Mining and Clustering

LVADs have emerged as a valuable therapeutic option for patients with advanced heart failure, significantly improving survival rates [18, 102]. However, the occurrence of adverse events (AEs) after LVAD implantation poses challenges to clinicians and necessitates a comprehensive understanding of their patterns and characteristics [20, 24, 102]. Analyzing these events can provide valuable insights that aid clinicians in making informed decisions and developing effective clinical decision support systems for post-LVAD care.

The INTERMACS Events database, a comprehensive longitudinal dataset comprising data from thousands of LVAD patients, provides valuable insights into the occurrence and outcomes of AEs [18,102]. The Events database, organized chronologically, contains a wealth of information about AEs and final outcomes of LVAD patients, all associated with precise time stamps.

One crucial aspect is the temporal analysis of the AE journey of LVAD patients. While previous studies utilizing this data set have primarily focused on individual AEs such as right heart failure [103], bleeding [42, 45, 49, 104], and infection [30, 31, 46, 105–107], stroke [43, 47, 48, 50, 108, 109] and others, there remains a research gap in investigating the timeline of the entire AE journey experienced by LVAD patients. Specifically, important aspects such as the duration between the first and last AE, the time gaps between AEs, and the most common occurrence time of the initial or final AE have yet to be thoroughly explored. Understanding the temporal aspect of AEs can provide valuable insights into the progression and trajectory of AEs and can help clinicians anticipate and manage AEs more effectively.

Traditional statistical methods have been utilized in previous clinical studies to explore the influence of risk factors, such as demographic information, preoperative status, hemody-

namics, and lab value, as well as individual AEs on outcomes of LVADs [49,51–53,56,84,103, 109]. However, these studies have certain limitations. They often lack generalizability due to their reliance on case studies or data from a single hospital with limited patient numbers, typically a few hundred patients, compared to the larger INTERMACS dataset. Moreover, these studies predominantly rely on traditional statistical methods like multivariable Cox-Regression Models to establish hazard ratios, odds ratios, and event rates, without incorporating modern data mining techniques to model patterns of post-LVAD AEs. Additionally, some studies are cross-sectional in nature and fail to consider the temporal transitions between AEs following LVAD implantation, treating each AE type as a separate event and disregarding their potential interrelation. However, LVAD-associated AEs rarely occur in isolation, as they are closely related in terms of their underlying causes. For instance, a gastrointestinal (GI) bleed can lead to subsequent GI bleeds due to factors like von Willebrand factor dysfunction and arteriovenous malformations introduced by continuous flow LVAD [110]. Similarly, bleeding and stroke can occur sequentially in LVAD patients due to challenges in finding the right balance of anticoagulation [111]. Therefore, it is crucial to not only explore effect measures such as hazard ratios [61, 112, 113] but also utilize data mining techniques to uncover the interplay and sequencing of AEs that commonly occur in LVAD patients, including their combinations of common pairs, triplets, or quadruplets. Recent studies have acknowledged that AEs themselves can act as risk factors for subsequent events, indicating the existence of chains of AEs. While some investigations have explored one-to-one relationships, such as the influence of infection on the risk of thrombosis [112,113], or one-to-many relations between AEs, such as the association of a primary AE with subsequent AEs [61], there has been limited exploration of complex patterns involving multiple AEs in a specific order.

Utilizing data mining techniques, particularly frequent sequential pattern mining, enables the discovery of recurring patterns and the identification of co-occurring AEs [114–119]. The analysis of the transition between AEs and their interrelation can provide valuable insights into the underlying causes and potential preventive measures. In addition to sequential pattern mining, clustering analysis plays a vital role in gaining insights into the AE journeys of LVAD patients. While sequential pattern mining focuses on discovering patterns in the

temporal order of AEs, clustering analysis provides a complementary perspective by grouping similar AE profiles together based on their characteristics [114–119]. These subgroups can provide valuable information about the underlying mechanisms and potential risk factors associated with specific sets of AEs. By identifying these distinct subgroups, clinicians can tailor their treatment strategies and interventions to address the unique needs and challenges faced by patients within each cluster.

Clustering analysis provides a holistic view of the overall landscape of AEs in the LVAD population by identifying subgroups with specific patterns of AEs, ranging from small to large groups of patients. These insights gained from clustering guide clinicians and engineers in optimizing the outcome of LVADs. For instance, it suggests allocating more effort and resources to AE patterns that involve a larger group of patients rather than a smaller group. Additionally, clustering analysis highlights the importance of subgroups whose specific patterns of AEs significantly impact the outcome of LVAD therapy. If a subgroup demonstrates a high rate of mortality or poor clinical outcomes due to their AE patterns, prioritizing interventions and optimization strategies for this subgroup becomes crucial. By understanding the specific challenges and risks faced by these patients, clinicians can tailor their approaches to mitigate adverse outcomes and enhance the overall effectiveness of LVAD therapy.

In summary, gaining insights into the patterns of patient AEs after LVAD implantation is critical for optimizing therapy and improving patient outcomes. The integration of clustering analysis alongside sequential pattern mining enhances our understanding of the complex nature of AEs in patients receiving LVAD therapy. It provides a comprehensive framework to identify distinct subgroups, explore associations between AEs, and inform clinical decision-making. These data-driven approaches empower clinicians to make timely and informed decisions regarding targeted strategies for AE prevention and management. Ultimately, leveraging these analyses helps optimize LVAD therapy, improve patient outcomes, and minimize the burden of AEs in this population.

2.4.3 Post-LVAD Management: Post-LVAD Risk Scores

To enhance the quality of life for LVAD patients and potentially expand treatment options for individuals in earlier stages of heart failure, improving patient management has become imperative. Despite technological advancements, the rate of rehospitalizations remains high, with 218 rehospitalizations per 100 patients within a twelve-month period [120]. To alleviate this burden, one potential approach is to proactively identify early signs of deterioration that can lead to significant consequences and promptly modify patient treatment [121].

There are various sources of data that can aid in this endeavor. Patient monitoring devices, such as pulmonary artery pressure sensors, left atrial pressure sensors, and pump controllers, play a vital role in improving LVAD outcomes. These devices enable the selection of optimal LVAD pump speed and the diagnosis of suction events [89–91, 122–124]. Additionally, unstructured data, such as photographic evidence of LVAD driveline wound infections, can facilitate timely intervention and response to therapy [125–127]. Structured data obtained from electronic health records, including national registry data from INTERMACS, also plays a significant role. By leveraging these diverse data sources, healthcare professionals can gain comprehensive insights into patients' conditions, enabling personalized interventions and ultimately improving patient outcomes.

Recent studies have also explored the incorporation of post-implant structured data into pre-implant models to investigate the improvement in accuracy for assessing the risk of AEs and mortality. *Kouroua et al.* developed a risk score that assessed the survival probabilities within a 12-month period after VAD implantation by incorporating both pre- and post-implant data [126]. Their model achieved a sensitivity of 87% and specificity of 82% for 1-year mortality prediction. Similarly, *Felix et al.* developed bleeding risk scores for different time frames (next 3, 7, and 30 days), with discrimination ranging from 0.78 to 0.79 AUC-ROC [127]. However, these studies face several limitations. Firstly, they were based on a single center with a limited number of patients, which raises concerns about generalizability. Secondly, the high risk of overfitting is evident as the data was not split into training and test sets, likely due to limited data availability. Although a k-fold cross-validation approach was used, the authors did not report the standard deviation. These issues highlight the

need for further validation and external validation studies involving larger, diverse patient populations. Furthermore, the rarity of mortality or bleeding events per month creates an imbalance in the data. This imbalance can lead to misleading results when using traditional evaluation metrics such as ROC analysis, as demonstrated by *Movahedi et al.* [128]. It is crucial to evaluate the models using precision-recall curves or F-scores that emphasize the performance of classifiers for minority classes. Addressing these limitations and refining the models will contribute to enhancing the accuracy and generalizability of risk prediction models in LVAD management. By leveraging machine learning techniques and incorporating both pre- and post-implant data, clinicians can improve patient outcomes, guide personalized treatment strategies, and optimize the overall care for LVAD patients.

2.5 Clinical Decision Support Systems

Clinical Decision Support Systems (CDSSs) are computer systems designed to provide personalized knowledge and information to clinicians, patients, and other individuals, aiming to enhance healthcare delivery [129, 130]. These systems leverage data and observations that may be inaccessible or incomprehensible to humans, contributing to improved decision-making. CDSSs serve multiple purposes, including diagnosis, treatment prediction, recommendation, prognosis, and risk prioritization. By combining human expertise with embedded knowledge, CDSSs act as a valuable "second set of eyes" for clinicians, promoting patient safety, improving care quality, and increasing healthcare efficiency [131, 132]. They are particularly useful in low-resource settings as they can help reduce healthcare costs [132, 133].

There are two main types of CDSSs, defined in the literature as knowledge-based and non-knowledge-based systems [131, 134, 135]. Knowledge-based CDSSs rely on medical guidelines and established knowledge, which can be literature-based or practice-related [135, 136]. On the other hand, non-knowledge-based systems utilize artificial intelligence, machine learning, or statistical pattern recognition [131, 135]. Non-knowledge-based CDSSs analyze historical clinical data to identify patterns and develop predictive models. These models can then provide recommendations to clinicians, enhancing the accuracy of decision-making and min-

imizing medical errors through logical processes based solely on input data [131]. However, it is crucial to understand the logic of these systems and critically appraise the recommendations given by CDSSs [137]. Additionally, the quality and quantity of the data used to train these models are essential factors, as biased or incomplete data can lead to biased or incorrect predictions and decisions [137, 138].

CDSSs have been developed and applied across a wide range of medical specialties and conditions. They have been utilized in areas such as rare diseases [139], oncology [140, 141], chronic obstructive pulmonary disease [142], Alzheimer's disease [143], diabetes care [144–146], and heart disease [147, 148]. The integration of CDSSs into clinical practice holds the potential to improve patient outcomes, enhance decision-making accuracy, and streamline healthcare processes.

In the field of LVADs, a CDSS named SensorART was proposed in 2011 with a focus on the management and remote treatment of patients with LVAD [149–152]. It offered an interoperable and extendable solution, independent of the specific type or brand of LVAD being used. The platform incorporated various hardware and software components to holistically enhance the quality of patient treatment and streamline the workflow of specialists. One key component of SensorART is the *Specialist's CDSS*, a web-based tool designed to assist specialists in designing therapy plans for patients before and after LVAD implantation. The SDSS analyzes patient data, extracts new knowledge, and aids in making informed decisions. It encompasses several tools covering different aspects of VAD therapy, including statistics, association rules, monitoring, treatment, weaning, and speed and suction detection. The SDSS and its modules have shown promising results through testing on numerous patients. However, despite its potential for success, SensorART and its CDSS component did not progress beyond the prototype stage. This could be attributed to its limitations, stemming from the fact that some CDSS components were trained with rather limited datasets (less than 100 patients) and experienced overfitting issues in the predictive components of CDSS. Moreover, they did not propose any plan of action regarding where, when, and how the CDSS should be utilized in clinical settings, nor did they address the evaluation of the practicality of their tools.

3.0 Areas of Investigations

3.1 A Deep Dive Into the INTERMACS Event Database

3.1.1 Motivation and Scope

The survival rate of patients with advanced heart failure receiving LVAD therapy has steadily increased [102]. However, the occurrence of adverse events (AEs) remains a concern [24, 102], leading to a growing focus on clinical studies. The publicly available INTERMACS Events database contains longitudinal data from over 18,000 patients, documenting more than 175,900 recorded AEs following LVAD implantation [20]. This dataset organizes AEs and final outcomes chronologically, providing a natural ordering and inherent semantic structure. Previous studies have focused on examining the timing of individual AEs, such as e.g., right heart failure [103], bleeding [104], infection [105]. However, there has been a notable gap in investigating the overall timeline of the entire AE journey experienced by LVAD patients. For instance, understanding the duration of the AE journey (time between the first and last AE) or identifying the most common occurrence times of the first or last AE remains unexplored.

Moreover, most previous statistical analyses have focused on exploring the influence of various risk factors, such as demographic information, preoperative status, hemodynamics, and lab values, on outcomes and specific AEs, such as right ventricular failure and bleeding [51–53, 56, 84]. Only recently have researchers started considering AEs themselves as risk factors for subsequent AEs, i.e., chains of AEs [61, 112, 113]. However, previous studies have primarily focused on one-to-one relations, such as the influence of infection on the risk of thrombosis [112, 113], or one-to-many relations, such as the relationship between a primary AE and subsequent AEs [61]. There has been limited exploration of chains of AEs, such as the probability of neurological dysfunction following an episode of infection that occurs after an episode of bleeding. Additionally, previous studies have predominantly focused on effect measures, such as hazard ratios [61, 112, 113], overlooking the identification of common

combinations of AEs, such as pairs, triplets, or quadruplets. Therefore there remains an opportunity to explore and discover how many times what types of AEs in what order and at what time points after implant occurred in the AE journey of patients receiving an LVAD. This is tantamount to exploring which AEs occur concomitantly; how many patients experienced the exact same AE journey or share one or more AEs in the same order, etc.

3.1.2 Plan of Actions

This study has two main intentions. Firstly, it aims to provide a comprehensive understanding of the temporal aspect of the INTERMACS Events dataset by posing specific descriptive questions about the timing of the first or last AE, the time span of AE journeys, the time gaps between AEs, and conducting time-to-event analysis. Secondly, the study intends to assess the similarity and diversity of AE journeys among LVAD patients by utilizing frequent sequential pattern mining on the multi-center National database of longitudinal data for post-LVAD AEs. Simple to advanced descriptive data mining methods are applied to 86,912 recorded AEs of 15,820 patients who received a continuous flow-LVAD between 2008 and 2016, extracted from the publicly accessible INTERMACS registry.

3.2 Pre-LVAD Predictive Risk Models

3.2.1 Motivation and Scope

The continuous evolution of LVAD technology and the pivotal role of patient selection in determining outcomes have led to the development of risk scores and predictive models, including risk scores for mortality and adverse events (AEs), specifically those associated with increased morbidity and mortality, such as right heart failure (RHF) [51–54, 56]. These tools serve to provide a structured approach to medical decision-making, aiming to improve the accuracy of patient selection, facilitate informed consent, and support shared decision-making among patients, families, and healthcare teams. However, previous risk scores and predictive models for LVAD therapy have shown limitations. They are often based on data

from single centers or limited selected populations, raising concerns about generalizability [52, 57]. Moreover, the poor overall performance of existing models, primarily due to poor-to-modest discrimination on validation data resulting from systematic overfitting [54, 56], further highlights the need for improved risk models for assessing mortality and RHF after LVAD. Furthermore, the neglect of class imbalance in the development and assessment of these models underscores the necessity for enhanced risk models that appropriately address the issue of class imbalance when evaluating mortality and RHF following LVAD.

3.2.2 Plan of Actions

This study aims to explore the possibility of improving the discrimination power and generalizability of risk models for mortality and RHF after LVAD. To achieve this, the study takes several approaches, including deriving models from big data using the publicly accessible INTERMACS national registry, which includes over 20,000 patients from more than 180 hospitals [20]. To avoid overfitting, the data is split into training and validation datasets, and repeated cross-validation is employed during the training process. Boosting-based methods such as random forest and XGBoost are utilized to reduce the chance of overfitting and model variance [153–155]. This study compares the performance of two classifiers for predicting 90-day mortality after LVAD implantation; the well-known HeartMate Risk Score (HMRS) and a Random Forest (RF) that was derived de novo from a large multi-center registry data. Additionally, the study highlights the imbalance problem in the context of LVAD classifiers for mortality and introduces an alternative metric that is sensitive to the data imbalance. This alternative metric better assesses the model's performance in predicting the minority class.

3.3 Post-LVAD Clustering

3.3.1 Motivation and Scope

The motivation for conducting clustering analyses for post-LVAD AEs stems from the findings of the previous section, "A Deep Dive Into the INTERMACS Event Database" (refer to section 4), which revealed the infrequency of identical AE patterns among patients and emphasized the need to explore alternative approaches. While identical patterns may be rare, it is still possible to identify similar patterns through different methodologies. One strategy is to cluster patients into subgroups based on shared trends and timelines of AEs. By grouping patients with comparable AE patterns, meaningful associations and patterns can be extracted within each cluster, enabling researchers and clinicians to gain a deeper understanding of the underlying factors and associations driving specific AE patterns.

3.3.2 Plan of Actions

In this thesis, two clustering analyses were conducted to address the rarity of identical AE patterns among patients. The first clustering analysis, presented in section 6.1, focused solely on grouping patients based on the similarity in types of AEs, without considering the timeline of AE journeys. The aim was to emphasize the grouping of patients solely based on AE types. The second clustering analysis in this thesis (section 6.2) takes into account the time intervals between AEs and the combinations of AEs within the same time interval. Furthermore, the second clustering analysis compares the resulting subgroups obtained from different choices of time granularity for aggregating AE records. This comparison aims to evaluate the effect of loose versus compact time granularity on post-LVAD clustering.

3.4 Post-LVAD Mortality Risk Models

3.4.1 Motivation and Scope

LVADs have demonstrated significant improvements in survival rates and quality of life for patients. However, they also carry a heightened risk of various adverse events (AEs), particularly in the early weeks following implantation, which can result in morbidity, mortality, and a diminished quality of life [61–65]. Identifying patients at high risk of mortality through their AE history enables close clinical supervision and necessary treatment modifications. While existing risk prediction models for LVAD patient mortality rely mostly on pre-implantation data [53, 54, 57–60], they fail to account for changes in patients' health status over time after LVAD implantation. Notably, one study demonstrated improved 1-year mortality prediction by incorporating only the most recent post-implantation laboratory values alongside pre-implant data [126]. Moreover, a few studies have highlighted the potential impact of post-LVAD therapies, such as anticoagulation, antiplatelet, and loop diuretic treatments, on the risk of AEs like bleeding and thromboembolic complications, which in turn can increase the likelihood of mortality [66–70]. This presents an opportunity to augment pre-implantation data with follow-up data, including laboratory values, hemodynamics, medication treatments, and patients' AE history, to enable real-time prediction of future risk.

3.4.2 Plan of Actions

This study builds upon previous research by integrating the history of adverse events (AEs) and pre- and post-implantation data to predict mortality within the early weeks following LVAD implantation, which is considered the most critical period associated with the highest rates of death and various types of AEs. Three random forest models are developed specifically for predicting death within distinct time intervals: the second to fourth week (W2-W4), the third to fourth week (W3-W4), and the fourth postoperative week (W4). Each model incorporates pre-implantation data collected 48 hours before the implant, data from the first week of post-implantation follow-up, and the patients' AE history up until the

time of prediction. These models are constructed using data obtained from the extensive International Registry for Mechanical Circulatory Support (INTERMACS), encompassing records of over 20,000 patients who underwent LVAD implantation across more than 180 hospitals [20].

4.0 A Deep Dive Into the INTERMACS Event Database

4.1 Timelines of Adverse Event [1]

4.1.1 Objective

The intention of this study is to provide a comprehensive panoramic insight into the temporal aspect of the INTERMACS Events dataset by posing six descriptive questions about the timeline of the adverse event (AE) journey in LVAD patients. These questions include: (1) How many occurrences of each type of AE were recorded in each post-LVAD month? (2) What are the most common time spans between the first and last AE? (3) What is the relationship between the time of the first AE and the last AE? (4) How many AEs did patients experience in different time spans? (5) What are the most common time gaps between AEs? (6) What are the percentages of patients who survived or were free from AE over time?

4.1.2 Methods

The descriptive analysis in this study ranges from frequency table and cross-tabulation to mining sequential patterns and association rules. This section includes the details of methodology including SPADE algorithm [156] for sequential pattern mining (4.1.2.3) and Kaplan-Meier approach [157] for survival analysis (4.1.2.2).

4.1.2.1 Study Population

The flow diagram in Fig. 5 outlines the criteria for inclusion of 86,912 recorded AEs of 15,820 patients (mean age of 57; 12,429 male vs. 3,378 female) who received a continuous flow-LVAD between 2008 to 2016, extracted from the publicly accessible INTERMACS *Event*

[1] The majority of this section is taken from our published work: Movahedi, Faezeh, Manreet K. Kanwar, and James F. Antaki. "Timelines of Adverse Event Journeys of LVAD Patients." *Artificial Organs* (2023).

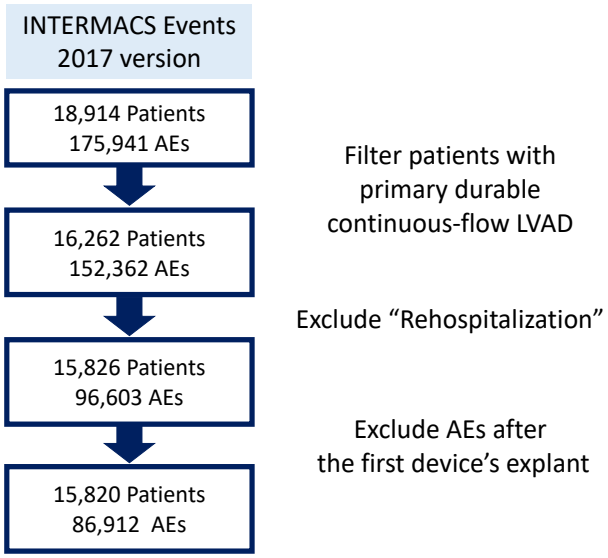


Figure 5: INTERMACS Events Data selection steps.

data set via BioLINCC (2017 version). Of the total 26 types of AEs, "Rehospitalization" was excluded from the study as it is a consequence of an AE rather the AE itself. For patients with multiple device implants, the AEs after the first LVAD explant are excluded. Patients with multiple subsequent LVAD devices are treated differently. Final outcomes, such as death, explant, and transplant, were also included in sequences of AEs as the last element. For the subset of patients who received a right-ventricular assist device (RVAD), the explant of that device was named "REXP". An additional 1,218 patients from the Patient dataset without AEs, hence not included in the Event data set, were used for the time-to-event analysis (including survival analysis) in this study.

4.1.2.2 Survival Analysis

The goal of survival analysis or time-to-event analysis is to describe the probability over time that the event of interest has not yet occurred. In INTERMACS data, the survival time is time from LVAD implant to death or any other final outcomes such as receiving a heart transplant. The survival analysis can be used to answer questions like what percentage of

the patients will survive past a certain time or how do particular characteristics such as age or gender influence the survival probability. The survival function is a function of time (t) and can be represented as:

$$S(t) = \Pr(T > t) \tag{4.1}$$

The survival function takes values in the range between 0 and 1 and is a non-increasing function of t.

One main concept in survival analysis is censoring. Patients are censored when the information about their survival time is incomplete. There are different kinds of censoring, including:right-censoring, interval-censoring, left-censoring. The most frequent one is right-censoring. For example, lets consider the survival analysis for post-LVAD patients. Some of the patients survived 10 years after LVAD implant and thus have not experienced the death event. At the same time, some patients may not come to follow-up visits, they might have actually died, but no confirmation was ever received. Those patients are considered as right-censoring, meaning their true survival time is $\geq$ the observed survival time (10 years after LVAD implant). Right-censored patients are included in estimates of survival probabilities at time points preceding their censoring time point; and excluded from the analysis thereafter.

In 1958, Edward L. Kaplan and Paul Meier published a paper about a non-parametric statistical method of estimating and plotting the survival probability as a function of time [157]. The Kaplan-Meier approach assumes that (1) the event of interest is well-defined and unambiguous (2) the survival probability of all observations is the same regradless of the time they have entered the study (3) censored observations have the same survival prospects as observations that continue to be followed. It should be mentioned that the Kaplan-Meier curve has no assumption on the distribution of survival times nor assume a specific relationship between covariates and the survival time. In reality, the true survival function is not known . That is why the Kaplan-Meier approximate the true survival function from the collected data. The estimation is defined as the fraction of patients who survived for a certain amount of time and is formulated as follow:

$$\widehat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \tag{4.2}$$

where, t_{i} is a time when at least one event happened, d_{i} is the number of events that happened at time t_{i}, n_{i} indicates the number of patients at risk at time t_{i}; patients who have not died event or not been censored yet. The survival probability at time t is equal to the product of the percentage chance of surviving at time $\leq t$. The Kaplan-Meier curve plots the estimated survival probabilities against time. The survival line is actually a step-function in which each vertical drop indicates the occurrence of one or more events. Right censoring of patients is typically indicated by a vertical mark at the censoring time.

4.1.2.3 Sequential Pattern Mining

Sequence mining is to discover patterns across time in a given temporal or sequential database. For example, consider INTERMACS dataset where the objects represent patients and the attributes represent adverse events. Table 2a shows 12 episodes of adverse events associated with 4 patients in INTERMACS where the "Patient ID" corresponds to the patients' medical record number and the "Event ID" corresponds to the time (month) of episodes of adverse events. The types of recorded adverse event (items) include Bleeding, Infection, and Hemolysis. The Table 2b presents all the frequent sequences with different lengths that is common among at least 3 out of 4 patients (a minimum *support value* of 75%). The goal of sequential pattern mining is to efficiently find all the frequent sequences as listed in Table 2b. Before entering in the SPADE algorithm description for sequential pattern mining, some terminology will be briefly reviewed below.

Definition 1 (Sequence database) A sequential database (SD) is a set of sequences $S_1, S_2, ..., S_{\mathrm{n}}$ where each sequence S_{i} is a set of events (itemsets) $(e_1, e_2, ..., e_{\mathrm{p}})$ in which repetitions of events are allowed and order of events is important. Each event (itemsets) is a subset of I where I is a set of m unique items $I_1, I_2, ..., I_{\mathrm{m}}$. In the post-LVAD AE domain, INTERMACS Event data consists of a set of sequences of AEs; one sequence per patient like S_1 is the corresponding sequence of AEs for the first patient ($Patient_1$) in database. The items in I can indicate the list of types of AE that patients can experience after LVAD implant (Bleeding, Infection, Neurological Dysfunction, etc. An AE event (itemset) is an unordered set of items (AEs) like $\{Bleeding, Infection\}$ that happened at the same time

Table 2: An example of sequential pattern mining.

(a) An example of INTERMACS Event database.

Database		
Patient ID	Event ID	Adverse Events (AEs)
1	10	{Infection, Bleeding}
	20	Bleeding
	30	{Infection, Bleeding}
2	20	{Infection, Hemolysis}
	30	{Infection, Bleeding, Hemolysis}
	50	Bleeding
3	10	Infection
	30	Bleeding
	40	Infection
4	30	{Infection, Bleeding}
	40	Infection
	50	Bleeding

(b) Frequent sequences.

Frequent set with 1 item	
(Bleeding)	4
(Infection)	4
Frequent set with 2 items	
({Infection, Bleeding})	3
(Infection) → (Infection)	4
(Infection) → (Bleeding)	4
(Bleeding) → (Infection)	3
(Bleeding) → (Bleeding)	3
Frequent set with 3 items	
({Infection Bleeding}) → (Bleeding)	3

interval. A sequence is a chronological ordered record of AE events (itemsets), for example, $S_1 = (Bleeding - \{Infection, Cardiac\ Arrhythmia\} - Neurological\ Dysfunction)$. S_1 indicates that $Patient_1$ experienced 3 AE events including first a Bleeding AE and then later Infection AE and Cardiac Arrhythmia AE at the same time, and finally had a Neurological Dysfunction AE. It should be noted that (1) brackets are omitted if an event (itemset) contains one item only (2) the order of items within an event is not important, it is usually taken in alphabetical order (3) In a sequence, the concept of the event e_1 occurs before e_2 is denoted as $e_1 \leq e_2$ (4) a type of AE (item) can occur only once in an AE event (itemset), but it can occur several times in different AE events (itemsets) of a sequence.

Definition 2 (Length of a sequence) The length of a sequence $S_{\mathrm{i}} = (e_1, e_2, ..., e_{\mathrm{k}})$ is defined as the total number of events (itemsets) k. For instance, the length of $S_1 = (Bleeding - \{Infection, Cardiac\ Arrhythmia\} - Neurological\ Dysfunction)$ is 3 ($len(S_1) = 3$).

Definition 3 (Subsequence) a sequence of $S = (e_1, e_2, ..., e_{\mathrm{len(S)}})$ is a subsequence of another sequence $S\prime = (e\prime_1, e\prime_2, ..., e\prime_{len(S')})$, denoted as $S \preceq S\prime$, if and only if there exists a sequence of integers $1 \leq i_1 \leq i_2..... \leq i_{\mathrm{len(S)}} \leq len(S\prime)$ such that $(e_1 = e\prime_{\mathrm{i}_1}) \wedge (e_2 = e\prime_{\mathrm{i}_2}) \wedge ..... \wedge (e_{\mathrm{len(S)}} = e\prime_{\mathrm{i}_{\mathrm{len(S)}}})$. For example, the sequence $(Bleeding - \{Infection, Cardiac\ \ Arrhythmia\})$ is a subsequence of $(\{Bleeding, Respiratory\ \ Failure\} - Hepatic\ \ Dysfunction - \{Infection, Cardiac\ \ Arrhythmia, Neurological\ \ Dysfunction\})$, since $Bleeding \subseteq \{Bleeding, RespiratoryFailure\}$ and $\{Infection, CardiacArrhythmia\} \subseteq \{Infection, CardiacArrhythmia, NeurologicalDysfunction\}$, and the order of events is preserved.

Definition 4 (Support) A sequence S is said to contain another sequence P, if P is a subsequence of the sequence S; $P \preceq S$. The total number of sequences in the database SD that contain a subsequence of P is the frequency or support value of P. If a sequence's support is $\geq$ a minimum support value defined by the user, it is considered a frequent sequence. If the frequent sequence is not a subsequence of any other frequent sequence, it is called maximal.

Definition 4 (Sequential pattern mining) Given a database of SD and minimum support value, sequential pattern mining is to find the complete list of all frequent sequences in the

SD with support values $\geq minimumsupport$.

Definition 5 (Sequence rules) Rules describe the interesting relationship between different sequence items and can be generated from the frequent sequences. For example, the sequence $(Bleeding - Infection)$ occurs in four sequences, while $(Cardiac\ Arrhythmia - Bleeding - Infection)$ in three sequences. Therefore, it can be said that if $(Bleeding - Infection)$ occurs together, then there is a 75% chance that $Cardiac\ Arrhythmia$ also occurs. In other words we say that the rule $(Bleeding - Infection) \Rightarrow (Cardiac\ Arrhythmia - Bleeding - Infection)$ has a 75% *confidence*. The confidence value ranges between 0% to 100%. The symbol of $\Rightarrow$ indicates a happens-after relationship between its left side and right side.

Definition 5 (confidence) The confidence of a rule $X \Rightarrow Y$, such as $Bleeding \Rightarrow Infection$, is the the probability of the subsequent occurrence of Y (Infection) following an occurrence of X (Bleeding); conditional probability $P(Y|X)$ $(P(Infection|Bleeding))$.

SPADE Algorithm

Several algorithms have been developed to find the frequent sequences in a sequential database. These algorithms consist of three main categorizes apriori-based, pattern-growth, and early-pruning algorithms which compete with each other to (1) minimize the I/O cost (2) eliminate number of data structures are needed all the time for only computing the support values. This study applied a benchmark sequential pattern-mining algorithm called SPADE (Sequential PAttern Discovery using Equivalence classes) which is developed by *Zaki* in 2001 [156]. SPADE algorithm uses the vertical layout of database and decomposes the search space (lattice) into sub-spaces (sub-lattice) that can be processed separately in main memory. The SPADE algorithm usually requires three scans of the sequence database which decreases I/O costs.

The SPADE algorithm formulates the sequential pattern mining problem as follow: (1) mining the sequences of subsets of items, and not just sequences with single item (2) mining sequences with arbitrary gaps among events (itemsets), and not just the consecutive subsequences. For instance, in Table 2b, the sequence $(\{Infection, Bleeding\}) \rightarrow (Bleeding)$ is a subsequence of AE sequence of Patient ID of 4, although there is an event of $Infection$ between the two events of $(Infection, Bleeding)$ and $(Bleeding)$. The SPADE algorithm

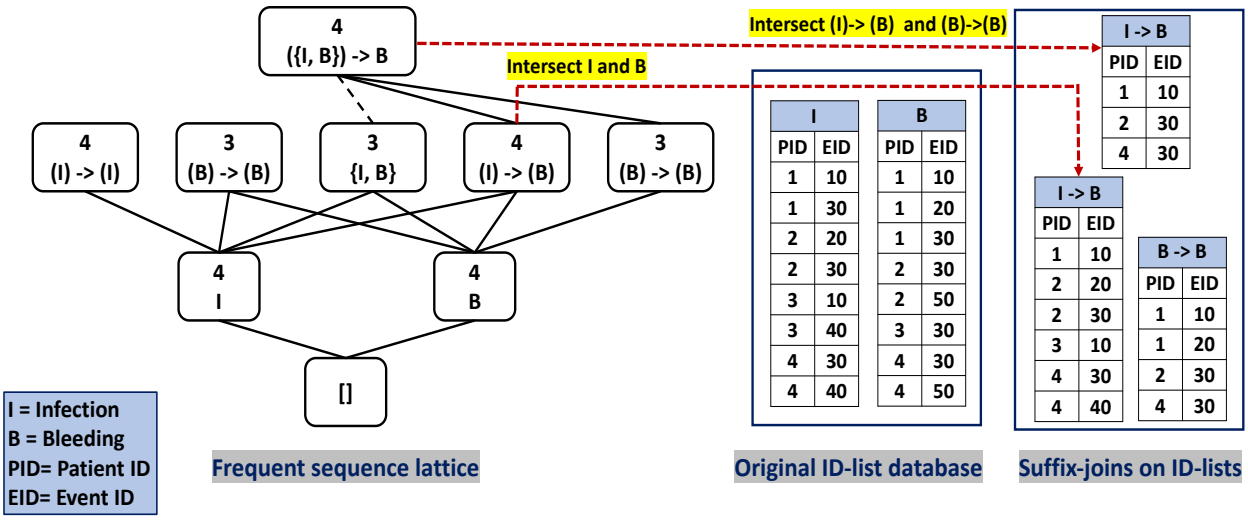


Figure 6: SPADE: a space-efficient temporal ID-list join to compute the support values.

proposes depth-first or breadth-first strategies to search in the search space that is extended by the $\preceq$ (subsequence relation), from the most general (single items) to the most the most specific frequent sequences (maximal sequences) as shown in Fig 6. The maximal sequences are $(\{Infection, Bleeding\}) \rightarrow (Bleeding)$, $(Infection) \rightarrow (Infection)$, and $(Bleeding) \rightarrow (Infection)$. The symbol of $\rightarrow$ indicates a happens-after relationship between its left side and right side.

4.1.3 Results

INTERMACS defines 25 different types of AEs, of which six are classified as *final outcomes*. Fig. 7 plots the distribution of count of AEs per patient. The 54% of patients experienced between 1 to 4 AEs, while a small percentage of patients experienced a very high number of AEs, like 5% of patients experienced 8 AEs. The time-related characteristics of these data were investigated by posing six descriptive research questions as follow:

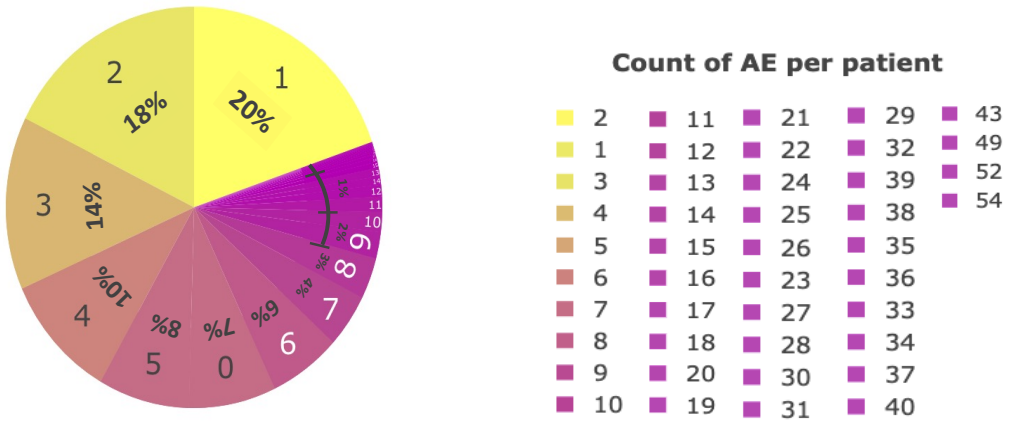
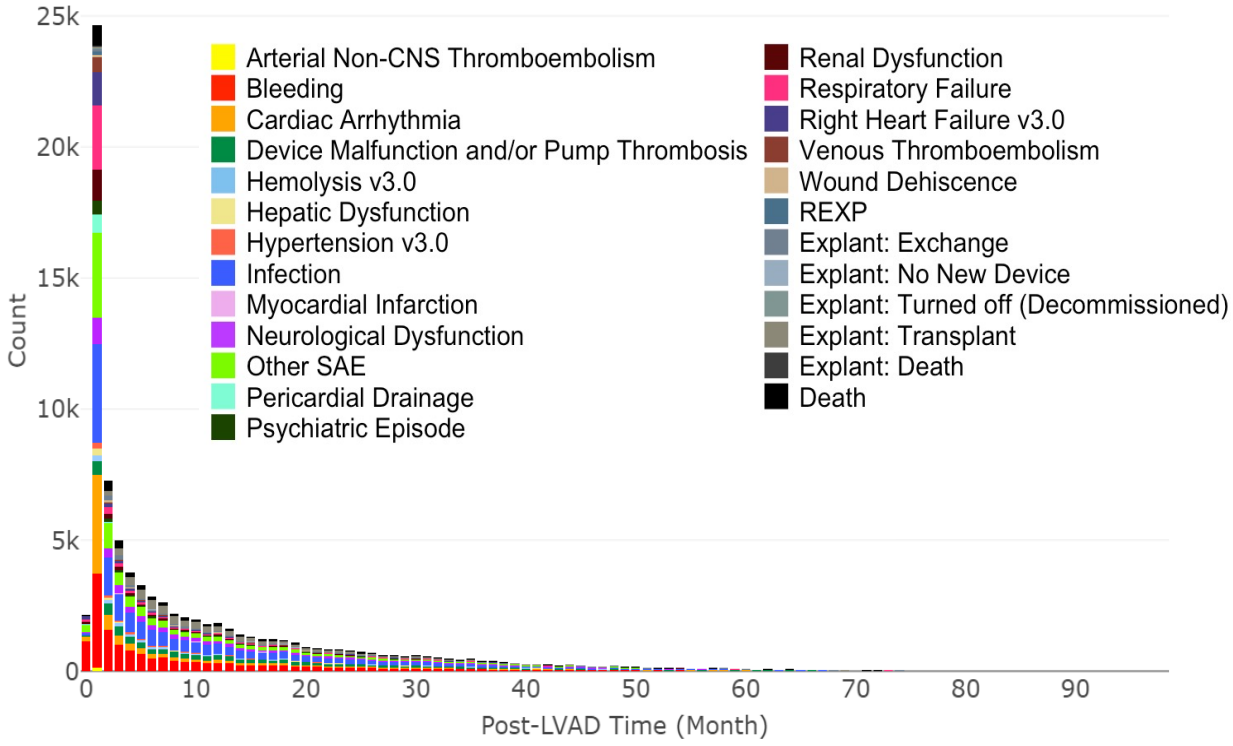


Figure 7: The count of AEs per patient. The pie chart and elements of the legend is ordered based on their frequencies.

4.1.3.1 How Many Occurrences of Each Type of AEs Were Recorded in Each Post-LVAD Month?

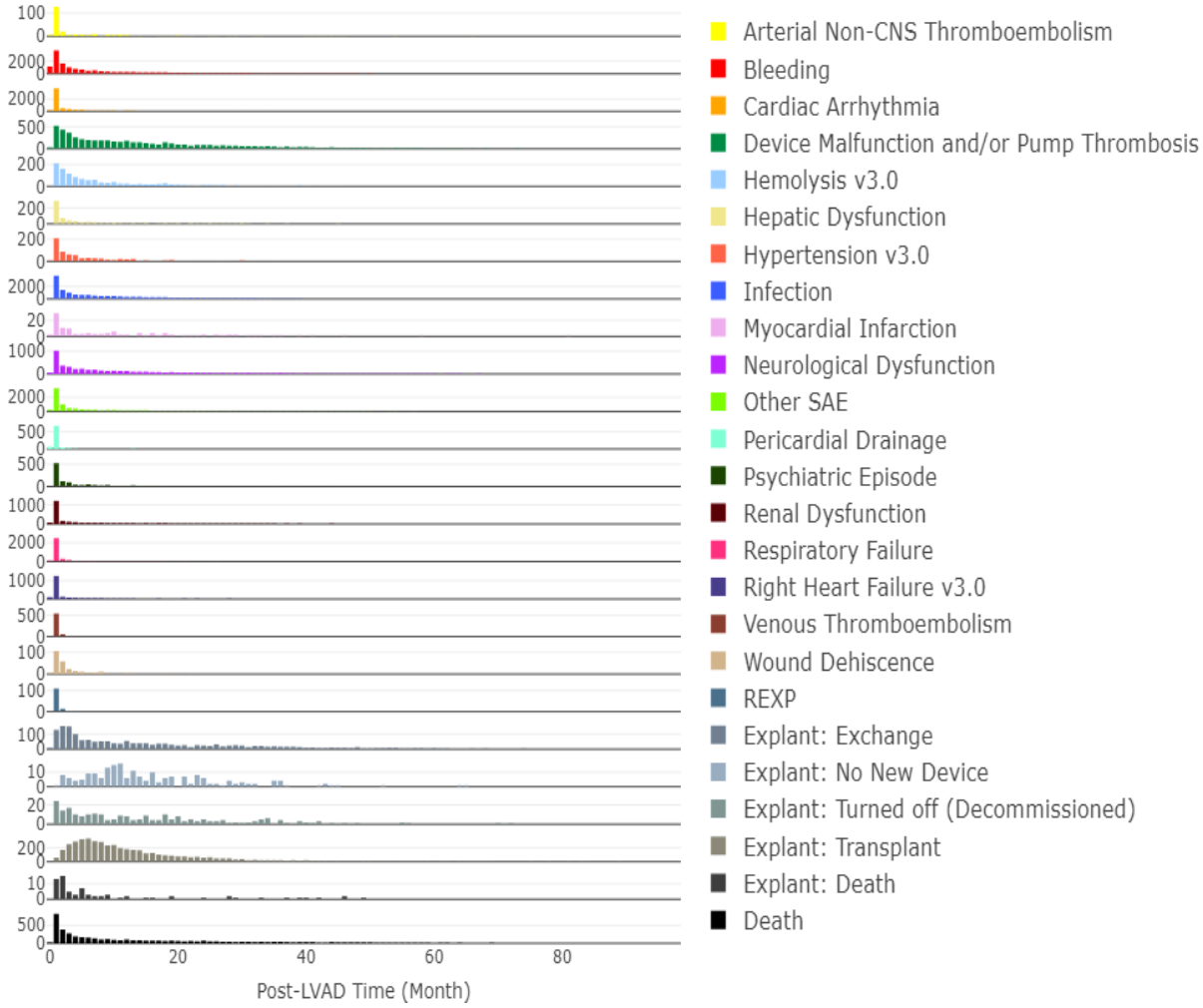
The stacked bar-plot in Fig. 8a provides an overview of all types of recorded AEs per month post-LVAD, color coded as shown in the legend. The majority of AEs occur before the first year, especially in the first month post implant with 24,666 recorded AEs, predominated by four common types of AEs including Bleeding, Cardiac Arrhythmia, Infection, and Other serious AE (SAE), each in excess of 3000 events recorded, while others were uncommon. There were 2,145 recorded AEs at the time of implant (time 0) mostly Bleeding (red). After the second month, the counts of AEs become much more sparse, and virtually imperceptible on this plot. This may pose a problem for predicting the timing of AEs. The overall preponderance of patients with no record of AE (majority class) creates an "imbalance class problem" for any predictive model that is trained on this data set. This will impose a bias towards *freedom from AEs* for most patients (imbalance class problem).

Fig. 46 depicts the distribution of each type of AE separately over the time. Contrasted to Fig. 8a, the individual counts of individual AEs for the first month are much less dramatic than for the aggregate count – due to the diversity of types of AEs. Similarly, the sparsity of



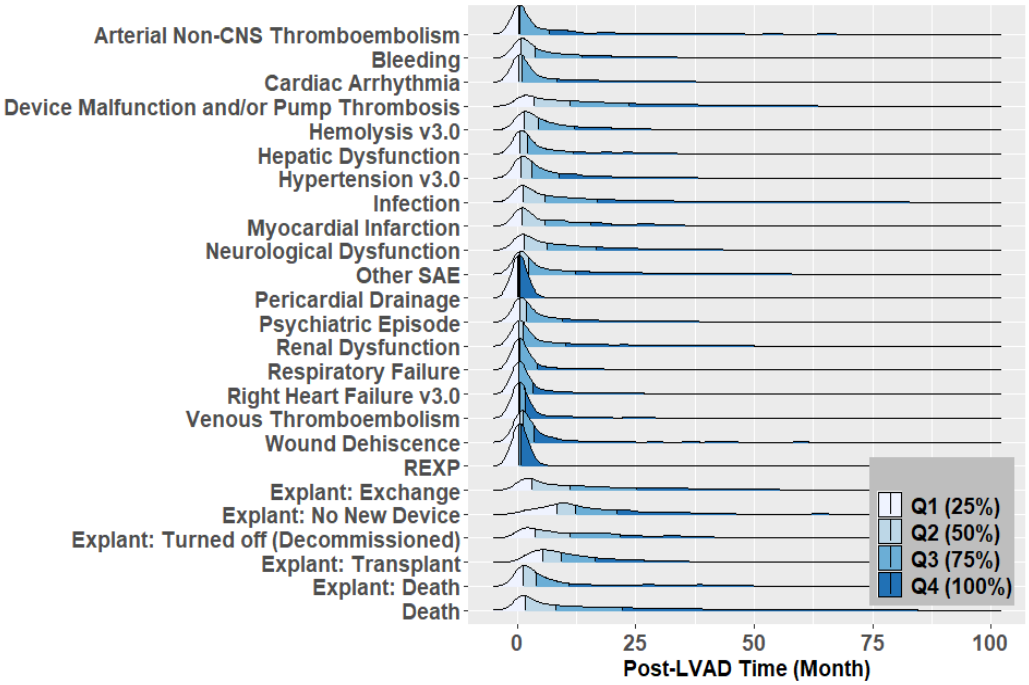
(a) Stacked bar-plot of the total number of recorded AEs in each post-LVAD month; color coded for each type of AE. The explant of a right-ventricular assist device (RVAD) was named "REXP". (SAE - Other Serious Adverse Event.) This refers to any adverse event that is serious but does not fall into any of the predefined categories of serious adverse events. 'v3' indicates the version of the definition of AE in INTERMACS. For complete list of definitions of AEs in INTERMACS see INTERMACS Appendices

Figure 8: Distribution of AEs over time.



(b) Individual bar-plots of distribution of each type of AE over the post-LVAD course.

Figure 8: Distribution of AEs over time.



(c) Density plots of the total number of each type of AE per post-LVAD month, colored based on the quartiles: Q1, Q2, Q3, Q4.

Figure 8: Distribution of AEs over time.

AEs in subsequent months is further apparent when split by type. It is also clear from Fig. 46 that the distribution of most AEs is right-skewed, indicating more occurrences of AEs in months close to the time of implant compared to the later months. In fact several types of AE (Neurological Dysfunction, Renal Dysfunction, Respiratory Failure, and Right Heart Failure) occur almost exclusively in the first month - indicated by a small spike in Fig. 46.

The density plots in Fig. 8c, color-coded by quartiles further illustrates the preponderance of AEs in months closer to the time of implant than in later months. For some AEs like various Explant final outcomes, their density plots are wider (occurrences of AEs are spread over longer period of the time) while the plots for AEs like RVAD explant (REXP) and Pericardial Drainage are narrower with a sharp spike in the first post-LVAD month. For example, almost 100% (up to the end of last quartile, Q4) Pericardial Drainage occurrences happened before the first year. Conversely, most final outcomes like Explant: Transplant have a wider bell-shaped density distribution with only 50% (up to the end of second quartile; Q2) of their occurrences before the first year. For all types of AEs, 75% (up to the end of third quartile; Q3) of occurrences happened before the second year.

4.1.3.2 What Are The Most Common Time Spans Between First and Last AE?

Fig. 9 provides a pie-chart (%) for the wide range of recorded time span (diversity) over which AEs occurred, computed as the time difference between the last and first AE entered into the INTERMACS Event database. The most common time span is 0 (27%; 4,219 patients) – corresponding to those patients who only experienced one or more AEs within one month. The next most common time spans are 1 and 2 months were experienced by 1,121 (7%) and 832 (5%) patients, respectively. A negligible number of patients ($< 1\%$) experienced multiple AEs over a period greater than 23 months – indicating sparsity of the data. The one patient with the longest recorded time span in which they experienced any AEs was 97 months. In conclusion, the timelines of AEs after LVADs tend to be short in duration.

It should be mentioned that the time of last AE here means the time of last recorded AE up to the point of follow-up time for each patient. The follow-up time is not equal for all

the patients because of their different implant years. For example, a patient who had LVAD implant in 2010 had 7 years follow-up data (INTERMACS 2017 version is used for this study) while a patient who had its implant in 2016 had only 1 year follow-up. However, considering that all patients in this study have at least 1 year follow-up and the most common recorded time of AE is during the first year and the most common number of AE per patient is 2 across all the time points, the result of this question would not be changed dramatically by having longer period of follow-up time for patients.

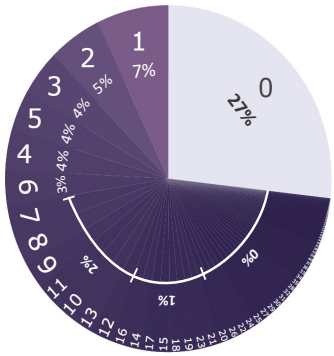
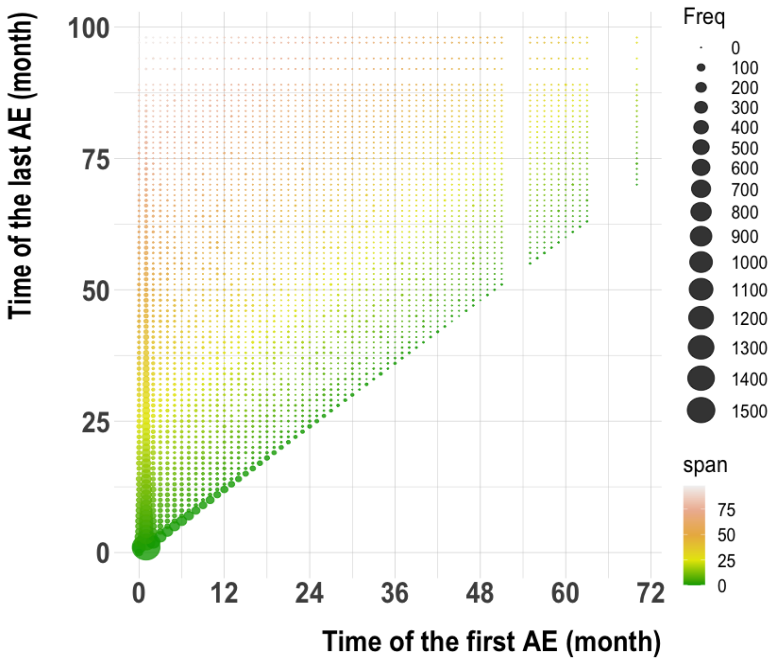


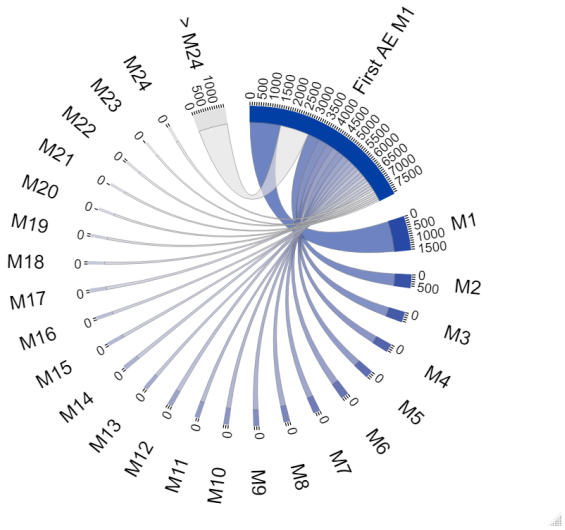
Figure 9: Frequency of time span (difference between the time of last AE and first AE) of AE journeys in months.

4.1.3.3 What Is the Relationship of the Time of First AE and Last AE?

In previous section, the time span between the last AE and first AE was discussed without considering the specific recorded time for the first and last AE within the timeline (history) of AEs. Fig. 10a depicts the relationship of the time of the first AE and the time of last AE. The size of the bubbles represents the frequency for the corresponding pair of first and last AE and their color represents the difference (time span). The time of first AE is distributed over 70 months, with the first month being the most common (n = 1,518) – obscuring 1,769 records of AEs at time 0 (time of implant). The largest bubble is found at the bottom of each column (time of first AE). The time of the last AE ranges from 0 (for patients with



(a) Bubble chart plots the time of first vs last AE. The size of bubbles indicates the frequency; color implies the length of time span.



(b) Chord diagram connects the time of first AE at the first post-LVAD month ("First AE M1") to their corresponding time of last AEs, colored based on the time of last AE. The size of arcs is proportional to the frequency of each connection.

Figure 10: The relationship of the time of first AE and last AE.

only one recorded episode of AE at the time of implant) to 98 months after LVAD. Thus, the wide range of time for the first and last AE illustrates the (diversity) of these data. Likewise the sparsity of these data is readily apparent by the clustering of larger bubbles within the first 24 months, and scattered small bubbles elsewhere. There is only 24% (3,237 patients) of the total 15,820 patients with recorded time of first AE or last AE greater than 24 months.

Further insight can be gained from the chord diagram of Fig. 10b focusing on the widest column in Fig. 10a, Time of First AE = 1 month. The arcs indicate the connection between the first AE (at 1 month) ("First AE M1") to all corresponding recorded times of the last AE, colored based on the time of last AE. The size of arcs is proportional to the frequency of each connection. The thickest and darkest arc corresponds to the largest bubble in Fig. 10a. Likewise, as the bubbles taper rapidly in the first column of Fig 10a, the thickness of arcs reduces rapidly following the first month. The number of patients corresponding to arcs "M1-M1" to "M1-M6" is greater than 300. The number of patients corresponding to arcs "M1-M7" to "M1-M12" is between 200 and 300. Thereafter the frequency below 100. The 1,440 patients with the time of last AE greater than 24^{th} months are aggregated in one arc.

4.1.3.4 How Many AEs Did Patients Experience In Different Time Spans?

Time span over which AEs occurred for a given patient is not necessarily proportional of the total number of AEs. For example, there is a meaningful difference between two patients with the same number and types of AEs in the same order, but one experienced the sequence of AEs in the time span of 2 months vs. the second patient over 23 months. Fig. 11 illustrates the relation between the count of total number of AEs experienced by patients and their time spans. The bubbles tend to be large when both the time span and the total number of AEs are low. For instance, the largest bubbles correspond to 4,219 patients who experience only 1 AE (time span of 0 month). Then 1,121 patients are associated with 2 total AEs with one month separation. 378 patients are associated with 3 total AEs over the

span of 2 months. 225 patients experienced 3 AEs over the time span of 3 months, and so forth. The remaining bubbles for subsequent combinations of time span and total number of AEs account for less than 500 patients in total - indicative of the (sparsity) of these data. The longest time span of 97 months belongs to one solitary patient who experienced 5 AEs. One patient with the greatest number of 26 AEs experienced them over the time span of 45 months.

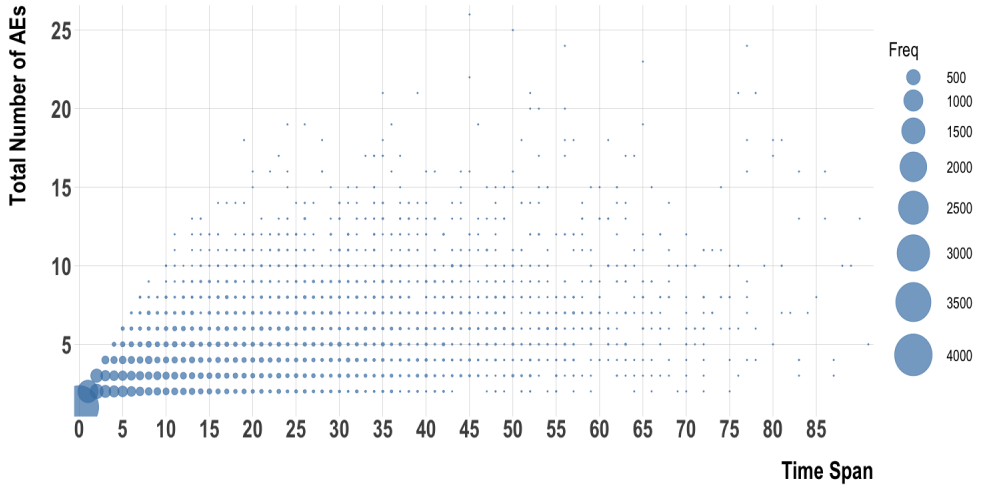


Figure 11: Bubble chart of the count of total number of AEs each patient experienced vs. corresponding time span (in month).

4.1.3.5 What Are The Most Common Time Gaps Between AEs?

For patients with greater than one AE, it could be informative to analyze the duration of freedom from AEs between successive AEs. While time span, above, emphasizes the total duration in which patients experience AEs, the time gap focuses on how scattered (or clumped together) the AEs during that time span are. Consider a case of two patients with the same time span of 40 months and the same number of 4 AEs in the same order but one had all the 3 AEs in the first month and one AE at month 40, while the other had 1 AE every 10-months. These two patients would be considered clinically different from

each other because the first patient experienced an unstable initial month after receiving an LVAD, while the second patient may be considered a "frequent flyer". There are 34,394 time gaps recorded for 14,634 patients with greater than 2 AEs (out of 15,820 total patients in this study). Fig. 12a plots the frequency of each recorded time gap, illustrating the wide range from 1 month to 86 months - indicative of the (diversity) of these data. 61% of the time gaps were less than 4 months; 35%, 16%, 10% for time gaps of 1 month, 2 months, and 3 months, respectively. Any time gap greater than 19 months are negligible ($< 1\%$) indicating the sparsity of these data. Fig. 12b and 12c plot the frequency of recorded min and max values of time gap per patient. For those patients with only two AEs, the min and max gaps are same. The min time gap varies between 1 month and 75 months. The max time gap varies between 1 month and 86 months. More than half (56%) of the total 14,634 patients had a min time gap of 1 month and 15% have 1 month as their max time gap. More than half of the patients (52%) have max time gap of less than 7 months including 15% of patients with max time gap of 1 month.

For patients with recurrence of the same type of AE, it is informative to know how rapidly the subsequent AE occurs - i.e. the time gap between AEs. This is presented in Fig. 12 for two common types of AE, Bleeding Fig. 12d and Infection Fig. 12e based on 6,200 and 6,822 records, respectively, including all types of Bleeding, including gastrointestinal bleeding, device and non-device related Infection. Both charts show that the recurrences of both AEs occurred in a short interval, with one month being the most common. However, recurrent bleeding within one month was more common than infection (35% vs. 23%). The maximum time gaps recorded for Bleeding and Infection were 66 and 80 months. Since the unit of time in this figure is month, it should be noted that any recurrences of AEs less than one month is not counted in this analysis. For example, if the time interval is changed to week, the number of occurrences for Bleeding will increase from 6,200 to 7,726 records.

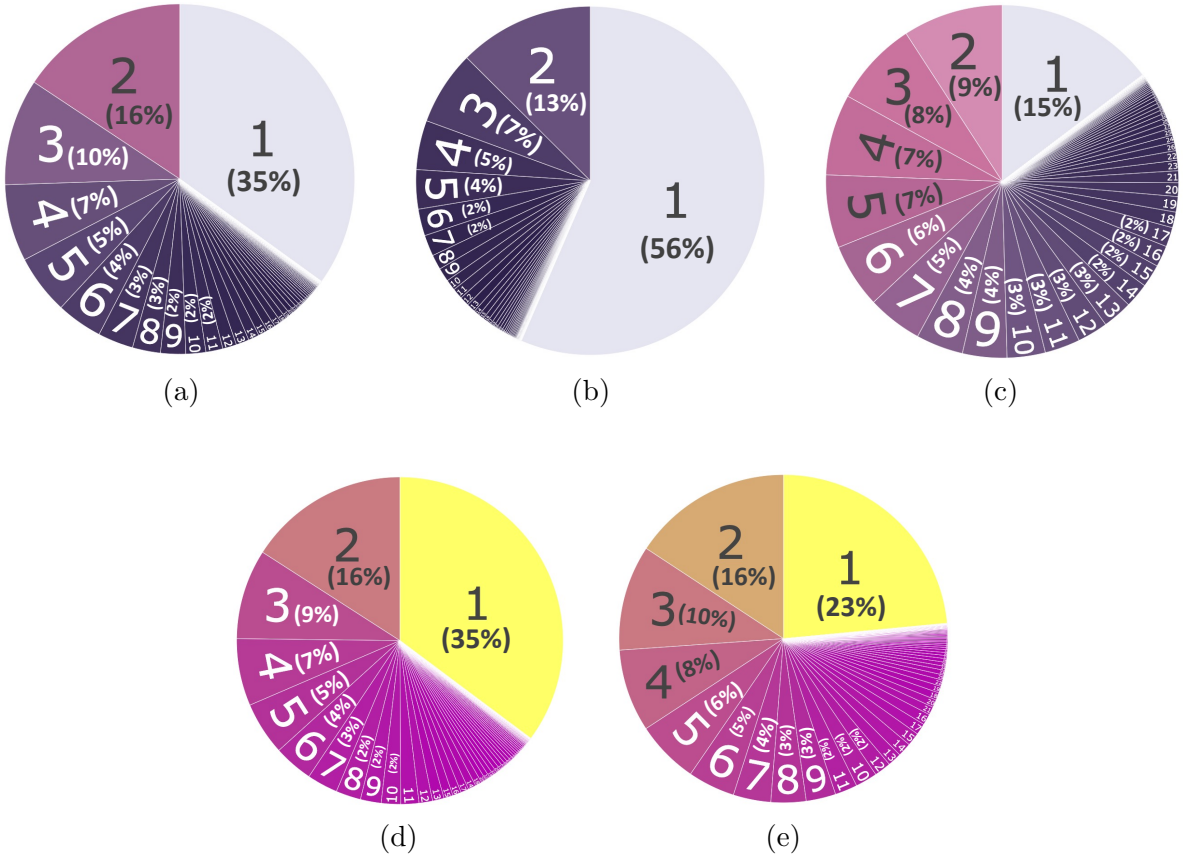


Figure 12: Frequency of (a) the time gaps (in month) between two consecutive recorded times of AEs (b) the minimum recorded time gap per patient (c) the maximum recorded time gap per patient (d,e) time gaps between the recurrences of two consecutive AEs of (d) Bleeding (n= 6,200) and (e) Infection (n= 6,822).

4.1.3.6 What Are The Percentages of Patients Who Survived Or Were Free From AE Over Time?

Fig. 13 shows the time-to-event analysis for Death and any type of AE (free from AE curve), respectively, for the first 15 months of after LVAD. The curves highlight the contrast between survival and freedom from AE. While survival gradually declines, freedom from AE drops rapidly in the first few months post-LVAD. For instance, there's a 92% probability of surviving at two months, but a 30% chance of not experiencing an AE by two months.

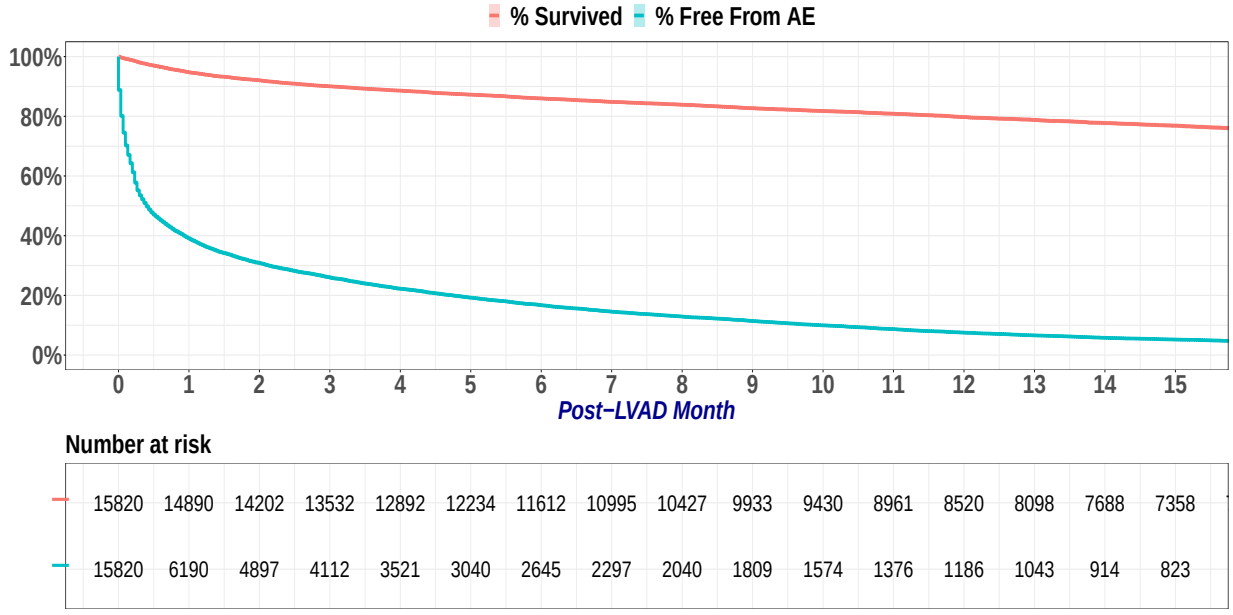


Figure 13: Kaplan-Meier survival curve and freedom from AE for patients with continuous-flow LVAD as the primary durable device therapy.

4.1.4 Discussion

Over the past two decades, several studies have explored the data in the INTERMACS Events registry, including the annual INTERMACS Reports [18,20,79,102,120]. This study, to the best knowledge of the authors, is the first to depict a panoramic view of timelines of AE journeys of LVAD patients. The results of this analysis also revealed the magnitude of

diversity and sparsity of these data. It was shown that the recorded times of AE extends up to the 98^{th} month after LVAD implant, but the first few months are most critical as they have the preponderance of the most common types of AEs. (See question 1; QES.1). The AE timelines tended to be short in duration (QES.2), even for patients with greater counts of AEs (QES.4), with the most common starting point at the 1^{st} month (QES.3). The most common time gaps between AEs ranged from 1 to 4 months (QES.5). Although the survival probability is high in LVAD patients, the probability of being free from adverse events is low (QES.6).

The great diversity and sparsity encountered in the data set posed challenges in the analysis, for example, the choice of post-LVAD time interval, hence granularity; e.g. day vs. week vs. month affected the perception of post-LVAD AE patterns. Characteristics of the AE journey, such as the time span and the number of AEs aggregated in each time interval (day vs week vs month), also affected the complexity of analysis and computational cost. There were tradeoffs to be considered when choosing the time interval. For INTERMACS data with 25 types of AEs, we found that aggregating a greater quantity of AEs by longer time interval (e.g. days to months) could make two similar patients easily dissimilar. On the other hand, by choosing a shorter time interval (day vs week), AE journey became lengthy, and less similar. It should be taken into consideration that these analyses are reported based on the absolute count of AEs, not normalized to the (decreasing) number of patients at risk. In other words, as AEs diminishes over time, so does the number of patients at risk. (See Fig. 8) In summary, we found that the choice of granularity of post-LVAD time should be tailored to the context of the analysis and/or the required sensitivity or specificity. It may be elucidating to consider a variety of time granularity to explore the data from different perspectives.

This data mining exercise endeavored to extract value from the INTERMACS Event database, vis a vis insight about patients' AEs journeys. However, the analysis suffered from limitations in the database: some of which cannot be improved, and others that point to potentially helpful modifications to the data collection protocols. For example, patterns of AEs that do not occur at an isolated point in time, such as Infection, could be better identified if the duration of each episode of infection was recorded. If a patient experienced

an Infection in the first post-LVAD month and then experienced a Neurological Dysfunction event, it would be valuable to know if Infection was resolved at the time of Neurological Dysfunction, or if the two overlapped. The authors also acknowledge that the publicly available INTERMACS registry used in this study does not fully reflect the most contemporary data and devices. As pump technology continues to improve, so does the AE profile as was shown in the recently published MOMENTUM trial data [23]. Future studies would benefit greatly from access to the most current registry data.

4.1.4.1 Clinical prospective

Researchers, industry professionals, and clinicians can benefit from these findings in the following ways: (1) Guidance on critical time points following LVAD implantation that require attention to anticipate AEs and gain a deep understanding of the underlying causes. (2) Assistance in allocating resources and scheduling follow-up appointments during prevalent time points with the highest frequency of AEs to address any potential complications effectively. (3) Awareness of the most common time gaps between AEs, facilitating anticipation of subsequent AEs and enabling the implementation of preventive measures or adjustment of treatment plans as necessary. (4) Knowing the percentage of patients with long AE journeys versus short journeys provides hospitals and the industry with a valuable estimate of the cost of post-LVAD management related to VAD, allowing for better financial planning and resource allocation.

4.1.5 Conclusion

This study sheds light on the timeline of AE journeys experienced by LVAD patients, as recorded in the INTERMACS Event database. Valuable insights are gained by examining the most common duration of AE journeys, the number of AEs within different time spans, the prevalent time of occurrence for the first and last AEs, the time points with the highest frequency of recorded AEs, and the most common time gaps between AEs. Researchers, industry professionals, and clinicians can benefit from these findings to optimize the outcome of VAD therapy, such as effectively allocating resources and scheduling follow-up appointments

during prevalent time points with the highest frequency of AEs or anticipating subsequent AEs through awareness of the most common time gaps between them.

4.2 In Search of Similarity in Adverse Events [2]

4.2.1 Objective

The INTERMACS Event dataset contains an extensive collection of longitudinal evidence detailing the course of AEs in over 15,000 patients who have undergone LVAD treatment. Within this vast dataset lies valuable knowledge that can offer a profound understanding of the characteristics of the "AE journey." Therefore, the objective of this chapter is to comprehensively analyze the Event data with the following aims to identify novel distributional information, relationships, and patterns.

4.2.2 Methods

The descriptive analysis conducted in this study encompasses a range of techniques, including frequency tables, sequential pattern mining, and association rules. For detailed information on the SPADE algorithm [156] used for sequential pattern mining, please refer to section 4.1.2.3. The study population for this analysis is identical to the population discussed in the "Timelines of Adverse Event" section of this chapter. To review the inclusion criteria for the 86,912 recorded AEs of 15,820 patients, please see section 4.1.2.1.

4.2.3 Results

INTERMACS defines 25 different types of AEs, of which six are classified as *final outcomes*, indicating a wide diversity. The characteristics of patterns of the AE journey of these

[2]The majority of this section is taken from our published work: Movahedi, Faezeh, Francis D. Pagani, and James F. Antaki. "In Search of Similarity in Adverse Events Journeys of Left Ventricular Assist Device Patients." *The Journal of Thoracic and Cardiovascular Surgery* (2023).

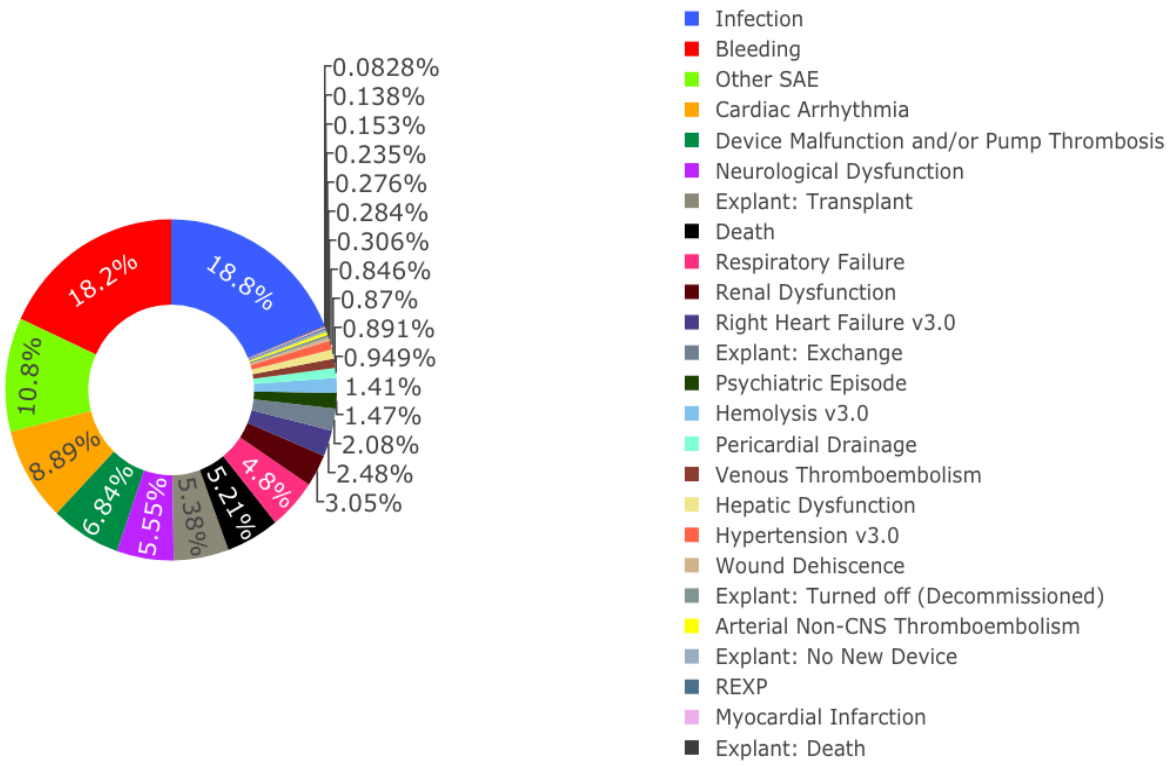
data were investigated by posing five descriptive research questions as follows:

4.2.3.1 What Are The Most Common Types of AEs?

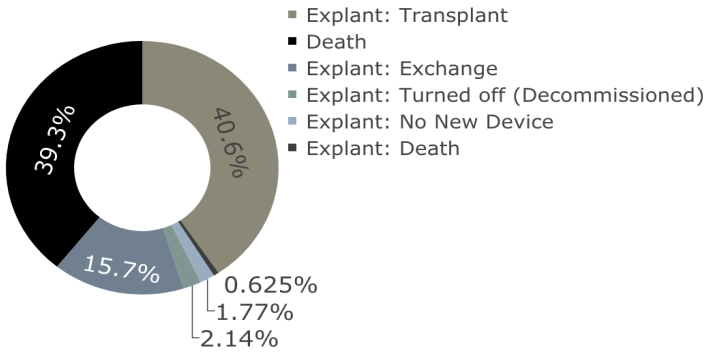
The pie chart of Fig. 14a presents the percentage of each type of AE from the total 86,912 recorded AEs. For the visualization of the results, each type of AE were color coded as shown in the legend. The four most common AEs are Infection (19%), Bleeding (18%), Other SAE (11%) and Cardiac Arrhythmia(9%) comprise nearly half (57%) of all recorded AEs. The most common *final outcomes* are Explant: Transplant and Death, each slightly more than 5% of the total recorded AEs and 40.6% and 39.3% of the total recorded final outcomes (see Fig. 14b). Eleven of the AEs occur less than 1% in the INTERMACS database (see Fig. 14a). Overall, it is clear that the majority of AEs are uncommon, indicative of a sparse data set.

4.2.3.2 What Are The Most Common Concomitant Post-LVAD AEs?

Post-LVAD AEs do not always occur in isolation, but concomitant with other AEs. The goal is to find out what types of AE frequently occurred together (at the exact same time) following LVAD, and specifically AEs coupled with final outcomes such as Death and Explant. Out of 76,590 recorded times of occurrences of AEs, only 11% (8,638) recorded an occurrence of more than one AE at the exact same time (concomitant AEs), corresponding to 5,768 (36%) patients. Collections of concomitant AEs per recorded time of AEs were assigned to Baskets (See Fig. 15a). Most baskets of AEs (85%) included only two concomitant AEs. The maximum number of concomitant AEs for one record of AEs' time was 9 (See Fig. 15b). The most frequent types of AEs that were found in baskets of concomitant AEs were: Other SAE, Infection, and Bleeding. Thus, the most common combinations of these AEs are shown in Fig. 15c including: $\{Bleeding, Other\ SAE\}$, $\{Infection, Other\ SAE\}$, and $\{Bleeding, Infection\}$ with support values of 9%, 8%, and 5%, respectively. The most common types of AEs concomitant with a final outcome were $\{Death, OtherSAE\}$, $\{Device\ Malfunction\ and/or\ Pump\ Thrombosis, Explant : Exchange\}$, and $\{Death, Neurological\ Dysfunction\}$ with support values of 3%, 3%, and



(a) Donut chart depicting the percentage of each type of AE from the total number of 86,912 recorded AEs. AEs in the legend are ordered based on their frequency.



(b) Donut-chart depicting the percentage of each type of final outcome from the total number of recorded final outcomes in the Event data set. The elements of the legend are ordered based on their frequency.

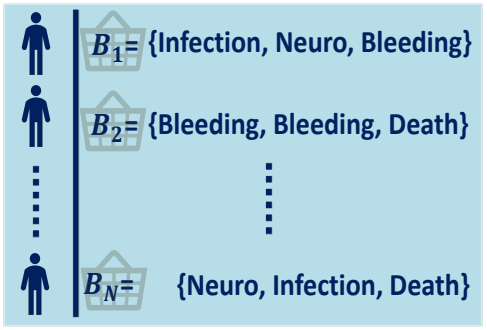
Figure 14: Distribution of different types of AEs in INTERMACS Event data.

2%, respectively.

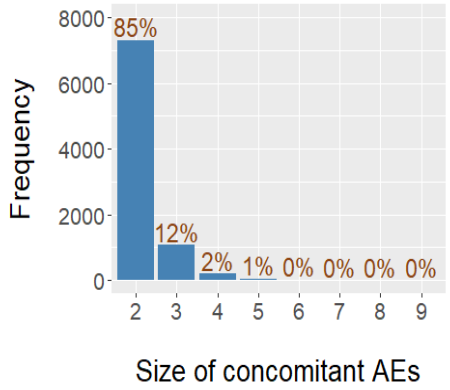
4.2.3.3 What are the most common sequences of AEs (The Exact Same Types of AE In The Exact Same Order)?

It may be interesting for clinicians to understand the common patterns of AEs that occur after LVAD. AEs can occur in various sequences, meaning that patients can experience the exact same types of AEs in the exact same order. For example, Fig. 16a displays the most frequent AE sequences in the first post-LVAD month that are common among the 1% of 9,786 patients who had at least one AE in their first post-LVAD month. In this analysis, having no AE (NAE) or freedom from AE is considered part of the "sequence". The Fig 16a displays a horizontal bar plot, in decreasing order from bottom to top. The y-axis is the cumulative frequency which shows these eleven AE sequences are related to only 20.6% of total 9,786 patients – indicating the diversity in these data. The most frequent AE sequence, common among 473 patients (4.8%), is $(NAE - Cardiac\ Arrhythmia)$. The second most common AE sequence is experiencing Bleeding during implant ($Bleeding$); 236 patients (2.4%). The rest of AE sequences include only one type of AE: Bleeding, Infection, Cardiac Arrhythmia, or Other SAE in one of the post-LVAD weeks.

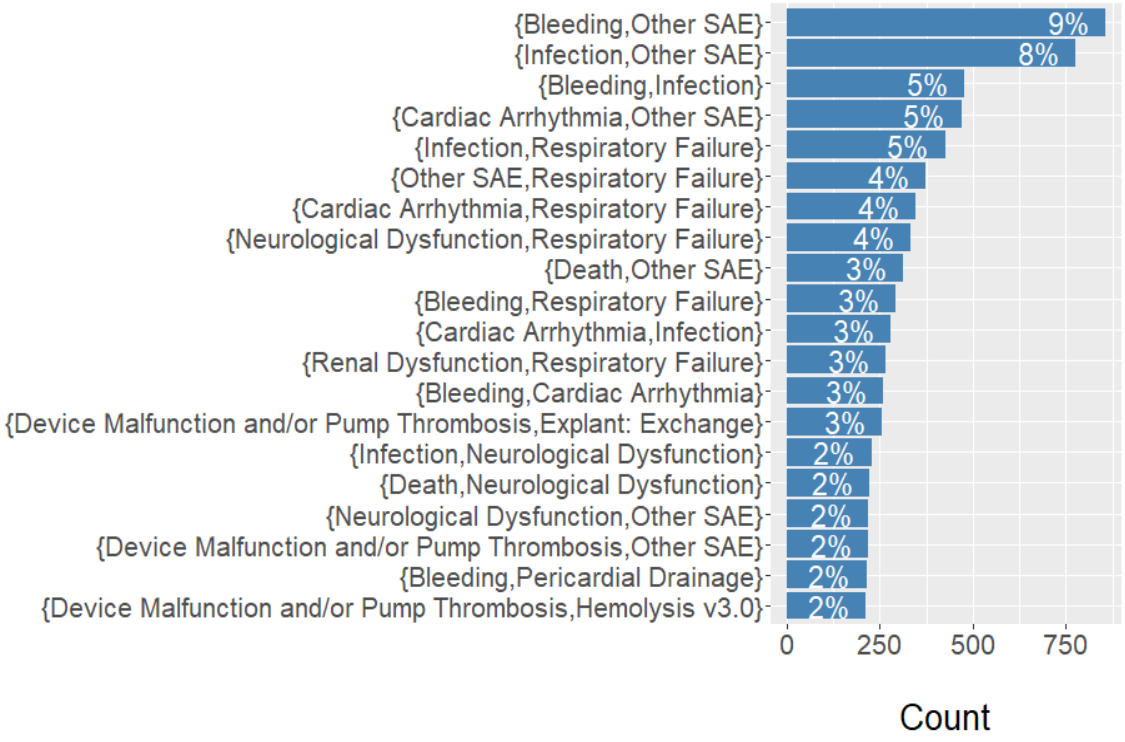
Observed AE sequences were influenced by the timescale used in the analysis. For example, a weekly timescale limits the ability to detect sequences occurring *within* one week. Increasing the time increment to year further reduces or obscures information about common sequences, owing to the sparsity of the data. This is illustrated in Fig. 16b that depicts the common patterns of AE associated with at least 1% of patients. The plot indicates that there are *only* two such patterns, both occurring in the first month: (1) NAE-Cardiac Arrhythmia, $(NAE - Cardiac\ Arrhythmia)$ is common among 218 patients (1.47%) (2) $(NAE - Death)$ is common among 158 patients (1.06%). This is despite the fact that there are 14,858 patients (out of total 15,737 patients) who experienced at least one AE during the first post-LVAD year. Due to the diversity of the data, aggregating AEs over the time interval of month vs. week causes the AE timelines (sequences) to be more dissimilar, as the chance of experiencing different types of AE over a month is greater than over the week. In



(a) Symbolic presentation of baskets of concomitant AEs identified in INTERMACS.



(b) Size of baskets of concomitant AEs



(c) The top 20 most common concomitant AEs

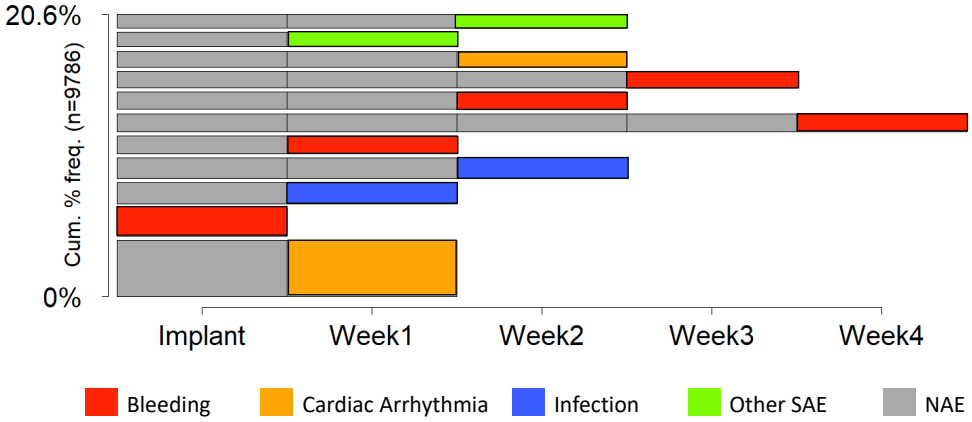
Figure 15: Common Concomitant Post-LVAD AEs.

summary, the diversity among AE timelines (sequences) exponentially increases by increasing the follow up time from 1 month to 1 year and changing the time interval from week to month.

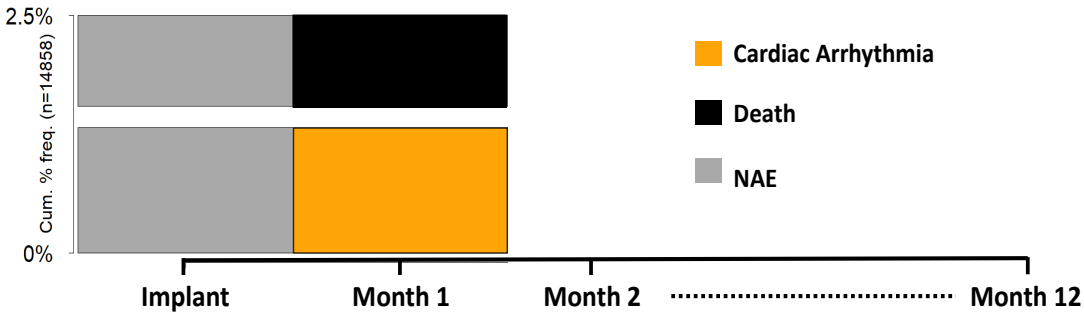
4.2.3.4 How Many Patients Share One or More AEs In The Same Order? (Common Subsequence of AEs)

Discovery of common patterns of AEs irrespective of specific week of occurrence can be elucidating. This question is related to the previous question by (1) relaxing the constraint of matching the entirely of AE sequences versus only doublets or triplets of AEs (i.e. *subsequences*) and (2) ignoring the time spans and time gaps to emphasize the type and order of AEs. For instance, if the goal is to found out how many patients experienced first Bleeding and then Infection, both of ($Bleeding - NAE - Infection$) and ($Bleeding - Infection$) will be counted for the subsequence of ($Bleeding - Infection$).

To accomplish this, the SPADE sequential pattern mining technique [156] was applied to all AE sequences to extract most common AE subsequences. Please see Appendix A for the details of SPADE algorithm. Table 3 lists the AE patterns/subsequences that are common to at least in 10% of the total 15,737 patients ordered by the count values. In this analysis, the time interval was month and there was no limit of follow-up time. The AE at the time of implant was not considered in this analysis, and thus 83 patients with only AE at the time of implant was removed from this analysis leading to total number of 15,737 patients. The first 8 common AE patterns (subsequences) are the single occurrence of Bleeding, Infection, Other SAE, etc. among patients. There were 51.3% and 41.4% of patients who had at least one occurrence of Infection and Bleeding after LVAD, respectively. Pattern mining also yielded two final outcomes, Explant: Transplant and Death, that were among the top AEs patterns with 29.7% and 28.6%, respectively. The first two common pairs of AEs (colored in green) are ($Infection - Infection$) and ($Bleeding - Bleeding$) which are common to 21.9% and 17.9% of the patients. The next sets of AEs are combinations of Infection and Bleeding in different orders; first Bleeding and then Infection ($Bleeding - Infection$) or vice versa ($Infection - Bleeding$). There are also combinations of Infection and Bleeding with the



(a) The most frequent AE sequences in the first post-LVAD month that are common among the 1% of 9,786 patients who had at least one AE in their first post-LVAD month. The AE sequences are displayed bottom-up in decreasing order of their frequencies.



(b) The most frequent AE sequences in the first post-LVAD year that are common among the 1% of 14,858 patients who had at least one AE in their first post-LVAD year. The AE sequences are displayed bottom-up in decreasing order of their frequencies.

Figure 16: Most common sequences of AE; time interval of week vs. month

final outcome of Death ($Infection - Death$) and ($Bleeding - Death$) common among 14.5% and 11.7% of patients, respectively. This analysis also revealed common pairs of AEs that occurred within a same time interval (colored in apricot) namely ($\{Infection, Bleeding\}$), ($\{Infection, Other\ SAE\}$), ($\{Infection, Respiratory\ Failure\}$). There is only one triplet in the top 10 % of AE subsequences, namely ($Infection - Infection - Infection$). Compared to the previous section, it can be appreciated that relaxing the constraints for time span, time gaps, and the complete AE sequence, resulted in only AE pattern (subsequence) that is common to at least 50% of patients. This further illustrates the influence of diversity among AE timelines after LVAD.

4.2.3.5 What Are The Interesting Relations (Association Rules) Between Various Types Of AEs?

The AE patterns (subsequences) found in the previous section are informative about the evolution of AEs, such as determining whether an occurrence of Bleeding forecasts a future occurrence of Infection. However, the counts of common AE subsequences (or percentage) in Table 3 are not sufficient to confirm a relationship (**association rule**) between occurrence of AEs. An association rule like $Bleeding \Rightarrow Infection$ indicates that Bleeding frequently preceded Infection. The **confidence** for such an association rule can be computed as the probability of the subsequent occurrence of Infection following an occurrence of Bleeding; $P(Infection|Bleeding)$. The confidence values such association rules range between 0 (0%) to 1 (100%). The rule with confidence of 1 implies that *whenever* the AE on the left occurs, the AE on the right side is guaranteed to follow 100% of the time. Table 4 lists only the first twenty association rules out of 55 with confidence of $\geq$ 50% common among at least 1% (about 182 patients) of total 15,737 patients. The first five rules with confidence values ranging from 58% to 68% are related to patients who experienced multiple Bleeding AEs with some other types of AE that leads to another Bleeding AE later. The first rule is indicates that 68% of patients who experienced 6 Bleeding AEs also experienced a 7^{th} Bleeding AE. The total number of patients who experienced this rule is 1.1% of total 15,737 patients. The remainder of rules are related to patients experiencing various combinations of Bleeding and

Table 3: The most common sequential AE patterns (subsequences) after LVAD. The time interval is month. The count value indicates the number of patients who had the AE pattern. The % indicates percent from total patients. The pairs of AEs are colored in green and the pairs of AEs that occurred within a same time interval are colored in apricot.

Rank	Common AEs Patterns	Count	%
1	(Infection)	8074	51.3%
2	(Bleeding)	6522	41.4%
3	(Other SAE)	5045	32.1%
4	(Cardiac Arrhythmia)	4830	30.7%
5	(Explant: Transplant)	4674	29.7%
6	(Death)	4501	28.6%
7	(Device Malfunction and or Pump Thrombosis)	4282	27.2%
8	(Neurological Dysfunction)	3656	23.2%
9	(Infection-Infection)	3452	21.9%
10	(Respiratory Failure)	3168	20.1%
11	(Bleeding-Bleeding)	2811	17.9%
12	(Bleeding-Infection)	2552	16.2%
13	(Infection-Bleeding)	2376	15.1%
14	(Infection-Death)	2279	14.5%
15	(Renal Dysfunction)	2221	14.1%
16	({Infection, Bleeding})	2087	13.3%
17	(Other SAE-Infection)	1934	12.3%
18	(Bleeding-Death)	1847	11.7%
19	(Cardiac Arrhythmia-Infection)	1820	11.6%
20	(Infection-Explant: Transplant)	1802	11.5%
21	(Explant: Exchange)	1795	11.4%
22	(Right Heart Failure v30)	1730	11%
23	({Infection, Other SAE})	1727	11%
24	(Infection-Other SAE)	1725	11%
25	(Infection-Device Malfunction and or Pump Thrombosis)	1705	10.8%
26	({Infection, Respiratory Failure})	1612	10.2%
27	(Infection-Infection-Infection)	1608	10.2%

Infection in different months or during the same month that lead to another Bleeding AE or Infection AE. It should be mentioned that decreasing % threshold below 1% will result in association rules with even greater confidence values (more than 90%). However, doing so would not provide much practical information clinically.

Whereas the association rules in Table 4 are relatively long, the continuation of the list of rules with $\leq 50\%$ confidence include shorter rules like $Infection \Rightarrow Infection$ that are less common - i.e. experienced by fewer patients, indicative of the diversity among AE patterns. For example, Table 5 provides a list of association rules, common among at least 1% of the total patients and having only one Infection AE on the left side of the rule. The rule with the greatest confidence is 42% but associated with a greater percentage of total patients compared to the rules in Table 4 with greater confidence. Despite the size of this cohort 15,737 patients it is seen that the confidence values for top association rules does not reach to even 70% indicating diversity among patients' AE patterns.

4.2.4 Discussion

This study provides potentially valuable insights about characteristics of "AE journeys" following LVAD implant that were extracted from the longitudinal data of $> 15,000$ patients by posing five questions. Although these analyses were performed on the publicly available INTERMACS database, which does include the most contemporary technology, this study nevertheless exhibits the great opportunity to apply advanced analysis to the AE database, as well as some related challenges.

The publicly accessible INTERMACS database comprises 25 types of AEs, however four common types encompass more than half of the recorded AEs. (See question 1; QES.1.) An obvious improvement in INTERMACS registry would be to provide more granular information on "Other SAE" which is one of the most common recorded AEs in INTERMACS. Although concomitant AEs are rare, the most common were combinations of Bleeding, Infection, and Other SAE (QES.2). This analysis revealed that patients rarely experience identical sequences of AEs as there were only two sequences shared by greater than 1% of patients in first year post-LVAD implant (QES.3). Similarly, the number of patients who

Table 4: The first twenty association rules out of 55 with confidence of $\geq 50\%$ common among at least 1% (about 182 patients) of total 15,737 patients. The time interval is month. The confidence for an association rule is the probability of the subsequent occurrence of right-hand side of the rule following an occurrence of left-hand side of the rule. The confidence value ranges between 0 to 1.

Rank	Rule	Confidence	Per
1	(Bleeding-Bleeding-Bleeding-Bleeding-Bleeding-Bleeding) ⇒ (Bleeding)	68%	1.1%
2	(Bleeding-Bleeding-Bleeding-Bleeding-Bleeding) ⇒ (Bleeding)	62%	1.7%
3	(Other SAE-Bleeding-Bleeding-Bleeding) ⇒ (Bleeding)	61%	1.3%
4	(Infection-Bleeding-Bleeding-Bleeding-Bleeding) ⇒ (Bleeding)	60%	1%
5	(Cardiac Arrhythmia-Bleeding-Bleeding-Bleeding) ⇒ (Bleeding)	58%	1%
6	({Infection, Bleeding}-{Infection, Bleeding}) ⇒ (Infection)	57%	1.2%
7	(Bleeding-Bleeding-Bleeding-Bleeding) ⇒ (Bleeding)	57%	2.7%
8	({Infection, Other SAE}-Infection-Infection) ⇒ (Infection)	56%	1.3%
9	({Other SAE, Bleeding}-Bleeding-Bleeding) ⇒ (Bleeding)	55%	1.1%
10	(Bleeding-Other SAE-Infection-Infection) ⇒ (Infection)	54%	1%
11	(Bleeding-Infection-Infection, Bleeding) ⇒ (Infection)	54%	1.1%
12	(Other SAE-Bleeding-Infection-Infection) ⇒ (Infection)	54%	1%
13	({Infection, Bleeding}-Infection-Infection) ⇒ (Infection)	54%	1.7%
14	({Infection, Bleeding}-Bleeding-Bleeding) ⇒ (Bleeding)	54%	1.5%
15	(Infection-Bleeding-Infection-Infection) ⇒ (Infection)	53%	1.7%
16	(Infection-{Infection, Bleeding}-Infection) ⇒ (Infection)	53%	1.3%
17	(Infection-Other SAE-Infection-Infection) ⇒ (Infection)	53%	1.2%
18	({Infection, Respiratory Failure}-Infection-Infection) ⇒ (Infection)	53%	1.1%
19	(Respiratory Failure-Infection-Bleeding) ⇒ (Infection)	53%	1.3%
20	(Other SAE-Infection-Infection-Infection) ⇒ (Infection)	53%	1.4%

Table 5: The association rules that are common among at least 1% of the total patients and having only one Infection AE on the left side of the rule.

Rank	Rule	Confidence	Per
1	(Infection) ⇒ (Infection)	42.8%	21.9%
2	(Infection) ⇒ (Bleeding)	29.4%	15.1%
3	(Infection) ⇒ (Death)	28.2%	14.5%
4	(Infection) ⇒ (Explant: Transplant)	22.3%	11.5%
5	(Infection) ⇒ (Other SAE)	21.4%	11%
6	(Infection) ⇒ (Device Malfunction and or Pump Thrombosis)	21.1%	10.8%
7	(Infection) ⇒ (Neurological Dysfunction)	17.3%	8.9%
8	(Infection) ⇒ (Cardiac Arrhythmia)	14.6%	7.5%

shared only part of their AEs' journey (subsequences) was minimal compared to the huge number of patients in this study (QES.4). The most common AE subsequence was shared by only 50% of patients. Consequently, the maximum confidence values for interesting associations between occurrences of AEs were below 70% (QES.5). All five questions revealed great diversity and sparsity in INTERMACS Event data with respect to type and timing and order of AEs impairing the ability to identify trends despite the large size of this dataset.

Although QES.4 and QES.5 revealed the rarity of identical AE patterns among patients, it is important to note that similar patterns could still be identified through a different approach. One possible strategy is to cluster patients into subgroups that share similar trends and timelines of AEs. By grouping patients with comparable AE patterns, meaningful patterns and associations can be extracted within each cluster. This clustering process enables researchers and clinicians to gain a deeper understanding of the underlying factors and potential associations that drive specific patterns of AEs.

Considering the goal of this study was to provide a big picture of INTERMACS Event data, it was necessary to manage to wide diversity in risk factors such as baseline patients characteristics, post-LVAD patient management, and device type. For instance, it is likely

that the AE journeys will look very different at high volume versus low volume centers [158, 159]. Therefore, it would be insightful to explore AE patterns within subsets of the INTERMACS Event data. This may lead to a more definitive, and practical guidelines for improved device design as well as patient management. In addition, this study ignored the sub-types of AEs, such as systemic infections vs device-associated infections, to control an already wide diversity in the data. Future studies may, likewise, segregate the data by sub-types of major types of AEs, such as infection and bleeding, and ignore some less common AEs, "AE noise". Lastly, the publicly available INTERMACS data (2017 version) includes a small percent of patients with new LVAD technology. Although the new technology still suffers from the most types of AEs like infection and bleeding but in lower rates, it would be valuable to mine the contemporary databases such as MOMENTUM3 for "AE journeys" related to the currently dominant device [18].

4.2.5 Conclusion

The findings of this study shed light on the diverse nature of "AE journeys" in LVAD patients, highlighting the variation in AE types, combinations, and sequences, thereby contributing to a deeper understanding of the complexities in AE management post-LVAD implantation. The findings indicate that among the 25 types of AEs in the INTERMACS database, four common types (infection, bleeding, cardiac arrhythmia, and other serious AE) account for more than half of the recorded AEs. The analysis reveals that concomitant AEs are rare, with the most common combinations involving bleeding, infection, and other serious AE. Additionally, patients rarely experience identical sequences of AEs, with only a small percentage sharing specific AE sequences in the first year post-LVAD implant. The study also highlights the minimal occurrence of shared subsequences among patients, despite the large sample size.

4.3 Guidelines for INTERMACS Event Database Investigation

In the process of addressing the eleven questions in the sections titled "Timelines of Adverse Event" (section 4.1) and "In Search of Similarity in Adverse Events" (section 4.2), a set of guidelines for investigating the INTERMACS Event database was compiled to emphasize the key factors that influence the interrogation of Event data.

4.3.1 Considerations regarding missingness and time gaps

When mining the INTERMACS Event database for sequences of AEs, missing data translates to a time gap (see "What Are The Most Common Time Gaps Between AEs?" in section 4.1.3.5) that can be classified as one of three main varieties. (1) Left gap: the absence of AEs from the time of LVAD implant (up to the the first AE) is a meaningful missingness as it reflects AE-free survival (**NAE**). Thus, the missing data could be replaced by the "NAE" in patients' AE sequences. (2) Between gap: absence of AEs two consecutive episodes of AEs indicates the duration of freedom from AE and can be filled by "NAE". (3) Right gap: The missingness after last AE up to the follow-up time depends on the last AE can differ (3.1) If the last AE is a final outcome like "Explant" or "Death" then the missingness after the final AE is a real missingness and should be left as a gap. (3.2) If the last AE is not a final outcome, then the missingness can be replaced with "NAE" as it reflects patient after the last AE is alive up to last follow-up time point. In general, the decision about the time gaps in patients' sequences is subjective to the goal and design of a study. For instance, pattern mining studies may ignore the one or more types of time gaps were explained above (1) when their emphasise is on the order of AEs' occurrences rather the timing of AEs (2) when dealing with huge number of time gaps due to the choice of time granularity which will be explained in the next part.

4.3.2 Considerations regarding granularity of time

The post-LVAD time can be segmented at different levels of time granularity; daily interval vs. weekly interval vs. monthly interval The choice of time interval for segmenting

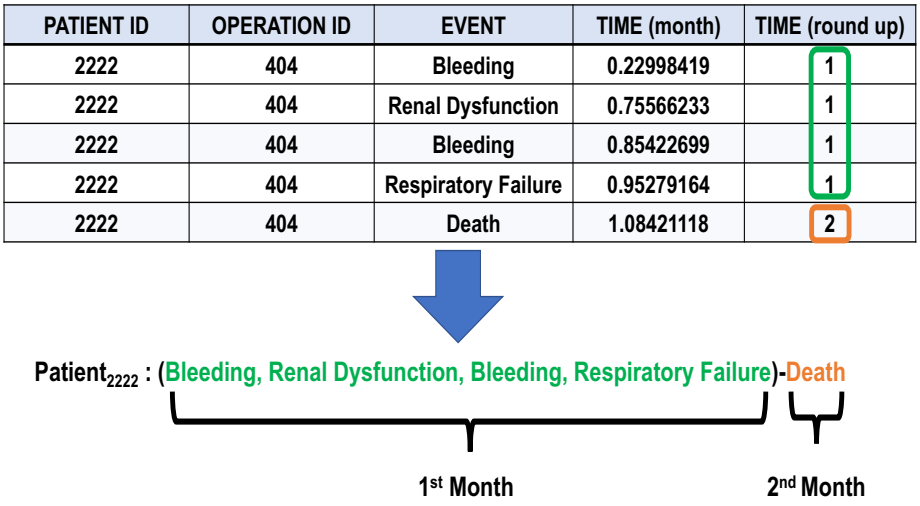


Figure 17: An example of converting INTERMACS Event data for a patient into a sequence of chronological ordered AEs.

the AE records (daily vs weekly vs monthly, etc.) will affect the perception of post-LVAD AE patterns. (See "What are the most common sequences of AEs?" in section 4.2.3.3). Time stamps in INTERMACS are provided as decimal numbers, measured in months. For example, Fig. 47 shows an portion of the Event data for a patient (P_{2222}) that includes recorded AEs spanning the first 1.08 months with five entries. By choosing a monthly granularity, these five entries become two episodes, or elements of their AE sequence: four entries in the first month, and one in the second. Fig. 18 contrasts the effect of different time granularity for the same patient: month vs. week vs. day. The corresponding AE sequences are provided in Fig. 19 which appear very different from one another in terms of length, max counts of events per element. Increasing granularity of time interval from month to Week decreases the number of AEs that are combined in each element and increases the sequence's length from 2 to 5 and introduces two time gaps ("NAE"). Further refining the time granularity to Day extends the length of the sequence to 33 including 28 time gaps ("NAE"). It is apparent that changing the time interval can vastly alter the sequence for the same patient.

TIME Month	TIME Month	TIME Week	TIME Day
0.22998419	1	1	7
0.75566233	1	4	23
0.85422699	1	4	26
0.95279164	1	4	29
1.08421118	2	5	33

Figure 18: Different time granularity for P_{2222}'s timing of AEs.

Time Unit	Sequence	Sequence Length	Max count of events per element
Month	(Bleeding, Renal Dysfunction, Bleeding, Respiratory Failure)-Death	2	4
Week	Bleeding-NAE-NAE-(Renal Dysfunction, Bleeding, Respiratory Failure)-Death	5	3
Day	NAE-NAE-NAE-NAE-NAE-NAE-Bleeding-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-NAE-Renal Dysfunction-NAE-NAE-Bleeding-NAE-NAE-Respiratory Failure-NAE-NAE-NAE-Death	33	1

Figure 19: Corresponding sequences for P_{2222} with respect to different time intervals.

Now, extrapolating this example to all 15,819 patients, it can be appreciated that their AE sequences can become more similar or dissimilar by changing the time granularity, thus effect statistical analysis like pattern mining. Characteristics of the AE sequences, such as the length and the number of AEs aggregated in each element, will also affect the complexity of analysis and computational cost.

There are tradeoffs to be considered when choosing the time interval. For INTERMACS data with 25 types of AEs, aggregating a greater quantity of AEs by longer time interval (e.g. days to months) can make two patients easily dissimilar by having only one AE in one of the aggregated AEs element in their sequences. On the other hand, by choosing a shorter time interval (day), AE sequences become lengthy, and thus slightly different timing of AEs occurrences, like 1 day, can increase the dissimilarity between patients.

In summary, the choice of granularity of post-LVAD time should be tailored to the context of analysis, the goal of the study, and/or the required sensitivity or specificity. It may be elucidating to consider a variety of time granularity to explore the data from different perspectives. (See "What are the most common sequences of AEs?" in section 4.2.3.3).

5.0 Pre-LVAD Risk Models

5.1 Pre-LVAD Mortality Risk Model Highlighting the Imbalance of Classes[1]

5.1.1 Objective

This study develops a random forest classifier to predict 90-day mortality after LVAD implantation, with the aim of achieving better performance compared to previous risk scores. The classifier is trained using data from the national registry of INTERMACS, and several approaches are employed to reduce the chance of overfitting. The performance of the random forest classifier is compared to a well-known 90-day mortality risk score in the LVAD field, referred to as the HeartMate Risk Score (HMRS). Furthermore, this study addresses the issue of class imbalance that arises during the development of LVAD-related risk scores, emphasizing how the commonly used evaluation metric, ROC, can lead to an overly optimistic assessment of risk scores. To mitigate this issue, the study suggests the use of a supplementary evaluation metric called the precision-recall curve (PRC), which considers the imbalance problem when assessing risk scores.

5.1.2 Methods and Background

5.1.2.1 Comparison of Two Classifiers for 90-day Mortality

This study compares the performance of two classifiers for predicting 90-day mortality after LVAD implantation; the well-known HeartMate Risk Score (HMRS) and a Random Forest (RF) that was derived de novo from a large multi-center registry data. The HMRS, a logistic regression, was derived from and validated within 1,122 patients red with %13 90-day mortality who received a HeartMate II as a bridge to transplant or destination therapy

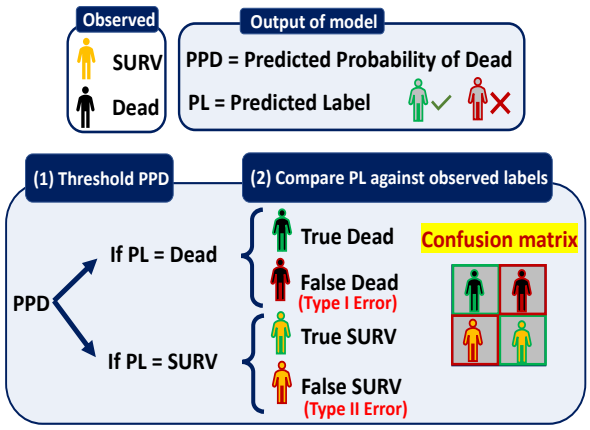
[1]The majority of this section is taken from our published work: Movahedi, Faezeh, Rema Padman, and James F. Antaki. "Limitations of receiver operating characteristic curve on imbalanced data: assist device mortality risk scores." The Journal of Thoracic and Cardiovascular Surgery (2021).

and computes the 90-day risk scores for mortality based on five variables [51]. The RF is a popular ensemble algorithm constructed by combining multiple decision trees based on "bootstrap" samples from data with random feature selection [160]. Each tree in RF will have a "vote" for a patient outcome, then the overall classifier is determined by majority of votes of the trees. For this study, a RF was derived based on 235 pre-LVAD clinical variables, such as lab values, demographic information, clinical history, etc., from 11,967 patients with advanced heart failure who received a continuous-flow LVAD recorded in the Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS). The data were randomly divided into a training (70%) and a test (30%) set. The HMRS score was computed for a subset of the test data set, censoring patients who received a heart transplant or had total recovery before 90 days, and for whom the data records did not contain all five variables required to compute HMRS. The resulting data set for computing HMRS included 800 patients (mean age of 59 years; 146 females vs. 654 males). The majority of these 800 survived to 90 days (SURV class) 92% and only 8% of patients were dead at 90 days (DEAD class). Thus, there is a high imbalance between the SURV class (majority class) and DEAD class (minority class) in these data.

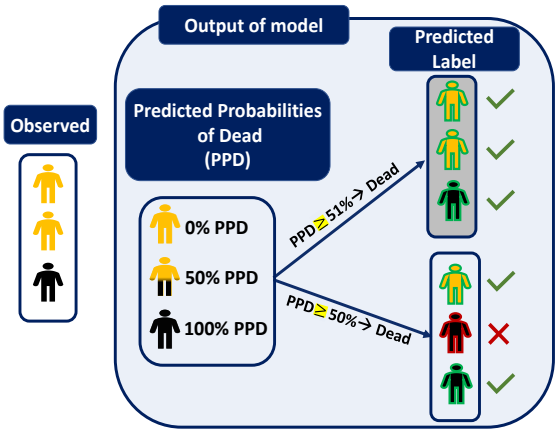
5.1.2.2 The Problem of Imbalance (and Overlap)

Without loss of generality, a classifier is a means of assigning the predicted probability of an outcome to a specific class, also known as a label. For example, if the predicted probability of a hypothetical patient being dead (PPD) is 70%, then this patient can be assigned the label "DEAD" by prescribing a cutoff value of, say 50%. By the same token, if PPD is below the cutoff, say 30%, then the patient would be assigned to the survival class "SURV." (See Fig. 20a). Then, to evaluate the performance of the classifier, the predicted label is compared to the actual outcome and summarized in the form of a confusion matrix, as shown in Fig. 20a (inset) containing four elements: True DEAD, False DEAD, True SURV, False SURV. From these four elements, several evaluation metrics can be computed, including sensitivity, precision, and specificity.

Choosing the best threshold is a challenging task that can highly affect the perception



(a) Transition of the outcome of a classifier from Predicted Probabilities of Dead (PPD) to Predicted Label (PL)- (1) Threshold the PPDs: If the PPD for a patient is greater than the threshold then the PL would be Dead otherwise PL would be SURV. (2) Compare generated PL against the observed class/label and form the confusion matrix.



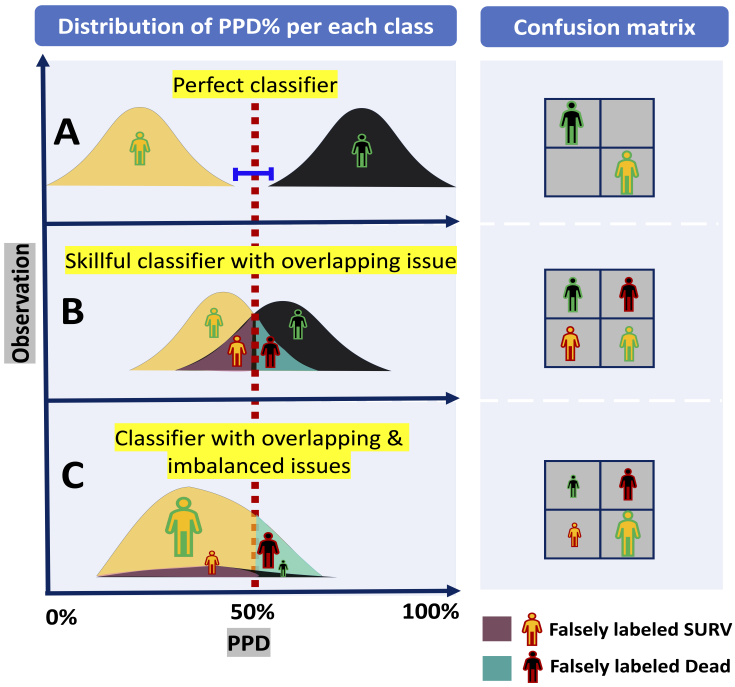
(b) An example shows the transition of the outcome of a classifier for three patients from PPDs to PLs using two slightly different thresholds- The PLs generated using the threshold of 51% (with gray background) are all correctly classified vs the threshold of 50% caused one misclassified label

Figure 20: Transition of the outcome of a classifier from predicted probabilities to predicted labels

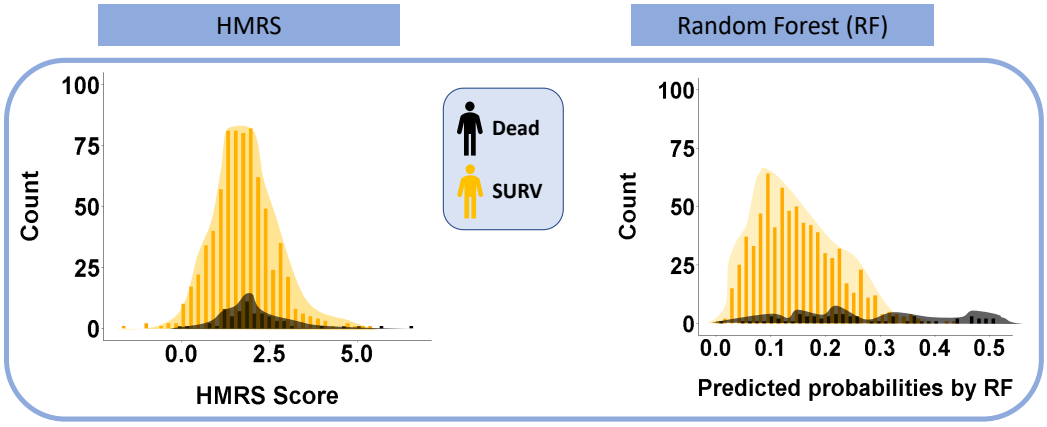
of model's performance. For example, Fig. 20b shows the predicted probabilities of being dead for three patients based on two potential thresholds of 50% and 51%. In this example the labels for the two extreme cases (PPD = 0% and 100%) are unambiguous. However, the patient in the middle with the 50% predicted probability of being dead (hence 50% of chance of being alive) can be classified with either label SURV or DEAD by merely altering the threshold by one percentage point. In this example, the performance of this classifier is achieved by assigning a threshold of 51% leading to the correct classification of all three patients. However, optimizing the threshold in real life is not as straightforward as this example.

When considering a larger population of patients, the distribution of PPD contains an ambiguous overlap in which an intermediate range of probabilities is associated with both classes. (See Fig. 21a, plot B). This is in contradistinction to the "perfect" classifier that does not contain any such overlap. (See Fig. 21a, plot A). Therefore, choosing the threshold involves a subjective trade off decision: which is worse, incorrectly predicting a patient as being dead (False DEAD, type I error), vs incorrectly predicting a patient as alive (False SURV, type II error)?

When the data are highly imbalanced, the unequal distribution of classes will compound the problem of overlap and make classification even more challenging. (See Fig. 21a, plot C). The LVAD 90-day mortality study introduced in the previous section is an example. Fig. 21b shows the histograms of HMRS Risk score (left plot) and RF probability of mortality (right plot) categorized by their actual mortality outcome (DEAD vs SURV). The predicted probabilities of being dead generated by RF for all 800 patients in this study (total of both DEAD and SURV) ranges from 0.01 to 0.52 with the mean of 0.15; while HMRS scores range between -1.80 to 6.50 with the mean of 1.71. The means of both distributions are closer to the lower part of their ranges because of the preponderance of alive patients (92%) in the test sample. This is clearly visible in Fig. 21b as the black bars (DEAD class) are much lower than the orange bars (SURV class.) HMRS has a tall bell-shaped distribution of scores around the means for both classes. On the other hand, the distribution of the SURV class in the RF histogram is right-skewed, whereas the distribution of the DEAD class is relatively flat, with no identifiable maximum. However, RF was more successful in



(a) Overlapping issue-Right Figure: A theoretical example of a classifier's output for 250 patients including: Predicted Probabilities of Dead (PPD) and Predicted Labels (PL). Left Figure: Top plot shows the outcome of a perfect classifier with no overlapping between the distributions of PPD of Dead class (colored in black) and SURV class (colored in orange). Bottom plot shows an imperfect classifier that generates PPDs with ambiguous overlap in which an intermediate range of probabilities is associated with both (either) class.



(b) Overlapping issue for the outcomes of 90-Day LVAD mortality classifiers- The underlying distributions of HMRS risk scores and RF predicted probabilities of DEAD for 800 patients in this study data are shown histogram plots. The histograms are categorized based on the observed labels for patients in this study.

Figure 21: Overlapping issue.

separating the classes for two reasons: first, the SURV class is clustered within a narrow band (approximately 0.0 to 0.3; median of approximately 10%.) Secondly, the band above 35% is dominated by the DEAD class. Nevertheless, for both of the plots in Fig. 21b, optimizing a cutoff point threshed that efficiently separate both classes is not straightforward. By default, any cutoff point threshold for both plots in Fig. 21b will result in a much greater number of True SURV and False SURV compared to True DEAD and False DEAD. On account of this dilemma (imbalance plus overlap), caution is needed when applying metrics of performance to these classifiers - such as the common ROC.

5.1.2.3 Receiver Operating Characteristic (ROC)

The ROC curve is defined as the LOCUS of True Positive Rate (***TPR***) and False Positive Rate (***FPR***) for ***all possible choices of cutoff thresholds*** as shown in Fig. 22a The color bar to the right of the ROC curve represents the threshold levels. Another term for TPR is sensitivity or recall. Another term for FPR is 1-Specificity or 1-True Negative Rate (TNR). The Y-axis (TPR) is the proportion of True Dead over all observed Dead. The x-axis (FPR) is one minus the proportion of True SURV over the total of observed SURV. In other words, ROC looks at the performance of a classifier in prediction of both of classes, Dead and SURV. The *overall* performance, for *all* threshold values, can be assessed by computing the Area Under the Curve of ROC (AUC-ROC).

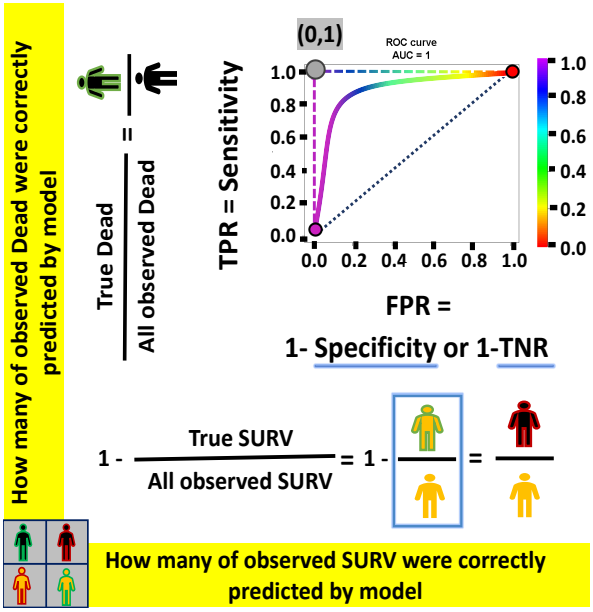
The shape of the ROC curve depends on the overlap between the distributions of predicted probabilities of the two classes. A perfect classifier with no overlapping will have an L-shape ROC curve (dashed line in Fig. 22a) with AUC-ROC equal to one, and will pass through the point of (FPR=0, TPR=1), as indicated with the gray dot, corresponding to the threshold where all patients are correctly identified by the classifier. The closer the ROC curve gets to the diagonal "line of unity" (representing random chance) such as the solid curve in Fig. 22a the worse the performance of classifier. The purple and red dots correspond to the upper and lower bounds of the threshold: > 1 and 0, respectively. At the lower bound (threshold = 0), the classifier identifies all patients as SURV (FPR= 0% or specificity=100%). At the upper bound, the classifier identifies all patients as DEAD

(sensitivity= 100%).

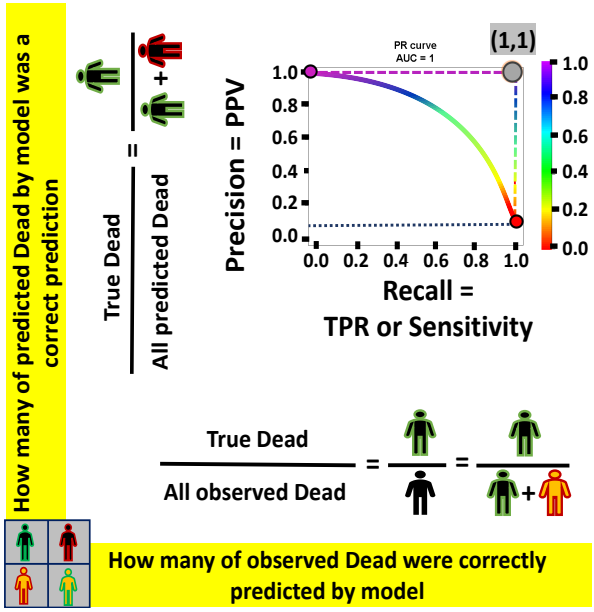
5.1.2.4 Precision-Recall Curve (PRC)

The PRC is a plot the LOCUS of ***Precision*** and ***Recall*** for ***all possible choices of cutoff threshold*** as pictured in Fig. 22b The X-axis in the PRC (recall) is the same as the Y-axis in the ROC. Other terms for recall are sensitivity and TPR, equal to True DEAD over all observed Dead. The Y-axis (precision) is also known as positive predictive value (PPV), equal to True DEAD over all predicted DEAD by the model. The color bar to the right of the PRC curve represents the threshold levels of the classifier. It is important to note that the formulas for precision and recall have the same numerator (True DEAD) and both include True DEAD in their denominator. Their only difference is one element in their denominators: False DEAD in precision and False SURV in recall. Thus, PRC focuses on the quantity of True DEAD over various cutoff thresholds considering errors of both classes: False DEAD and False SURV. Therefore, PRC is beneficial when dealing with imbalanced data as it focuses on the performance of model in only the minority class (DEAD) and it is sensitive to skewness in the imbalance data. Similar to ROC, the PRC can be summarized by computing the Area Under Curve of PRC (AUC-PRC). In PRC, the random classifier would be a horizontal line with precision equal to proportion of minority class; for instance, 8% for 90-day post-LVAD mortality.

A perfect classifier will have an L-shape PRC curve (dashed line in Fig. 22b) with the AUC-PRC of 1, and will include the point (recall= 1, precision= 1), as indicated with the gray dot in the figure, corresponding to the cutoff threshold where all patients are correctly identified by the classifier. However, when there is overlap between predicted probabilities of classes, as illustrated in Fig. 21b, the PRC curve approaches the dotted horizontal line corresponding to a random classifier. The red and purple dots in Fig. 22b corresponds to the two extreme thresholds. The red dot (threshold = 0.0) indicates the point where recall = 0, and precision =0/0, indicated by 1.0. The purple dot corresponds to the threshold 1.0 wherein the classifier identified all patients as DEAD; hence the recall = 1.0 (no False SURV) and the precision = overall proportion of the minority (DEAD) class.



(a) ROC- The example of ROC curves for a perfect classifier (L-shape dashed-curve), an imperfect classifier (solid curve), and a random classifier (diagonal dotted-line).



(b) PRC-The example of PRC curves for a perfect classifier (L-shape dashed-curve), an imperfect classifier (solid curve), and a random classifier (diagonal dotted-line).

Figure 22: Evaluation metrics: ROC and PRC.

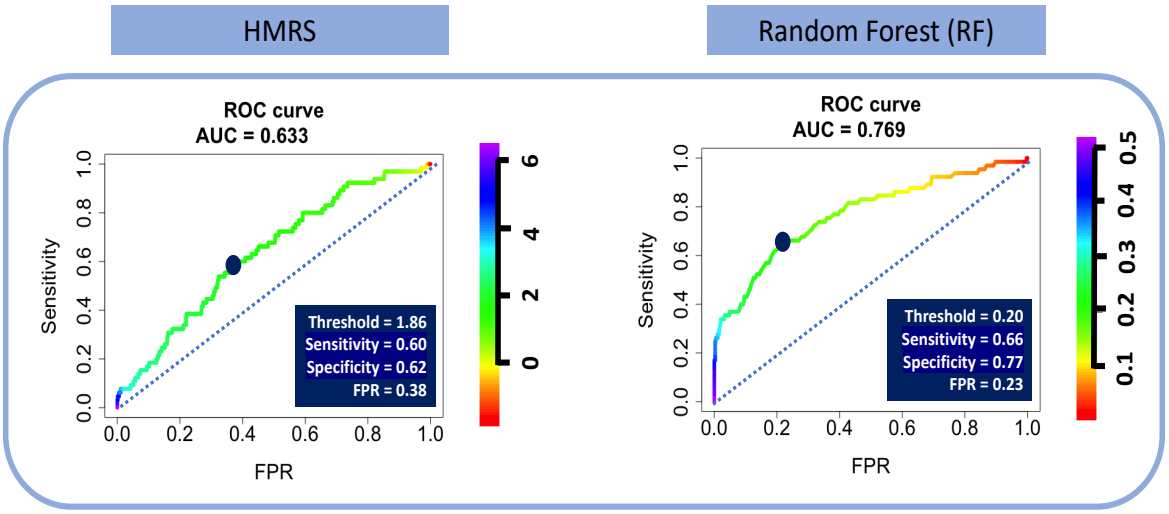
5.1.3 Result

5.1.3.1 Limitations of ROC due to imbalanced LVAD mortality

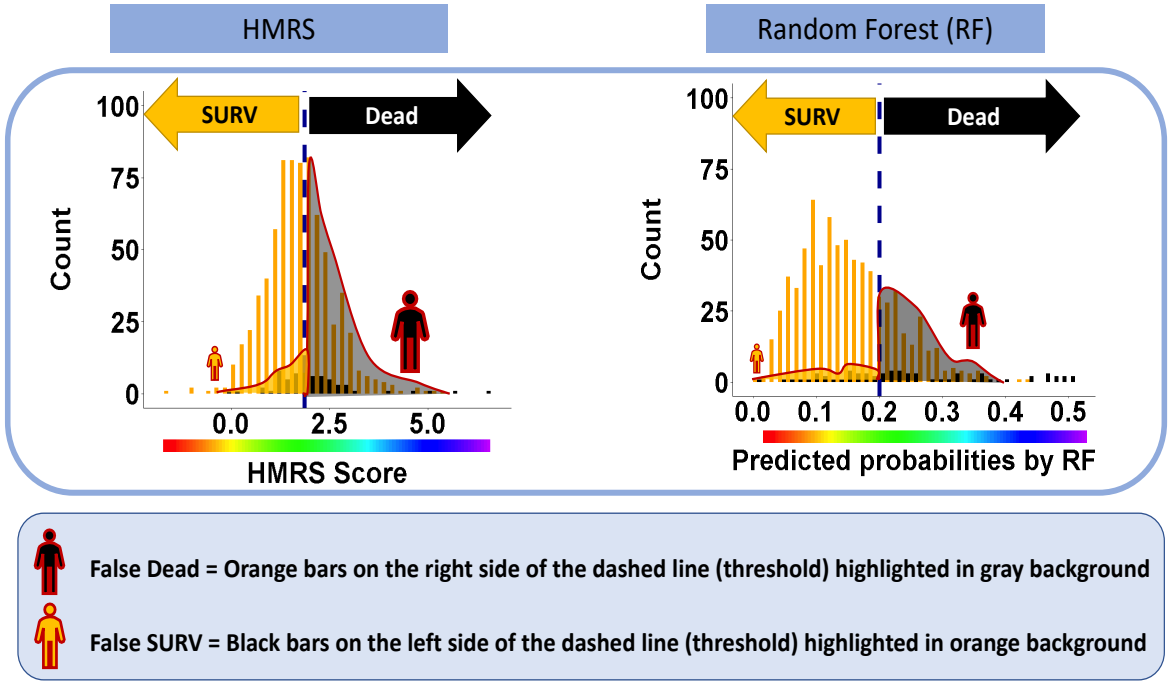
Fig. 23a shows the ROC curves for the two classifiers, HMRS and RF, for prediction of 90-day mortality after LVAD implantation. The color of the curves corresponds to the values of cutoff thresholds for each classifier shown in their corresponding legends (from 0.01 to 0.52 for RF vs -1.8 to 6.50 for HMRS). The dominant color in the ROC curve for HMRS is green corresponding to the compact (tall and narrow) distribution of scores around the mean of 1.71 as shown in Fig. 21b. Therefore, a small change in cutoff threshold above or below the mean may dramatically change the performance of classifier. On the other hand, the ROC curve for RF illustrates its performance over a more uniformly distributed range of thresholds, especially for the lower part of the range (less than 30%), corresponding to the right-skewed distribution of predicted probabilities shown in RF's histogram for SURV class (orange bars) in Fig. 21b. The area under the curve (AUC) for these two ROC curves are comparable, although RF is slightly greater (0.77) vs. 0.63 for HMRS, indicating better overall performance of RF in separating DEAD vs. SURV.

The two dark blue points on the curves indicates the optimized threshold points where the values of sensitivity and specificity are effectively equalized (1.86 and 0.21 for HMRS and RF, respectively). Although the values of sensitivity for HMRS and RF at the optimized threshold are similar, 0.60 and 0.66, respectively, the corresponding specificity of RF, 0.77, is notably greater than for HRMS, 0.62. Translating these optimized thresholds to histograms of Figure 21b illustrates the efficacy of each classifier in separating classes. (See Fig. 23b) Comparison of the two types of errors: False SURV (dead patients incorrectly classified as alive) and False DEAD (alive patients incorrectly classified as dead) reveals that the proportion of False DEAD is much greater than the proportion of False SURV for both classifiers. This is due to a combination of the imbalance of the data (about 92% alive patients) and relatively poor performance of the classifiers. However, the False DEAD is visibly larger for HMRS compared to RF due to the huge overlap between distributions of HMRS scores for DEAD and SURV classes.

The stark differences revealed by the histograms in Fig. 23b are not discernible from



(a) The ROC for HMRS and RF- The dark blue points indicate the optimal cutoff thresholds, detailed in the inset tables.



(b) The distributions of false predictions for HMRS and RF classifiers- These histograms are the same histograms in Fig. 21b. The dashed lines are corresponding to optimized cutoff thresholds chosen based on ROC curves in Fig 23a. The two types of errors, False Dead and False SURV associated with this threshold are reflected in black and orange regions with red outline, respectively.

Figure 23: Evaluation of HMRS and RF classifiers.

comparison of the corresponding ROC curves. For example, a small change of the threshold in the ROC curves in Fig. 23a corresponds to a small change in both Sensitivity and FPR. However, Fig. 23b reveals that the shifting the cutoff from the optimal point (left or right) will result in a much greater change in False DEAD vs False SURV. This is because the denominator of FPR in the ROC curve plots, the total number of SURV which is a huge number, thus attenuating the effect of changes in the numerator, False DEAD. In terms of the confusion matrix, this can be restated as number of False DEAD is overwhelmed by the much larger number of True SURV – considering that the total observed SURV is the sum of True SURV and False DEAD. (See Fig. 21a plot C). Consequently, the ROC curves in Fig. 23a does not reveal a dramatic difference in performance between RF and HMRS, Fig. 23b clearly shows that RF suffers much less from error of False DEAD than HMRS. In addition, it can be seen that choosing the threshold only based on the ROC curve may cause unintentional effects in the perception of model with respect to the minority class. In conclusion, when the ROC is dominated by the majority class (the large proportion of patients that survive), it poorly reflects the performance of the model with respect to the minority class (dead patients), and thus may be a deceptively optimistic evaluation tool in the case of imbalanced data. Therefore, there is clearly a need for a *supplemental* evaluation tool that is sensitive to skewness in the data and emphasizes the performance on the minority class. One such evaluation tool is the Precision-Recall Curve (PRC) [100, 101, 161].

5.1.3.2 Solution: PRC for imbalanced LVAD mortality

Fig. 24 shows the PRC curves for the two classifiers of HMRS and RF for prediction of 90-day mortality after LVAD implant. The color legends indicate the same thresholds values as presented in the ROC curves above (Fig. 23a). Unlike the ROC curve, the distribution of colors is more uniform, visible by the broader spectrum of colors in in the curves, especially the blue colors toward the upper range, which are virtually absent in the corresponding ROC curves in Fig. 23a. This is important because the blue portion of the curve relates to the upper bound of threshold values (high predicted probabilities of being DEAD) in which the classifiers have the greatest precision, i.e. when more of the predicted DEAD by the classifier

are True DEAD. This reveals a striking difference between HMRS and RF inasmuch as the blue region (high precision) of HMRS is limited within a very narrow band of recall, and virtually vertical. The PRC reveals that the precision drops precipitously to approximately 10% (close to the random classifier: blue dotted line) as recall (sensitivity) increases from 0% to about 10%. This corresponds to the severe overlap between the classes in the histogram of HMRS (Fig. 21b), even for the greatest scores, which leads to the huge proportion of False DEAD (alive patients incorrectly identified as DEAD). This is contrasted with the PRC of RF which decreases more gradually in precision with increasing recall. Also, it is noted that the PRC of RF remains at nearly 1.0 over a wider range of threshold, i.e. between 0.44 (44%) and 0.51 (51%) corresponding to a range of recall (sensitivity) from 0% to 17%. Overall, from the perspective of precision-recall, RF outperforms HMRS with AUC-PRC of 0.43 vs. 0.16.

The dark blue dots in Fig. 24 correspond to optimized thresholds chosen based on the ROC curves in Fig. 23a. It is readily seen that the precision of both classifiers at these thresholds is very low, although RF has a better precision, 38%, for achieving the sensitivity of 66% than HMRS with precision less than 10% for sensitivity of 60%. Using these thresholds, the HMRS classifier will correctly identify only 38 out of 64 dead patients (60% sensitivity) in the 800-patient test data set, yet will *incorrectly* label 308 patients (90% of the 342 patients labeled as DEAD) that are actually alive!

On the other hand, if we assert that precision and recall are equally important, the corresponding optimal cutoff would be indicated by the red dots on PRC curves in Fig. 24 for which both precision and recall of HMRS and RF is 15% and approximately 38%, respectively. At these optimized points, the harmonic mean of precision and recall (F1-Score) equals to both precision and recall. These optimized points are not necessarily the best way of choosing the threshold since it results in very low levels of recall; however, it illustrates that the choice of threshold highly depends on the comparative "importance" of sensitivity and precision; hence acceptance/consequence of errors (False DEAD vs False SURV).

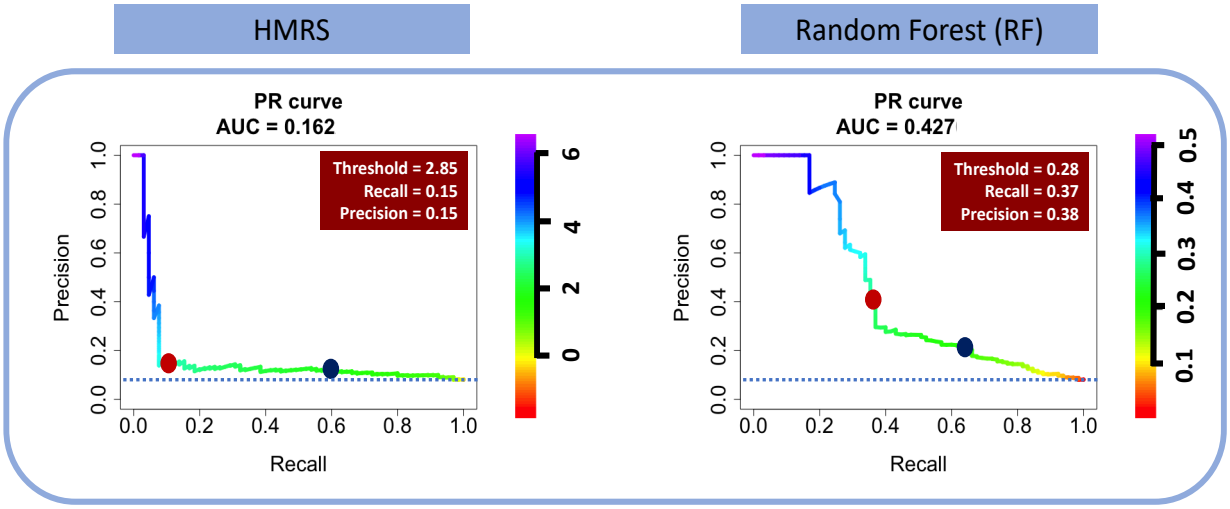


Figure 24: The PRC for HMRS and RF classifiers- The dark blue point on the PRC curves are corresponding to optimized thresholds chosen based on the ROC curves in Fig 23a The red boxes are the corresponding specifications of red dot points on PRC curves presenting the optimized cutoff thresholds of PRC curves.

5.1.4 Discussion

The clinical utility of a risk score or classifier for mortality following LVAD implantation depends greatly on the degree of separability between predicted probabilities of the two classes: DEAD vs SURV. (See Fig. 21a). Overlap between the distributions of classes creates an intermediate range of probabilities that is associated to both classes. This results in two types of errors: False DEAD (alive patients who are incorrectly labeled as DEAD) and False SURV (dead patients who are incorrectly labeled as SURV). Therefore, the choice of a threshold is tantamount to choosing between these two types of error. This dilemma is accentuated when the data are highly imbalanced, as is the case of 90-day mortality post-LVAD. (See Fig. 21b). The overwhelmingly large size of the majority class, SURV class, amplifies the False DEAD error much more than False SURV. Thus, when choosing a threshold and evaluating the performance of these classifiers, it is very important to focus on the minority class (both True DEAD and False DEAD).

This study illustrated that the ROC, a well-known evaluation tool used for most LVAD risk scores, in the case of imbalanced data, leads to an overly-optimistic perception of the performance of the classifier. This is due to the intuitive but misleading interpretation of specificity: where the large number of False DEAD error is overwhelmed by the huge number of All observed SURV in its denominator. Neglecting the full magnitude of False DEAD generated by a classifier or risk model, i.e *precision*, could give the clinician false confidence in the prediction of DEAD by the classifier. Unfortunately, most of published pre-LVAD risk scores and classifiers have not reported their precision. Therefore, these scores should be used with extreme caution.

The *Precision Recall Curve (PRC)* was shown here to be a useful tool to reveal the performance of a classifier for minority class. The PRC plots the proportion of True DEAD to both errors: False DEAD and False SURV. This is in contradistinction with the ROC which has an equal emphasis on both minority and majority classes. PRC is not affected by the overwhelming number of True SURV (majority class), and thus it does not generate a misleadingly optimistic perception performance, as does the ROC. The utility of the PRC was illustrated with two classifiers for 90-day mortality following LVAD implantation that

both suffer from imbalanced data: the well-known HMRS, and a de-novo RF classifier derived from INTERMACS, much larger data.

The preceding is not an indictment of ROC, but a revelation that ROC fails to paint a complete picture of a classifiers' performance. Therefore, ROC provides a view of classifiers' performance with both minority and majority classes while PRC provides a view of classifiers' performance on minority class which becomes more important and informative when dealing with imbalanced data.

5.1.4.1 Clinical Perspectives

Using any classifier for mortality following LVAD implantation inevitably involves choosing a threshold. From a clinical perspective this translates to a conscious decision between risk of inserting an LVAD in a patient who will die due to misplaced faith in the classifier (False SURV); versus denying a patient from a potentially life-saving LVAD because of a false presumption of death (False DEAD) by the classifier. This is an ethical dilemma. If the clinician chooses a conservative threshold, so as to avoid False SURV, he/she will mitigate the risk of accelerating a patient's death by inserting and LVAD, however he/she is at a loss for a classifier to evaluate the alternatives. This situation begs for a more holistic approach to stratification of patients with severe heart failure, to provide comparison, or ranking of alternatives, for example the use of a temporary support device as a bridge to VAD. Because VADs are one of the most expensive therapies in medicine, overly optimistic projections of survival could adversely affect cost (per quality adjusted life years, QALY), and potentially return to haunt the field in the future if costs are much higher than had been predicted.

Another consideration that highly affects the tradeoff between False SURV (False Negative) and False DEAD (False Positive) is the intended role of classifier in the clinical assessment of the pre-LVAD patients. For example, the initial screening test for HIV has a high sensitivity because of the importance of avoiding False Negative. But among those with positive initial screening test, there exists patients who do not actually have HIV (False Positive). Thus, patients with positive initial tests are reassessed with a much more precise diagnostic test with lower False Positive rate to confirm the HIV diagnosis. Therefore, as a

screening tool, sensitivity is most important (avoiding False Negative); but as a diagnostic tool, precision is more important, to avoid False Positives. By analogy to the pre-LVAD classifier, the choice of threshold might be situation-specific: more conservative as a *screening tool*, and less so as a definitive *diagnostic tool*. In conclusion, there is a need for future studies to comprehensively investigate the role of pre-LVAD risk assessment in clinical medical decisions by considering all-inclusive aspects of clinical settings of pre-LVAD.

5.1.4.2 Limitations

The problem of classifier development with imbalanced data is well-known area of research in many disciplines, including medicine [162–165], and was most recently recognized by Ishwaran in the context of cardiovascular surgery [166–168]. Accordingly, there exists a variety of approaches to mitigate the effects of imbalance such as resampling methods, assigning weights to minority samples, one-class classifier, etc. [95–97, 169]. In addition, there have been studies investigating optimization of threshold choice for imbalance data such as the quantile-classifier proposed by *Ishwaran et al.* to optimize the G-mean [170]. This study did not attempt to employ any of these methods; however, it would be beneficial in future studies to explore various strategies to achieve the best performance of LVAD classifiers. We also acknowledge that there exist other evaluation metrics, such as G-mean, PRC, and relative PRC, recently recommended by Ishwaran [166] as well as cost curve [169] and concentrated ROC [171], which were not explored in this study, but worthy of future consideration.

5.1.5 Conclusion

ROC has become an entrenched evaluation tool for assessing the performance of classifiers and risk scores in the medical arena. However, when the data is highly imbalanced, ROC can provide a misleading optimistic view of the performance of the classifiers. In such circumstances, it is imperative to employ evaluation metric such as precision-recall curve (PRC) to precisely evaluate the prediction of the minority class. This study showed this point by developing a predictive model for 90-day mortality in patients with advanced heart

failure after receiving an LVAD implant trained on the publicly accessible INTERMACS data registry. While the model exhibited a good value of 0.77 AUC-ROC, the AUC-PRC was 0.43, suggesting this highly sensitive model could be a useful screening tool for prediction of mortality at 90-day, but not a diagnostic tool because of its low precision.

5.2 Pre-LVAD Right Heart Failure Risk Model

5.2.1 Objective

The aim of this study is to explore the possibility of improving the discrimination power and generalizability of an RHF model. This will be achieved by: (1) deriving a model from big data using the publicly accessible INTERMACS national registry, which includes over 20,000 patients from more than 180 hospitals; (2) avoiding overfitting by splitting the data into training and validation datasets, utilizing repeated cross-validation during the training process, and employing a boosting-based method called XGBoost to reduce the chances of overfitting and model variance. Additionally, this paper employs more appropriate evaluation metrics that account for the data imbalance and are more sensitive to false positives.

5.2.2 Methods

5.2.2.1 Study population

A cohort of 11,967 patients was derived from INTERMACS data registry (mean age 57; 21% female and 79% male) with primary durable continuous flow-LVAD between 2008 to 2016. The database documents 245 variables including demographic details, history of medical conditions, prescription drugs, lab results, quality of life questionnaire, etc. In addition, Body Mass Index (BMI) was computed and included as an input. For patients with multiple device implants, the RHF after the first LVAD explant were excluded.

The incidence of RHF was only 9% at 1 year (1,079 patients). Thus, these data were found to be highly imbalanced.

5.2.2.2 Training predictive model

Data were randomly divided into a training (80%) and a testing (20%) sets. Then, a decision tree-based method called XGboost (Extreme Gradient Boosting) model was derived and tuned from training data using 5 repeated 10-fold cross-validations, with the data split differently for each iteration. Boosting methods are machine learning techniques that improve performance by iteratively correcting wrong predictions [155].

During construction of boosted decision trees an *importance score* is computed by the magnitude of improvement in the performance measure by the split point of the feature in a tree weighted by the number of patients in each node [172]. The overall importance score is taken as the average of all scores across all the decision trees.

5.2.2.3 Evaluation metric

The trained model was evaluated on a test data set from two perspectives: the discrimination power and calibration. The former measures how often the probability of having RHF for a randomly chosen patient with RHF is greater than that of the randomly chosen a patient who are free from RHF. The latter measures the degree to which a predicted probability of having RHF agrees with the actual probability of RHF. The ROC and Precision-Recall Curve (PRC) were reported to evaluate the discrimination power of the model by considering the imbalance issue in the dataset. (Please refer to sections of 5.1.2.3 and 5.1.2.4 for details of ROC and PRC). Specific details for calibration plot are provided below.

The calibration plot illustrates the agreement between the predicted probability and true probability of observations [173]. First, the predicted probabilities of having RHF generated by the model are discretized to bins. Then, for each bin, the mean or median of the predicted value is plotted against the true fraction of patients with RHF. If the model is well calibrated the points will be located along a diagonal line. If the points are below or above the diagonal line, it means the predicted probabilities generated by the model are too large or too small, respectively.

5.2.3 Results

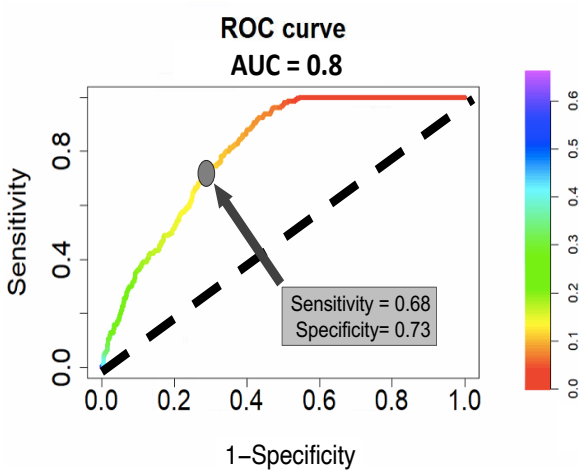
5.2.3.1 Discrimination power

Fig 25 plots the ROC and PRC for the performance of trained XGBoost on test dataset consisting of 2,392 patients (215 patients with RHF, 2,177 patients free from RHF-FreeRHF). The color-bars in both plots in Fig 25 show the range of predicted probabilities of RHF. The ROC in Fig. 25a indicates good performance of the RHF classifier with Area Under Curve (AUC) of 0.8. This is in contrast with the PRC in Fig 25b with AUC of 0.24 that shows the precision of the classifier drops rapidly from 1 (100%) to 0.4 (40%) as recall (sensitivity) increases to 0.05 (5%).

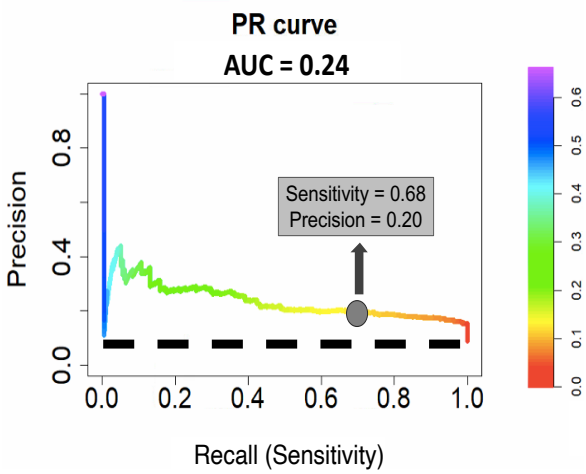
The gray dot in Fig 25a indicates the "optimized" threshold value of 0.13 for predicted probability of RHF that corresponds to the maximum value of F-score ($2 \times [(Precision \times Recall)/(Precision + Recall)]$). The F-score emphasizes the performance of the XGBoost on the minority class (patients with RHF). At this "optimized" threshold, the sensitivity and specificity are 68% and 73%, respectively. This means that XGBoost correctly classified 68% of patients with RHF (147 out of 215 patients) and 73% of patients that were free of RHF (1,597 out of 2,177 patients) in the test dataset. This is also illustrated in the confusion matrix in Fig 26 which is the table of true and false predictions for all the 2,392 patients in the test dataset. In contrast, the corresponding precision of the classifier for the same sensitivity (68%) is only 20%. (See gray point in Fig 25b). This means that only 20% of predicted RHF by this classifier is correct (147 True RHF in Fig 26). Thus, the preponderance of patients predicted to experience RHF are incorrectly classified (False RHF in Fig 26); 580 out of 727(580+147). The substantial predicted False RHF is not captured by ROC because False RHF in calculation of specificity is overwhelmed by the large number of observed patients in the denominator who are free from RHF.

5.2.3.2 Calibration plot

A calibration plot for XGBoost that relates observed risk vs predicted risk is provided in Fig 27. The plot indicates that the model is well-calibrated as the majority of the points



(a) ROC for 1-year RHF prediction. The gray box corresponds to optimized gray point in ROC where the sum of sensitivity and specificity is maximized.



(b) PRC for 1-year RHF prediction. The gray box corresponds to gray point on PRC which has the same sensitivity (recall) as the optimized point on the ROC.

Figure 25: Discrimination power of XGBoost model in prediction of 1-year RHF for patients in test dataset evaluated by ROC and PRC.

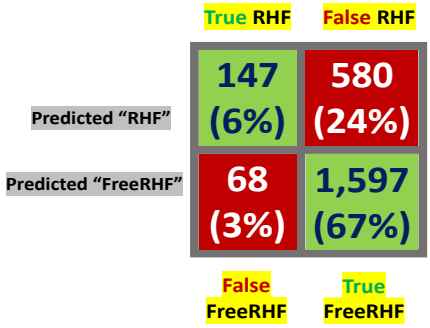


Figure 26: Confusion matrix corresponding to optimal threshold based on F-score (gray points in ROC and PRC in Fig 25.) The green elements are true predictions and the red elements are false prediction. The sum of the elements in confusion matrices is the total number of 2,392 patients in the test dataset. The percent of each element from the total number of patients is presented in parentheses.

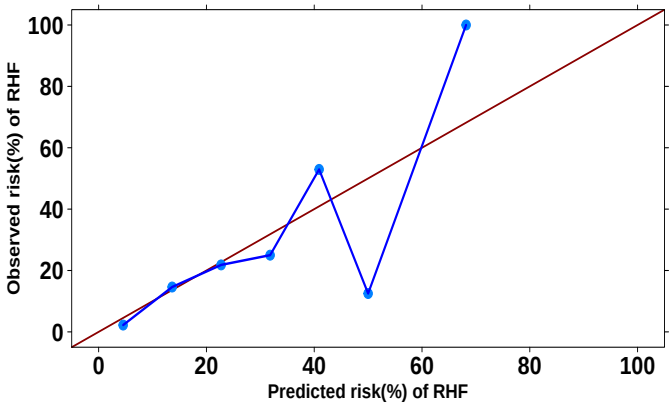


Figure 27: Calibration plot for predicted probabilities of RHF by XGBoost model for test dataset.

are closely aligned with the diagonal line except the two far-right points corresponding to predicted probabilities of having RHF more than $\geq 50\%$ for only 6 patients. The furthest-right point corresponding to only one patient that fall into the range of predicted probabilities of $(0.64\%, 0.73\%]$ is above the diagonal indicating that the observed risk is greater than the predicted risks i.e. 100% vs 68%. On the other hand, the second far-right point that includes 5 patients with predicted predicted probabilities of $(46\%, 0.55\%]$ are below the diagonal indicating that the observed risks are less than the predicted risks i.e. 12.5% vs 50%.

5.2.3.3 Feature importance

The relative importance of the top twenty pre-LVAD features are presented in Fig 28. These values range from 0-100 reflecting the relative influence of a variable in the XGBoost model. The top seven features with greatest relative importance are mostly related to secondary diagnoses. These findings are supported by previous studies such as *Ochiai et al.* that identified non-ischemic etiology (also referred to as *dilated*) as a predictor of RHF after LVAD [174]. *Wang et al.* considered valvular heart disease as a major concern at the time of surgery because of development of worsening RV failure after LVAD implantation [175]

and *Celik et al.* noted that optimal coronary artery flow may reduce postoperative adverse events [176].

To further investigate the two top features, Fig. 29 illustrates the proportions of "Secondary Diagnoses: None" and "Secondary Diagnoses: Coronary Artery Disease" associated with RHF and freedom of RHF. Both the secondary diagnosis of coronary artery disease and the *absence* of secondary diagnosis are more common among patients who are free from RHF; 29% vs 2% and 10% vs 0.3%, respectively. The results of a chi-squared test of independence is indicated in the right plots in Fig. 29. The p-values of $< 2.2\ e^{-}20$ indicate both of secondary diagnoses are significantly negatively associated with RHF (dark and medium-red circles in their residual Pearson's matrices in Fig 29). This indicates that both absence of a secondary diagnosis *and* secondary diagnoses of coronary artery disease is associated with freedom RHF.

5.2.4 Discussion

The main goal of this study was to develop a predictive model of RHF having improved discrimination power compared with existing models, while mitigating the adverse effect of overfitting – a deficit of existing models [56]. Specifically, this study considered four strategies: (1) splitting the data into training and test groups, (2) using repeated cross-validation for training the model, (3) employing big data that includes approximately 12,000 patients, (4) applying XGBoost which has been effectively used in the healthcare field for disease diagnosis and risk prediction. [177, 178]. In this study, a well-calibrated XGBoost model achieved a high AUC-ROC (C statistics) of 0.80 compared to AUC of 0.53-0.65 based on a recent external validation study for several models reported by *Frankfurter et al.* [56]. The test dataset in this study is effectively an external test dataset inasmuch as the 2,393 patients were randomly chosen from a large national registry from > 180 hospitals. This assures sufficient diversity in the profiles of patients and hospitals to account for potential confounding factors such as post-operative management.

The reported model in this study trained based on 245 variables – based on feature selection methods – gave the best performance but may not be practical for clinical usage.

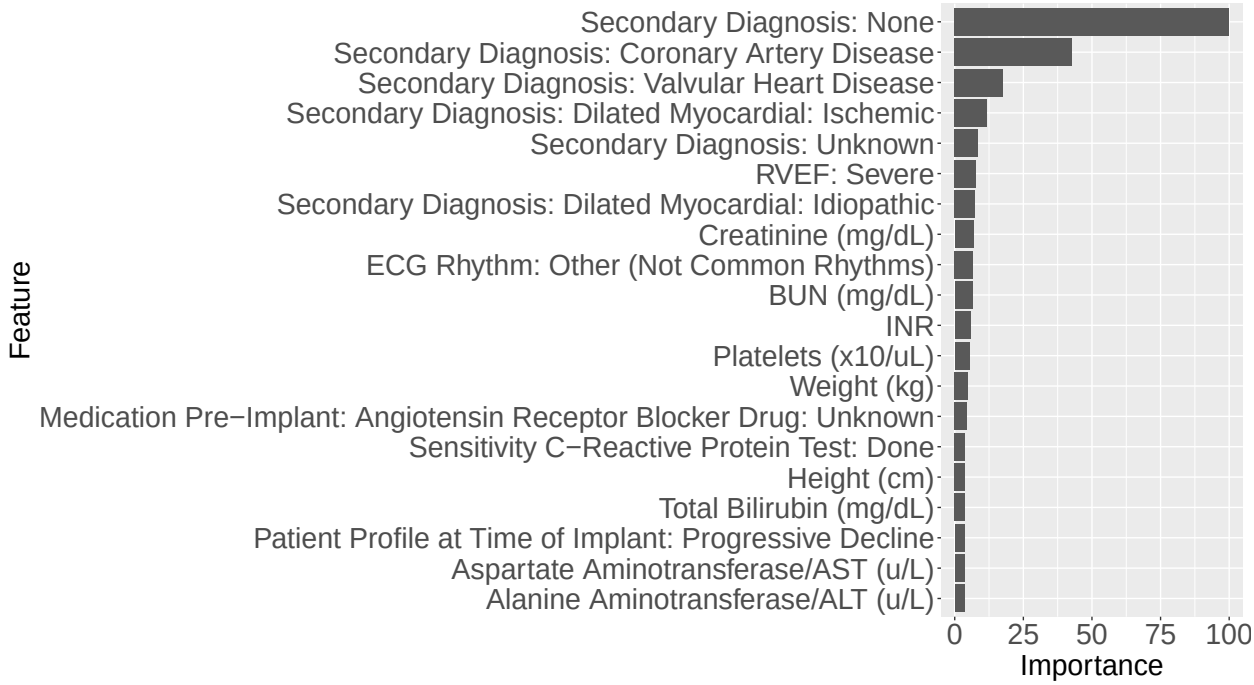
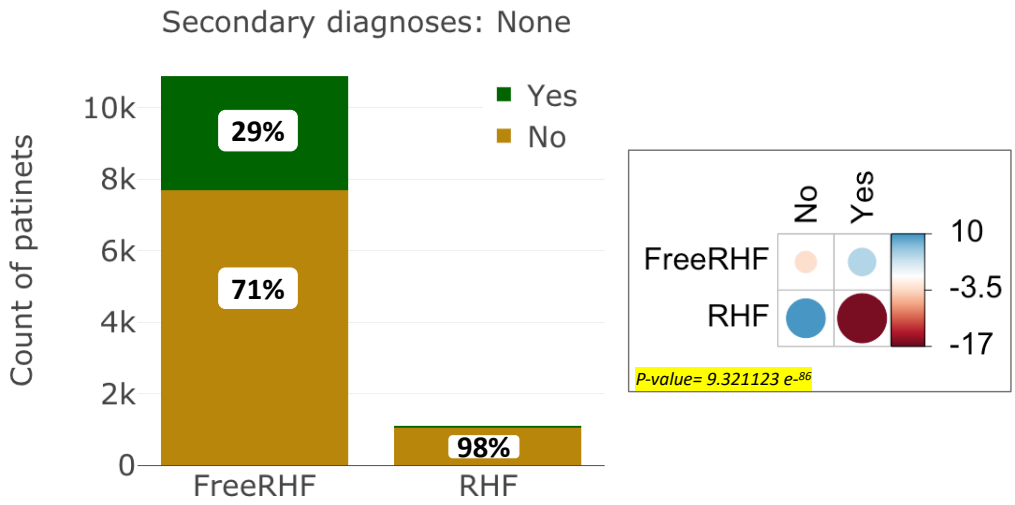
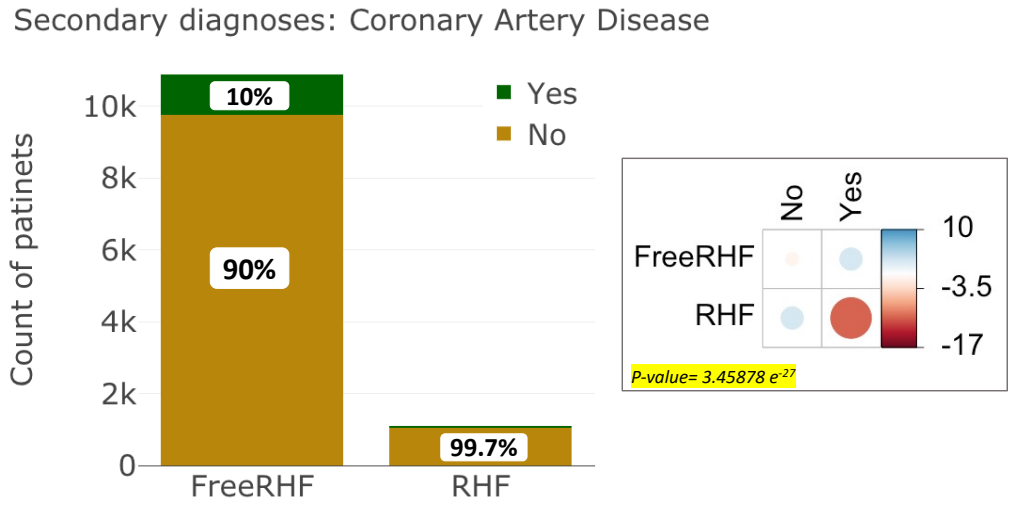


Figure 28: Plot of the top twenty relative importance of features for prediction of RHF.



(a) The associations between RHF and "Secondary Diagnoses: None". Left figure: stacked bar plot. Right figure: the Pearson residuals of Chi-square test showing the positive or negative associations.



(b) The associations between RHF and "Secondary Diagnoses: Coronary Artery Disease". Left figure: stacked bar plot. Right figure: the Pearson residuals of Chi-square test showing the positive or negative associations.

Figure 29: The association between RHF and secondary diagnoses.

Thus, a search was conducted for a model based on less features having similar performance. Various combination of top variables ranked by feature selection methods were tested. The resulting search achieved a model with 8 variables: second diagnosis, blood urea nitrogen (BUN), hemoglobin, white blood cells count (WBC), platelet count, total bilirubin (BILI), right ventricular ejection fraction (RVEF), and patient profile at time of implant. This model achieved virtually equivalent AUC-ROC of 0.8 although with decreased precision (AUC-PRC of 0.21 vs. 0.24) which was mostly affected by the greatest predicted probabilities of RHF by the model.

To the best knowledge of the authors, this paper is the first to report performance of a risk model for RHF in terms of *precision* to raise the alarm about the perils of imbalanced data. This study shows that the ROC, a common evaluation metric, can portray an overly-optimistic performance of a classifier or risk score when applied to imbalanced data – by neglecting the magnitude of false positives (false RHF) generated by a model. In this study, the PRC provided an evaluation metric that explicitly focuses on the minority class (RHF). The PRC indicates that the precision of the model is not acceptable – in contradistinction with its excellent sensitivity and specificity. The high sensitivity means that the RHF model reported here can be used as an effective *screening* tool; thus, it will identify patients with greatest risk of RHF. But the model cannot be relied upon as *diagnostic* tool as it suffers from an overwhelming false positive rate (patients without RHF that are mistakenly labeled by the model to be at risk of RHF).

Equally as important as choosing the proper evaluation metric when dealing with imbalance datasets, it is advisable to employ techniques to compensate for the imbalance [95–97]. In the current study, the investigators applied several such techniques, including re-sampling and weighting, but they were not effective in improving the outcome of the risk model. This may imply the low discrimination power of the data itself – suggesting that a solitary pre-implant time point may not be adequate for making precise predictions of the post-LVAD events, such as RHF, due to complex nature of both surgery and post-operative management.

Another confounding factor that hinders development of any RHF risk model is the heterogeneous definitions of RHF among the studies [56]. In this study, RHF is determined by the diagnostic criteria in INTERMACS (2017 version), but these criteria are different

among the clinical centers and also evolving over the time [56, 79, 179]. Another limitation of this study is that the publicly available INTERMACS registry does not fully reflect the most contemporary data and devices. Future studies would benefit greatly from access to the most current registry data.

5.2.5 Conclusion

This study developed a predictive model for 1-year RHF in patients with advanced heart failure after receiving an LVAD implant trained on the publicly accessible INTERMACS data registry. To reduce the risk of overfitting, the model employed the XGBoost algorithm. The model exhibited an AUC-ROC of 0.8 with a well-calibrated prediction of RHF. This study evaluated the discrimination power of the RHF model from the perspective of precision. These findings suggest that this highly sensitive model, based on data 48 hours prior to LVAD implant could be a useful screening tool for prediction of RHF at 1-year, but not a diagnostic tool because of its low precision.

6.0 Post-LVAD Clustering

6.1 Clustering Without Considering Timeline of Adverse Events [1]

6.1.1 Objective

The findings in Section 4.2.5, titled "In Search of Similarity in Adverse Events," indicate the diverse nature of "AE journeys" in LVAD patients. These findings highlight the variation in AE types, combinations, and sequences, which contribute to a deeper understanding of the complexities involved in AE management after LVAD implantation. Given the rarity of identical AE patterns among patients, one possible strategy is to cluster patients into subgroups that share similar trends and timelines of AEs. By grouping patients with comparable AE patterns, it is hypothesized that meaningful patterns and associations can be extracted within each cluster. Therefore, the main objective of this study is to explore and model sequential patterns of post-LVAD AEs and specifically differentiate distinct groups of patients based on their sequences of AEs.

6.1.2 Methods

6.1.2.1 Framework

The overall approach for this study is motivated by the methodology introduced by Zhang, *et al.* [180, 181] and has three main steps, as shown in Fig. 2. First, selected data from INTERMACS were transferred into sequences of AEs for each patient through multiple preprocessing tasks. Next, patients' AE sequences were clustered into groups with similar sequences using hierarchical clustering. Lastly, patterns of chains of transitions between AEs for each group were extracted using Markov Modeling.

[1] The majority of this section is taken from our published work: Movahedi, Faezeh, et al. "Sequential pattern mining of longitudinal adverse events after Left Ventricular Assist Device implant." IEEE journal of biomedical and health informatics 24.8 (2019): 2347-2358.

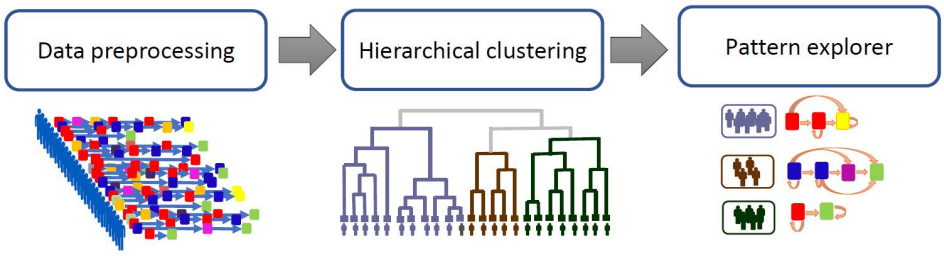


Figure 30: General work-flow: 1. Data preprocessing: Forming patients' sequences. 2. Hierarchical clustering:cluster patients into groups of patients with high similarity between patients' sequences. 3. Pattern explorer: Extracting patterns of post-LVAD sequential AEs in each group using Markov Molding [180].

6.1.2.2 Patient Selection

This study included 58,575 recorded AEs of 13,192 patients (median age of 50-59; 10,333 male vs. 2,859 female;) with advanced heart failure who received a continuous flow LVAD between 2006 to 2015, extracted from INTERMACS. For patients with multiple device implants, AEs after the first LVAD explant are excluded as patients with multiple subsequent LVAD devices are clinically treated differently.

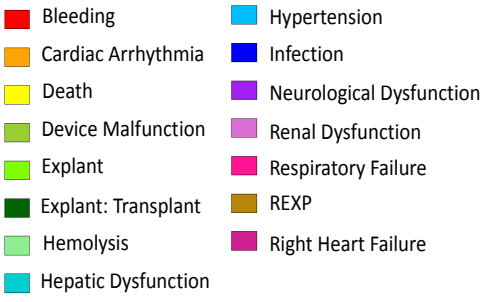


Figure 31: Color code for 15 types of AEs and final outcomes

Final outcomes, such as death, explant, and transplant, were included as the last elements in the sequences of AEs. For the subset of patients who received a right-ventricular assist

device (RVAD), the explant of that device was named to "REXP". For the visualization of the results, each type of AE and final outcomes were color coded, as shown in Fig. 31.

6.1.2.3 Forming patients' sequences

The sequence of AEs for each patient was identified and unified as a single record. A sequence for j_{th} patient (P_{j}) can be presented as follow:

$$P_j : AE_1 \rightarrow AE_2 \rightarrow ..... \rightarrow AE_i \tag{6.1}$$

Where:

$$AE_1 : \text{First adverse event}$$

$$\vdots$$

$$AE_i : i_{th} \text{ adverse event}$$

As INTERMACS provided no information related to the order of concurrent AEs (negligible percent of AEs), they were alphabetically ordered in patients' sequences.

A consequence of the large number of event types (15) is the possibility of a large number of sparse patterns. For example, if we only considered sequences of length = 3, there would be 2,940 ($14 \times 14 \times 15$; "Death" could be considered for only the last element) possible combinations. To overcome this issue, hierarchical clustering was used to divide the space of patients' sequences into more dense sub-spaces that represent patients with relatively similar sequences which would therefore be more amenable to pattern mining.

6.1.2.4 Hierarchical clustering

The goal of clustering is to identify clinically meaningful groups of patients with relatively similar sequences of post-LVAD AEs.

Defining the measure of dissimilarity:

A distance matrix, $\boldsymbol{d}$, comprised of distances between each pair of patient sequences was defined as:

$$d(P_n, P_m) = |P_n| + |P_m| - 2LCS(P_n, P_m) \tag{6.2}$$

where, $|P_n|$ and $|P_m|$ are the lengths of the sequences for patients of P_n and P_m, respectively; $LCS(P_n, P_m)$ is the Longest Common Subsequence between P_n and P_m, as formulated below:

$$LCS(P_n, P_m) = max\{|l| : l \in SB(P_n, P_m)\} \tag{6.3}$$

Here, $SB(P_n, P_m)$ is a set of all common subsequences between P_n and P_m and $|l|$ is the length of common subsequence. A subsequence is a secondary sequence derived from another (primary) sequence by deleting some or no elements while maintaining the same order of the remaining elements of the primary sequence. A *common subsequence* is a subsequence that is common to both P_n and P_m. As an example, the only common subsequence between the P_1 , P_2, shown in Fig. 32, is the subsequences of (Bleeding)-(Infection). Thus, in this example, the LCS is 2 and $d(P_1,P_2)$ is 3 which means by 3 movements (deletions of respiratory failure and death from P_1's sequence and insertion a bleeding AE to P_1's sequence) P_1's sequence becomes similar to P_2's sequence.

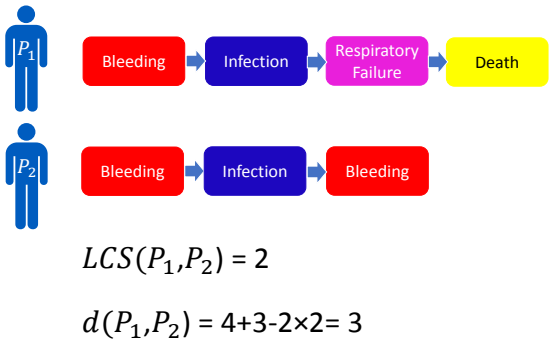


Figure 32: An example of computing dissimilarity score between two patients' sequences.

Defining linkage method for hierarchical clustering:

After forming dissimilarity matrix of AE sequences, they were clustered using bottom to top hierarchical clustering with Ward linkage. Hierarchical clustering merges sequences with lowest distance, d into a single group, and updates the distance matrix for the newly merged group and remaining patients. This merging process repeats using Ward's method in which groups of patients with lowest post-merging in-group variance (sum of squares) are merged until all patients are in one group. The Ward linkage distance between clusters is computed

using Lance–Williams recurrence algorithm [182]. Briefly, considering two clusters of C_i and C_j, the distance between new cluster C_{ij} $(= C_i \cup C_j)$ and remaining clusters such as C_k is formulated as follows [182]:

$$D_{(ij,k)} = \alpha_i D_{(ik)} + \alpha_j D_{(jk)} + \beta D_{(ij)} + \gamma |D_{(ik)} - D_{(jk)}| \tag{6.4}$$

Here, $D_{(ij,k)}$ is the distance between new cluster C_{ij} and cluster C_k. Lance–Williams coefficients α, β, and γ are different for various linkage methods and are defined for Ward linkage as follows:

$$\begin{aligned} \alpha_z &= \frac{|z| + |k|}{|i| + |j| + |k|}\ ,\ z = i, j \\ \beta &= -\frac{|k|}{|i| + |j| + |k|} \\ \gamma &= 0 \end{aligned} \tag{6.5}$$

Here, $|.|$ indicates absolute value, and i, j, and k are numbers of patients in each cluster.

Define criteria for choosing a number of clusters:

The clustering algorithm was implemented to maximize similarity between sequences within a group (internal validation), and minimize similarity between groups (external validation).

Internal validation (the within-group similarity) was performed by extracting the most common subsequences and their *support values*. This is the proportion of sequences in a group that contain that specific subsequence and ranges from 0 to 1 (0% to 100%). The most common subsequences were extracted from AE sequences by applying a prefix-tree-based search algorithm using "TraMineR" package from R [183, 184]. It should be noted that this algorithm computes the support value for a given subsequence by including all longer sequences containing that subsequence. For instance, patients with the subsequence of (Infection)-(Bleeding) are counted among the patients with the subsequence of (Infection).

External validation (the between-groups dissimilarity) was performed by identifying the subsequences that best differentiate two groups of patients' sequences using the Pearson Chi-square test (p-value of ≤ 0.01).

Step-wise evaluation of clustering: The clustering evaluation for choosing the number

of groups (n) started with evaluating the two-cluster solution (n= 2) and evaluation was continued for bigger numbers of groups until both internal and external criteria were satisfied for all the groups. At each step of clustering evaluation, n was increased by 1; only one group was divided into two new groups (G1 & G2) and their qualities (high internal similarity and low external similarity) were evaluated. First, external validation was checked between the G1 and G2. If the external criteria were satisfied, internal validations will be checked for each of the G1 and G2. If any of G1 or G2 satisfied the internal validation, it was considered as a qualified group, otherwise it was considered as an unqualified group that needed to be split into sub-groups with more similar sequences in the later steps of clustering evaluation (bigger values of n).

Interactive visualization evaluation: To help visualize the composition of AEs within each group (or sub-group), a histogram was constructed using the same color coding from Fig. 31. in which the proportions of each category of AE was plotted for each position in the sequence. Fig. 33 provides the histogram for the aggregate of all the 13,192 sequences of AEs over their chronological positions in the sequences. The first column of the graph, AE_1, presents the proportions of various types of AE that patients experienced as their first AEs (the first element of the sequences of AEs), the second column for the second AEs, and so forth, through the thirty sixth AE. The uniformity of the distribution in the first set of columns of the histogram (e.g. AE_1 through AE_{20}) is contrasted with the heterogeneity of the subsequent columns. This reflects the increasing diversity of AE's as there is a decreasing number of patients with longer sequences. For instance, there is only one patient with a sequence of 36 AEs - the last AE being death (yellow). This type of visualization is helpful to get a general quick view of the distribution of AEs in a cluster of patients.

The final clusters were evaluated by our clinical experts to determine the reasonableness of the clusters.

6.1.2.5 Markov Chain Models of AEs

Following clinical confirmation of the results of hierarchical clustering, patterns of AEs for each group (cluster) were analyzed using Markov modeling (MM). This has been shown to be

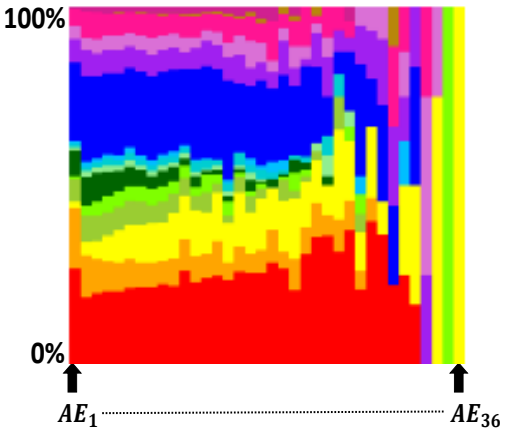


Figure 33: The proportions of various types of AE in all the 13,192 sequences of AEs in this study for temporally ordered AEs of patients sequences (see Fig. 31. for color coding).

a useful tool to model repetitive events, where the timing and the order of events are important [185]. For instance, the transition of bleeding to infection can be considered different if bleeding occurred as the 1st AE followed immediately by infection, (Bleeding1)-(Infection2), versus bleeding occurred as the 2nd AE, and then infection occurred, (Bleeding2)-(Infection3). Accordingly, the transitions between sequential AEs were assessed for likelihood of transitions as a function of the chronological position in the sequence of AEs. As this is the first attempt to model the sequential AEs after LVAD implant, it was preferred to start with a simple solid model such as the first-order Markov chain. MM was defined for a sequence of AE_1, AE_2, AE_3,... for discrete points of time: $1, ..., n$ was defined as follows:

$$\begin{aligned} &P(AE_{n+1} = ae \mid AE_1 = ae_1, AE_2 = ae_2, ..., AE_n = ae_n) \\ &= \\ &P(AE_{n+1} = ae \mid AE_n = ae_n) \end{aligned} \tag{6.6}$$

where ae was an AE from 15 different types of AE that could occur in various orders in a sequence of AEs. The above formula assumes that the probability of transitioning to the next state, AE_{t+1}, depends only on the present state, AE_t. Another assumption of MM is homogeneity, which assumes all patients in the same state have the same risk of transition.

MM considers each types of AE as unique Markov *state*, and transitions between states as *events*. As an example, a sequence of (Bleeding1)-(Infection2)-(Death3) was represented as (Bleeding1 $\rightarrow$ Infection2) $\Rightarrow$ (Infection2 $\rightarrow$ Death3) that includes 2 events presenting 2 transitions between 3 states/AEs in the sequence. It also shows the transitivity relation between 2 events as the target state in the first event, Infection2, is the source state in the second event. A transition matrix was formed by computing transition probabilities between each pair of states/AEs from all the sequences of AEs in each group. Then, chains of events were extracted from the transition matrix by connecting events that have transitivity relations. Finally, thresholding of the extracted chains was performed, based on the distributions of transition probabilities and frequency of occurrence to eliminate the "noise" of numerous infrequent or rare transitions. The thresholds were chosen subjectively as 0.1 for transition probability and between 30 to 50 for frequency of occurrences, respectively, to achieve a compromise between reducing noise versus over-simplifying the resulting collection of MMs.

6.1.3 Results

This study was performed with data from INTERMACS for 13,192 patients with advanced heart failure who underwent continuous flow LVAD implant between 2006 and 2015. A total number of 58,575 AEs, including 15 various types of AE, were included in this study. Table. 6 summarizes the 15 types AEs and final outcomes, and indicates that "Bleeding", "Infection", and "Cardiac Arrhythmia" are the most common AEs.

The time of recorded AEs ranged between 0 (at the time of implant) and 87 months (7.25 years) after LVAD implant with mean of 9.95 months. A great proportion of AEs (81% of total AEs) occurred within the first 18 months after LVAD with the peak of AEs (27%) during the first month.

6.1.3.1 Sequences of AEs

The length of AE sequences ranged from 1 to 36, however 94% of the sequences were less than or equal to 10. The distribution of lengths $\leq$15 are provided in Fig. 34. (Lengths 16-32 were very rare, <1%, and therefore not shown).

Table 6: Frequency of various types of events, including AEs and Final outcomes, and their percentages of the total number of events. The cells in the table that contains the final outcomes including "Death", "Explant", and "Explant: Transplant" are highlighted in light orange color.

Event	Frequency	Percent
Bleeding	12,877	22.0
Cardiac Arrhythmia	6,361	10.9
Device Malfunction	3,726	6.3
Hemolysis	797	1.4
Hepatic Dysfunction	850	1.4
Hypertension	701	1.2
Infection	13,399	22.9
Neurological Dysfunction	3,873	6.6
Renal Dysfunction	2,248	3.8
Respiratory Failure	3,614	6.2
REXP (RVAD explant)	92	0.1
Right Heart Failure	747	1.3
Death	3,675	6.3
Explant	1,819	3.1
Explant: Transplant	3,796	6.5
Total recorded events	58,575	100.0

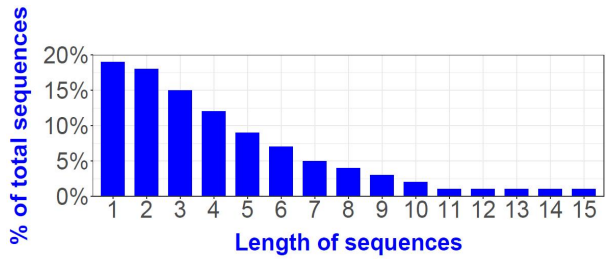


Figure 34: length of AE sequences from 1 to 15. (Length of 16-32, not shown, were rare < 1%.

6.1.3.2 Hierarchical Clustering

Fig. 35 shows the distribution of dissimilarity scores between AE sequences, which is left-skewed. The min/max of scores were 0/53 with the mean of 10 and the median of 8. The numbers of dissimilarity scores less than 2 and greater than 22 were negligible (≈0%).

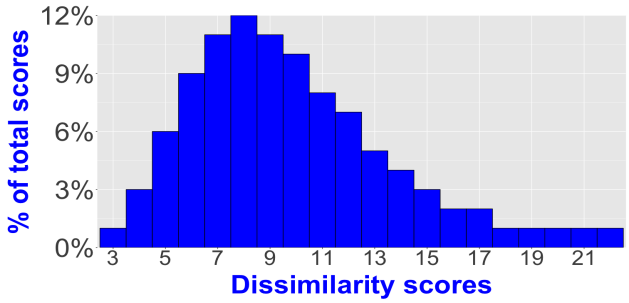
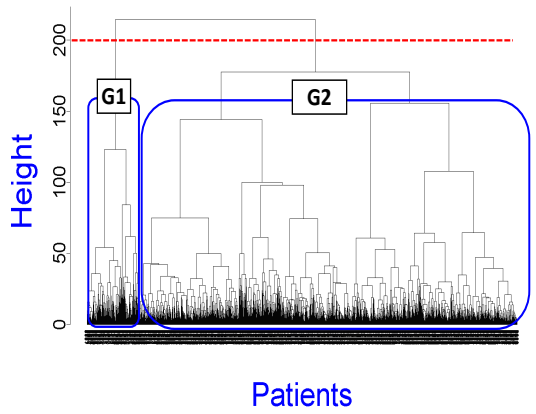
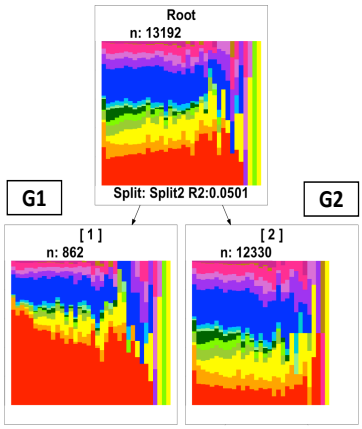


Figure 35: The dissimilarity scores distribution of all the sequences of AEs

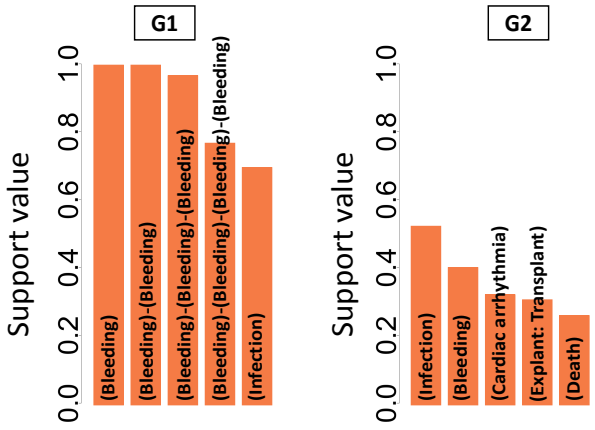
The result of hierarchical clustering is presented as a dendrogram, shown in Fig. 36a, which depicts the taxonomic relationship between clusters formed at each level of grouping. The bottom of the dendrogram represents all the patients' dissimilarity scores, which is a dense dark area because of the large number of patients. Moving from bottom to top of the dendrogram, one cluster is formed at each level of hierarchical clustering by grouping two sub-clusters until, at the top of the dendrogram, all patients are in a single group. The y



(a) Cutting the dendrogram into the two clusters.



(b) A regression tree plotting distributions of various types of AE over temporally ordered AEs in patients' sequences.



(c) Internal validation: support numbers for most common subsequences in two new-formed clusters.

		G1	G2	P-value
1	(Bleeding)-(Bleeding)-(Bleeding)	+	-	0.01
2	(Bleeding)-(Bleeding)-(Bleeding)-(Bleeding)	+	-	0.01
3	(Infection)-(Bleeding)-(Bleeding)	+	-	0.01
4	(Bleeding)-(Bleeding)-(Infection)	+	-	0.01

(d) External validation: the first four discriminative subsequences between G1 and G2. +/- determines the sign of Pearson's residual.

Figure 36: First step of step-wise cluster evaluation: evaluation of the two-cluster solution

axis, "Height", represents the dissimilarity between clusters (groups of patients) at that level of the tree, which is measured using Ward linkage.

The eventual partitioning of the dendrogram was based on the three criteria: the internal validation (within-group similarity), external validation (between-groups dissimilarity), and clinical interpretation. This was performed iteratively by "cutting" the dendrogram horizontally. An example is shown in Fig. 36a in which a horizontal cut results in two groups (G1 and G2). The first group includes 862 patients (7% of total patients), and the second group that includes 12,330 patients (93%). Their corresponding histograms (similar to Fig. 33) are provided in Fig. 36b, in which the proportions of various types of AE are temporally ordered. Visual inspection of the histogram corresponding to G1 reveals an obvious dominance of the Bleeding AE (red color). This also implies a degree of similarity of patients in this group. By contrast, the histogram corresponding to G2 does not present any single dominant AE. This is understandable since this group is comprised of a much larger proportion of the total number of patients (93%). Thus, the patients in the second group are not similar and require further stratification.

Fig. 36c represents the support values of the five most common subsequences in these two groups. In the first group (G1), the unary sequence (Bleeding) and the binary sequence (Bleeding)-(Bleeding) are the most common, both with support value of 1, indicating that 100% of patients in this group had a minimum of two bleeding AEs in their sequences. The next most common sequences were found to be (Bleeding)-(Bleeding)-(Bleeding) and (Bleeding)-(Bleeding)-(Bleeding)-(Bleeding), with support values of 0.98 and 0.73. It was concluded that the sequences in the first group all shared common subsequences, indicating high within-group similarity. In contrast, the maximum support value for the second group (G2) was approximately 50% corresponding to the (Infection) subsequence. This indicated low similarity within this group, and hence does not pass the internal validation.

As a final test, Fig. 36d shows the external validation in which subsequences which best discriminate sequences of the two groups via Pearson Chi-square test. The table shows the first four most discriminative subsequences, in which the plus (+) and minus (-) sign indicates that the observed frequencies of the subsequences were higher (+) or lower (-) than equally distributed frequencies. Here, p-values below 0.05 were taken to indicate

discrimination between groups. The subsequence of (Bleeding)-(Bleeding)-(Bleeding) was the most discriminating subsequence with p-value of 0.01 and support percentage for G1 greater than 98%. All of the remaining discriminative subsequences consisted of at least two bleeding AEs, emphasizing that multiple bleeding AEs was responsible for differentiating G1 from G2. Accordingly, it was concluded that the groupings were externally validated.

In summary, the initial two-cluster solution passed the external validation, but only G1 passed the internal validation. However, the second group which includes the remaining 93% of patients required further subdivision. This was accomplished in a similar manner, by bisecting the dendrogram at a lower level, effectively separating the second group into two sub-groups of 6,168 and 6,162 patients. This was followed by the same process to evaluate the external and internal validation. This procedure was repeated until all the resulting groups passed validation, resulting in a final number of seven groups (summarized in Table. 7). For convenience, each of the seven groups was given a mnemonic name based on visual inspection of the histogram including: GRP1: "Recurrent bleeding", GRP2: " Trajectory of device malfunction & explant", GRP3: " Infection", GRP4: "Trajectories to transplant", GRP5: " Cardiac arrhythmia", GRP6: "Trajectory of neurological dysfunction & death", and GRP7: " Trajectory of respiratory failure & renal dysfunction & death". For example, this histogram of GRP1 reveals an obvious dominance of the Bleeding AE (red color), and was therefore given the name "Recurrent Bleeding." In a similar fashion, G2 revealed a dominance device malfunction (forest green) and explant (lime green.) Since the two are always related sequentially, this group was named "Trajectory of device malfunction & explant."

Fig. 37 demonstrates clustering results through a regression tree. There is histogram associated with each of the groups; the ordinate of which reflecting the proportion of each color-coded AE type (from 0% to 100%) and abscissa reflecting the location in the respective sequences. Since each group has a unique maximum length, this is reflected in the varied width of these plots. Each of these groups were assigned a descriptive title that reflected the dominant AE or AEs therein. For instance, the dominant colors in the GRP2 plot are yellow green (representing device malfunction) and spurious green (representing Explant outcome). Sequence analysis for GRP2, similar to sequences analysis for GRP1 in Fig. 36c& 36d,

reveals that 1,097 out of 1,193 patients who experienced device malfunction eventually had the device explanted, indicating the temporal pattern (subsequence) of (Device Malfunction)-(Explant). The sequence analysis was performed for each of the seven groups.

Table 7 provides statistics of each patient's group. GRP3 and GRP4 had the highest number of patients and AEs by having 26% and 25% of total number of patients, respectively, and 27% and 17% of total recorded AEs, respectively, in this study. On the other hand, GRP1 had the lowest number of patients, 7% of the total number of patient (862 patients), while they had 14% of total AEs. In addition, patients in GRP1 had the greatest number of AEs with average number of 9.73 AEs, and minimum and maximum numbers of 3 and 36 AEs. The average numbers of AEs in other groups were $\leq$5 and minimum numbers of AEs were 1. Columns of 6 and 7 of Table. 7 shows information related to the time of AEs occurrences measured by the months after the LVAD implants. The distributions of post-LVAD time (month) of AEs occurrences were skewed to the right in all the groups as the means were greater than the medians. The average time of AEs occurrences were less than 13 months in all the groups and the median time were less than 7 months. AEs of GRP7 occurred at the earliest post-LVAD time by average of 6.06 months after LVAD and median of 1 month after LVAD. The last two columns of Table. 7 presents information related to the time span of patients' AEs (time of last AE - time of first AE) measured in month. The distributions of time span of patients' AEs were also skewed to right indicating higher number of patients experienced AEs in a short time span. GRP1 had the longest time span of AEs by average of 19.48 months between the first AE and last AE, while, the GRP6 and the GRP7 had the shortest time span by average of approximately 6 months and median of 1 month.

6.1.3.3 Markov Chain Models of AEs

The following sections presents results of Markov modeling (MM) within each of the groupings of patients presented above. The chains of transitions between AEs are presented with the graphs in which the size of the circles represent the frequency of AEs at each position in the chain and the thickness of the arrows reflecting the frequency of each AE that are followed by the subsequent AE.

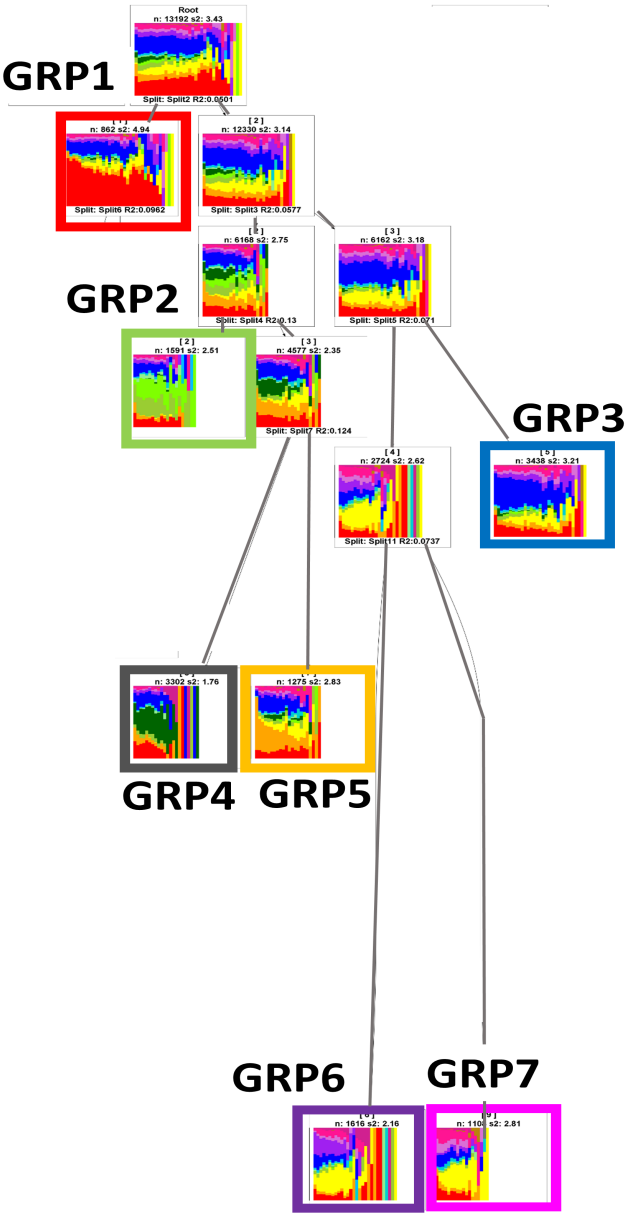


Figure 37: Regression tree showing groups formed at each step of hierarchal clustering. Groups are split to smaller groups until both internal and external criteria are satisfied. Groups are numbered based on the steps in which they are formed.

Table 7: Summary of statistical information related to groups resulted from hierarchical clustering.

		Number of AEs			AEs post-LVAD time (month) †		AEs time span (month) ††	
Number of patients (*)	Number of AEs (**)	Min/Max	Mean	Median	Mean	Median	Mean	Median
GRP1:Recurrent bleeding								
862 (≈7%)	8,388 (≈14%)	3/36	9.73	9	12.55	7	19.48	16
GRP2:Trajectory of device malfunction & explant								
1,591 (≈12%)	7,049 (≈12%)	1/21	4.43	4	11.19	6	9.89	4
GRP3:Infection								
3,438 (≈26%)	15,771 (≈27%)	1/31	4.59	3	11.31	6	10.25	4
GRP4:Trajectories to heart transplant								
3,302 (≈25%)	10,093 (≈17%)	1/22	3.06	3	7.98	5	7.31	4
GRP5:Cardiac arrhythmia								
1,275 (≈10%)	5,715 (≈10%)	1/22	4.48	3	8.63	3	9.48	4
GRP6:Neurological dysfunction & death								
1,616 (≈12%)	5,911 (≈10%)	1/28	3.66	3	9.55	5	6.42	1
GRP7:Trajectory of respiratory failure & renal dysfunction & death								
1,108 (≈8%)	5,648 (≈10%)	1/18	5.10	5	6.06	1	5.78	1

* % of the total 13,192 patients in this study

** % of the total 58,575 recorded AEs in this study

†AEs post-LVAD time is based on the month after LVAD implant (0 post-LVAD month means AE occurred at the time of LVAD implant)

†† AEs time span = time of the last AE (post-LVAD month) - time of first AE (post-LVAD month)

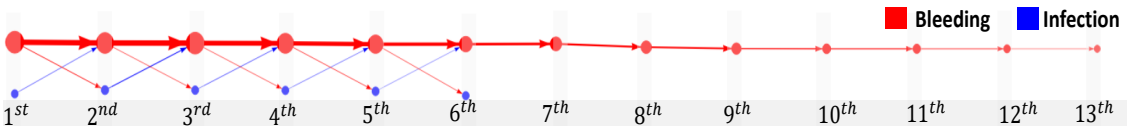


Figure 38: MM of GRP1: Recurrent bleeding (n= 862 patients)

- ***GRP1: Recurrent bleeding***

  All the 862 patients in GRP1 had at least two bleeding AEs, and among them, 98% (847 patients) had at least three bleeding AEs. The Markov chain for GRP1, shown in Fig. 38 is characterized by a long sequence of recurrent bleeding AE's (red circles), with a limited amount of branching involving an intermediate infection AE (blue circles). The frequencies of transitions depicted by the thickness of the arrows is seen to diminish progressively along the chain over the time. For example, the transition frequency for the 1st through 5th bleeding exceeded 300; but beyond the 5th bleeding event, the frequency reduced to the range 100-250. This analysis revealed that the probability of recurrent bleeding exceeded 50% for all instances up to the 13th AE. The transitions probabilities from infection AEs to bleeding AEs were 50% to 65%, while from bleeding AEs to infection AEs were 14% to 18%.

- ***GRP2: Trajectory of device malfunction & explant***

  75% of 1,591 patients in GRP2 experienced a device malfunction and 94% had Explant as their final outcome. The Markov chain for GRP2 (Fig. 39) was found to be much more diverse than the chain for GRP1. Multiple paths involving device malfunction (dark green) were found, although the terminal event was most commonly (>50%) device explant (light green). Only a small number of patients in this group (n=111, approximately 7%) experienced device explant as the initial, isolated AE (indicated by the small light green node at the bottom of the 1st column of Fig. 39). The majority of patients for which device explant was recorded was preceded by another AE, most commonly bleeding and infection (probability between 20-30%). There were also about 18% and 19% probabilities of recurrent device malfunction as the 2nd AEs or the 3th AEs, respectively.

The frequency of transition from Infection as the 2nd AE to Explant was 21%, but with no reported device malfunction AE (n=211, the spurious green circle in the 1st column).

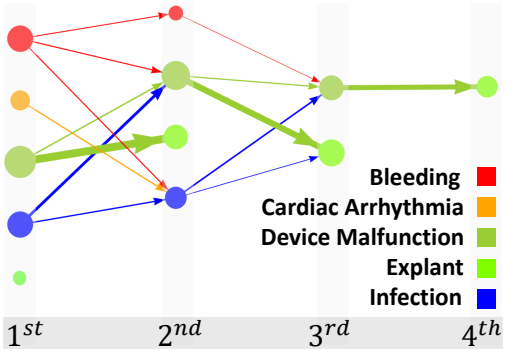


Figure 39: MM of GRP2: Trajectory of device malfunction & explant (n= 1,591 patients)

- ***GRP3: Infection***

  The Markov chain for GRP3 (Fig. 40) is characterized by a long sequence of infection AEs (blue circles), corresponding to a total number of 6,462 recorded infection AEs for 3,438 patients in GRP3, with a limited amount of branching involving an intermediate bleeding AE (red circles). This analysis revealed that the probability of recurrent infection ranged from 34% to 49%. The frequencies of recurrent infection AEs depicted by the thickness of the arrows is seen to diminish progressively along the chain. For example, the transition frequency for the 1st through 4th infection AEs ranged 300-420; but beyond the 4th infection event, the frequency reduced to the range of 50-300. The transition probabilities from bleeding to infection were 39% to 45%, while from infection to bleeding were less than 18%. There were also 285 patients (8%) who received a heart transplant after experiencing an infection AE (represented by the dark green circle in the 2nd column in Fig. 40). There were also transitions from cardiac arrhythmia and respiratory failures as the 1st AEs to infection as the 2nd AEs.

- ***GRP4: Trajectories to transplant***

  The Markov chains for GRP4 (Fig. 41) represents AE trajectories of 3,302 patients who ended in receiving a heart transplant. The majority of patients who revived heart transplants was precede by various types of AE, most commonly bleeding, infection, and cardiac arrhythmia. Only 970 patients (29%) in this group received a heart transplant

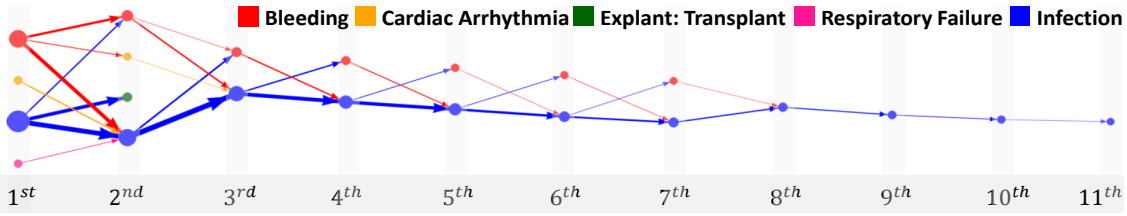


Figure 40: MM of GRP3: Infection (n= 3,438 patients)

as the initial, isolated event (the dark green circle in the first column of Fig. 41). AE trajectories to heart transplants from some specific types of AE like bleeding, infection, or cardiac arrhythmia were more likely than other types of AE like neurological dysfunction AE or device malfunction AE. For instance, 520 patients who experienced only one AE and then received a heart transplant were preceded by mostly bleeding or cardiac arrhythmia (cumulative 388 patients), and minimally by neurological dysfunction or device malfunction (cumulative 132 patients).

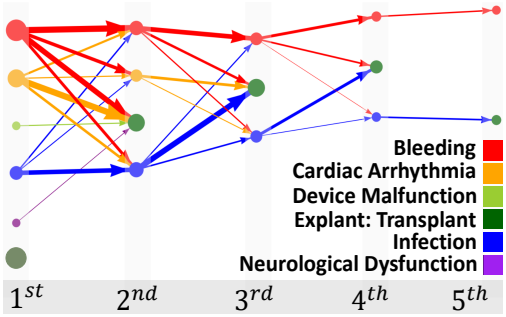


Figure 41: MM of GRP4: Trajectories to transplant (n= 3,302 patients)

- ***GRP5: Cardiac arrhythmia***

  Fig. 42 represents a chain of recurrent cardiac arrhythmia AEs (orange circles) in GRP5 (1,275 patients), with a limited amount of branching involving an intermediate infection AE (blue circles). The frequency of transitions depicted by the thickness of the arrows is seen to diminish progressively along the chain. As an example, 919 patients in GRP5 (72%) experienced cardiac arrhythmia as 1st AE and only 242 of them experienced cardiac

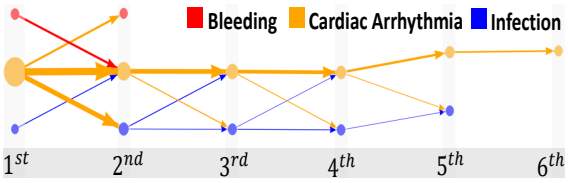


Figure 42: MM of GRP5: Cardiac arrhythmia (n= 1,275 patients)

arrhythmia as the 2nd AE too; beyond the 2nd AE the frequency gradually decreased to 71 for cardiac arrhythmia as the 6th AE. The probability of transitions for recurrent cardiac arrhythmia AEs were between 34% to 49% over the time. There were also transitions between cardiac arrhythmia and infection with probabilities from 16% to 36%.

- ***GRP6: Trajectory of neurological dysfunction & death***

  The Markov chain of GRP6 (Fig. 43) shows the trajectory of 1,616 patients who died (yellow circles) after suffering from neurological dysfunction AEs (purple circles). The rate of death for patients who suffered from neurological dysfunction AEs as the 1st AE trough 3th AE ranged from 27% to 39%. The majority of patients for which neurological dysfunction AE was recorded was preceded by other types of AE, most commonly bleeding and infection. Only a small number of patients in GRP6 (n=229, approximately 14%) died with no reported AEs (the yellow circle in the 1st column of Fig. 43). There were also a small number of patients who died after one or recurrent infection AEs with no recorded neurological dysfunction AE (thin blue arrows from blue circles to yellow circles).

- ***GRP7: Trajectory of respiratory failure & renal dysfunction & death***

  The Markov chain of GRP7 (Fig. 44) illustrates two different AE trajectories to death. One main trajectory represents 929 patients who died after suffering from respiratory failure (934 recorded respiratory failure AE) and/or renal dysfunction AEs (712 recorded renal dysfunction AEs) with transition probabilities between 22% to 44%. It also revealed that patients with reported respiratory failure and renal dysfunction were preceded by other types of AE, most commonly by infection AE and bleeding AE. The transition probabilities from renal dysfunction or infection to respiratory failure ranged from 32%

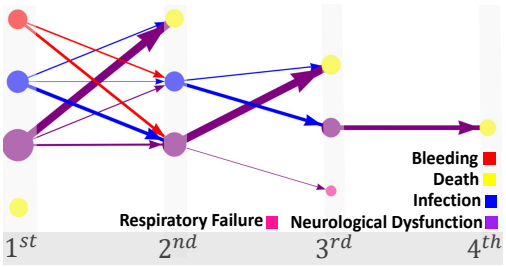


Figure 43: MM of GRP6: Trajectory of neurological dysfunction & death (n= 1,616 patients)

to 36%. Another trajectory of this group presents 179 patients who died after suffering from one or recurrent bleeding AEs with transition probabilities from 19% to 24%.

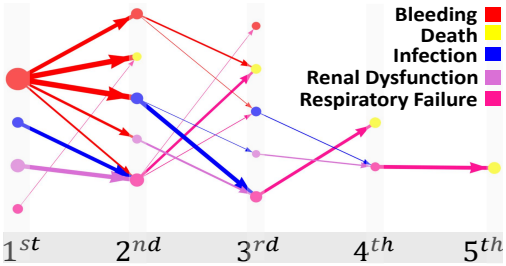


Figure 44: MM of GRP7: Trajectory of respiratory failure & renal dysfunction & death (n= 1,108 patients)

6.1.4 Discussion

This is the first study that shed the light to the entire AEs trajectories after LVAD implant in patients with advanced heart failure. In contrast with previous studies which isolated each AE in their analysis, these trajectories look at the patterns of sequential transitions between various types of AEs and their relationship with the final outcomes of LVADs. These patterns gave clinicians its first intimation of the potential inter-relation or bootstrapping effect between specific types of AEs related to LVAD.

Table. 8 summarized the highlights of each AE pattern and also provided findings from post-LVAD AEs studies that intuitively back up our findings. For example, clinical findings

Table 8: Summary of Clinical Insights of the Groups Resulted from Hierarchical Clustering.

GRP1:Recurrent bleeding
Patients who experienced recurrent bleeding AEs after LVAD. Recurrent occurrences of various types of bleeding such as gastrointestinal bleeding is commonly reported in clinical studies of patients who receive an LVAD [42, 45, 110]
GRP2:Trajectory of device malfunction & explant
Patients who had device malfunction and commonly had their LVADs removed (explanted). This trajectory was found to be preceded by the two types of AEs including infection and bleeding. Two others, less common, trajectories within GRP2 were 1. patients who had an explant without device malfunction and 2. patients who had a device malfunction without ending in explant. In clinical literature, device malfunction is defined as failure of one or more parts of LVAD that cause the inability to maintain adequate circulatory support. Device malfunction might be deadly or could be solved by replacing the device (explant) [186, 187]. It is important to note that within INTERMACS patients there may be some who had serious pump malfunction such as internal thrombosis, but the patient was not considered a candidate for pump exchange and the LVAD may have been simply turned off.
GRP3:Infection
Patients who suffered mostly from infection AEs. The most recent INTERMACS annual report indicated infection as the most frequent AE after bleeding during the first three months and the most common AE thereafter [20].
GRP4:Trajectories to heart transplant
Patients who received a heart transplant and the pump was explanted as part of the procedure. AE trajectories to heart transplant were mostly consisted of bleeding AE, infection AE, and cardiac arrhythmia AE. In practical terms these AEs resulted in an upgraded listing for cardiac transplant resulting in a higher likelihood of achieving a heart transplant. It is not uncommon for LVAD complications to drive more urgent listing status of a candidate and this analysis conforms that strategy. The most recent INTERMACS annual reports indicated slightly more than 30% of heart transplant candidates who received continuous-flow LVAD received a heart transplant [20, 188].
GRP5:Cardiac arrhythmia
Patients who experienced one or recurrent cardiac arrhythmia AEs after LVAD implant that accompanied mostly with infection AEs and bleeding AEs. Patients with ventricular arrhythmias tend to have recurrent episodes of these rhythm disturbances one they begin to manifest them.
GRP6:Neurological dysfunction & death
GRP6 trajectory is supported by clinical literature highlighting the high mortality in LVAD patients with neurological events especially with hemorrhagic strokes [47, 48, 108, 109]
GRP7:Trajectory of respiratory failure & renal dysfunction & death
GRP7 trajectory is predominated by both respiratory failure AEs and renal dysfunction AEs that ended in death although some of the trajectories also were accompanied by a bleeding AE and an infection AE. The eighth INTERMACS annual report named renal dysfunction and chronic pulmonary disease among the non-cardiac system commodities that impact the LVAD survival rate [20].

showed the high rate of death after neurological dysfunction which supports AE pattern in *GRP6 (Trajectory of neurological dysfunction & death)* [47, 48, 108, 109]. Another example is related to the imbalanced distribution of patients among the groups that resulted from clustering which implies high incidence of specific AE patterns, *GRP3 (Infection)* and *GRP4 (Trajectories to transplant)*, among the patients with LVAD which has been supported by findings from clinical studies indicating infection as the most common AE after bleeding and the 30% rate of heart transplant after LVAD [20].

This study was also as an illustration for challenges raised when applying well-known methodologies in large diverse clinical data and creative ways to overcome those challenges. The greatest challenge encountered in this study was the high diversity between patients AEs collected from 13,192 patients from over 150 clinical centers. Although, we used the pipeline described in Zhang et al. [180] that includes clustering and Markov modeling, we found out that its extension to such a large, diverse data set of this study was not straightforward. For instance, the choice of criteria to evaluate clustering results and decide about the number of groups was a challenging task. Zhang et al. [180] used Silhouette values to determine cluster number and support values for internal evaluation of clustering. However, the number of patients in this study (13,192 patients) was more than 10 times greater compared with the number of patients in [180] (1,576 patients). Thus, considering only support values was inadequate. It is obvious that high numbers of groups will result in increasing similarity of patients within each group; however, the number of groups should be limited to preserve the clinical utility of the results. This was the motivation to the step-wise clustering evaluation that was implemented that considers both within-group similarity and between-groups dissimilarity criteria.

Markov models in this study considered the order of occurrences for AE transitions in patient's sequences. For instance, the transition from bleeding as the first AE to infection as the second AE was considered different from transition from bleeding as the fifth AE to infection as the sixth AE. Thus, the probabilities of transitions were evaluated by considering where in the patient's sequence transitions occurred. This temporal constraint is very useful since AEs have a different effect on LVAD final outcome when they occurred in immediate succession as compared to a sequence with a different intermediate AE.

The time of occurrence following implant is very important to the analysis of AEs. It is highly valuable for physicians to know how many days or months after LVAD implant AEs are likely to occur or how quickly a series of AEs may occur for a patient. The time analysis of AE sequential patterns (Table. 7) indicates various timing characteristics among the groups. For instance, AEs in the GRP7 with respiratory and renal failure occurred in the immediate months after LVAD implant with median time of 1 month after LVAD implant reflecting typical post-operative timing of these events. This is in contrast to AEs in the bleeding group (GRP1) with median time of 9 months after LVAD implant, more typical of onset of gastrointestinal bleeding. This analysis also revealed differences in time span over which AEs occurred between groups. For example, AEs for patients with neurological events (GRP6) occurred over a short period of time (median time span of 1 month) contrasted with GRP5 patients whose AEs spanned a median of 4 months. These different timing characteristics of AE sequential patterns may have implications in guiding post-LVAD medical interventions or preventative measures.

Limitations:

One main limitation of this study is related to the voluntary collection and reporting of INTERMACS data. As an example, some AEs like infection is a longitudinal AE that might last for a while, but INTERMACS only records the occurrence of AE without recording its duration. Another issue is that there is no information regarding the order of concurrent AE (events occurred at the same day). The problem was exacerbated by the fact that this registry is comprised of contributions of over 100 centers, and hence involves differences in interpretation of definitions, omissions, and data entry errors. As an example, the ongoing change in the definition of right heart failure (RHF) causes inconsistency between studies about analysis of RHF, and therefore, reduces the confidence in results. Consequently, it would be helpful to pull in expertise from field to learn more about INTERMACS definitions and workflow of data collection, to avoid bias in the future studies.

The time gaps between sequential AEs were not considered in clustering and Markov modeling to minimize the diversity of sequences. One solution that was evaluated was to segment post-LVAD time, based on the critical points based on the AEs distribution and physicians' suggestions. This reduced the maximum number of time points in the sequences

from 36 to 7. However, it was not adequate to prevent enormously increasing diversity. Further consolidating the timeline into short-term and long-term could be another solution that requires some iterative process to find an optimum. Adding time gaps will also help concurrent AEs issue by considering basket of AEs with 0 time gaps as one element in the sequence [189].

Clinical Translation and Future Work:

The AE patterns found in this study can be used as the basis for personalized medical management of the LVAD patient. To reach this goal, future studies should investigate the clinical profile of patients in each AE pattern to explore the clinical scenarios behind these patterns. In addition, the probabilistic Markov models of these patterns can be used as a prognostic tool to predict the next AE (AE_{n+1}) by considering previous AEs (AE_1 to AE_n). This would be an important contribution, and a valuable tool for physicians to optimize treatment to minimize the risk of future AEs.

6.1.5 Conclusion

This study, to the best knowledge of the authors, was the first exploratory to discover sequential chains of AEs following LVAD implant. Mining of the AE sequences of 13,192 patients with advanced heart failure derived from the INTERMACS registry revealed the existence of seven groups of sequential chains of AEs, each characterized by a dominant AE or multiple AEs and occurring in a unique order. The discovered chains of AEs disclose potential interdependence between AEs and provide clinicians a valuable insight into the patient oriented post-LVAD AEs evidence. It is hoped that this analysis may support post-LVAD follow-up by alerting medical providers of the likelihood of impending AEs - based on a combination of independent factors, and patterns of prior AEs.

6.2 Clustering Considering Timeline of Adverse Events

6.2.1 Objective

The clustering analysis in Section 6.1 disregarded the timeline of patients' AE journeys and instead focused more on grouping patients based solely on the similarity of AE types and their order. In contrast, the main objective of this research is to conduct a cluster analysis of post-LVAD AEs within the first year after implantation, considering the time gaps between AEs and combinations of AEs that occur within the same time interval. Another objective of the research is to determine the optimal time granularity for aggregating AE records in the cluster analysis. Two different time granularities were compared: a compact granularity and a relaxed granularity. The compact granularity involves aggregating AEs occurring within each month into twelve separate records that are evenly spaced in time. On the other hand, the relaxed granularity involves aggregating AEs occurring in the first month into one record, and AEs occurring in all subsequent months into a second record. Overall, this study aims to achieve two main objectives: conducting a cluster analysis of post-LVAD AEs by considering time gaps and combinations of AEs, and determining the best choice of time granularity for aggregating AE records in the cluster analysis.

6.2.2 Methods

6.2.2.1 Framework

First, the AE records of each patient were transformed into multiple sequences (channels), with one sequence for each type of AE. Subsequently, hierarchical clustering was performed on the AE sequences to group patients with similar AE patterns. Finally, hidden Markov modeling (HMM) was employed to extract the patterns of transitions between AEs within each group.

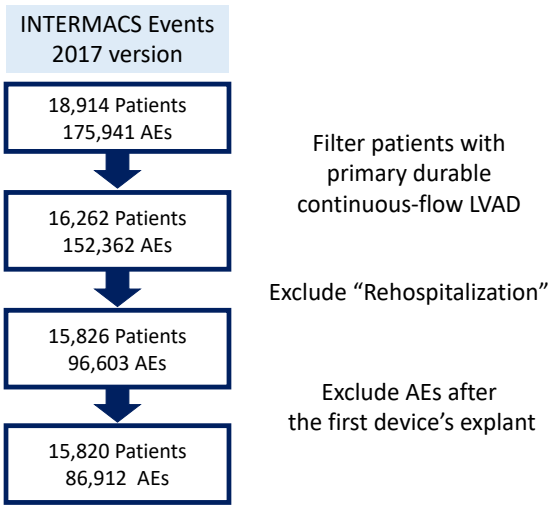


Figure 45: INTERMACS Events data selection for this study.

6.2.2.2 Study population

The flow diagram in Fig. 45 outlines the criteria for inclusion of 86,912 recorded AEs of 15,820 patients (mean age of 57; 12,429 male vs. 3,378 female) who received a continuous-flow LVAD between 2008 to 2016, extracted from the publicly accessible INTERMACS *Event* data set via BioLINCC (2017 version). Fig. 46 [190] depicts the distribution of each type of AE separately over the time. The majority of AEs occur before the first year, especially in the first month post implant with 24,666 recorded AEs. Bleeding, Cardiac Arrhythmia, Infection, and Other SAE were common types of AEs. Out of a total 26 types of AEs, "Rehospitalization" type was excluded from the study as we deemed it a consequence of an AE rather the AE itself. For patients with multiple device implants, the AEs after the first LVAD explant were also excluded.

6.2.2.3 Construction of patients' multi-channel AE sequences

Fig. 47 shows a small extract of the INTERMACS data for a patient (P_{3333}) that includes recorded AEs spanning the first 3 months with five entries. By choosing a monthly

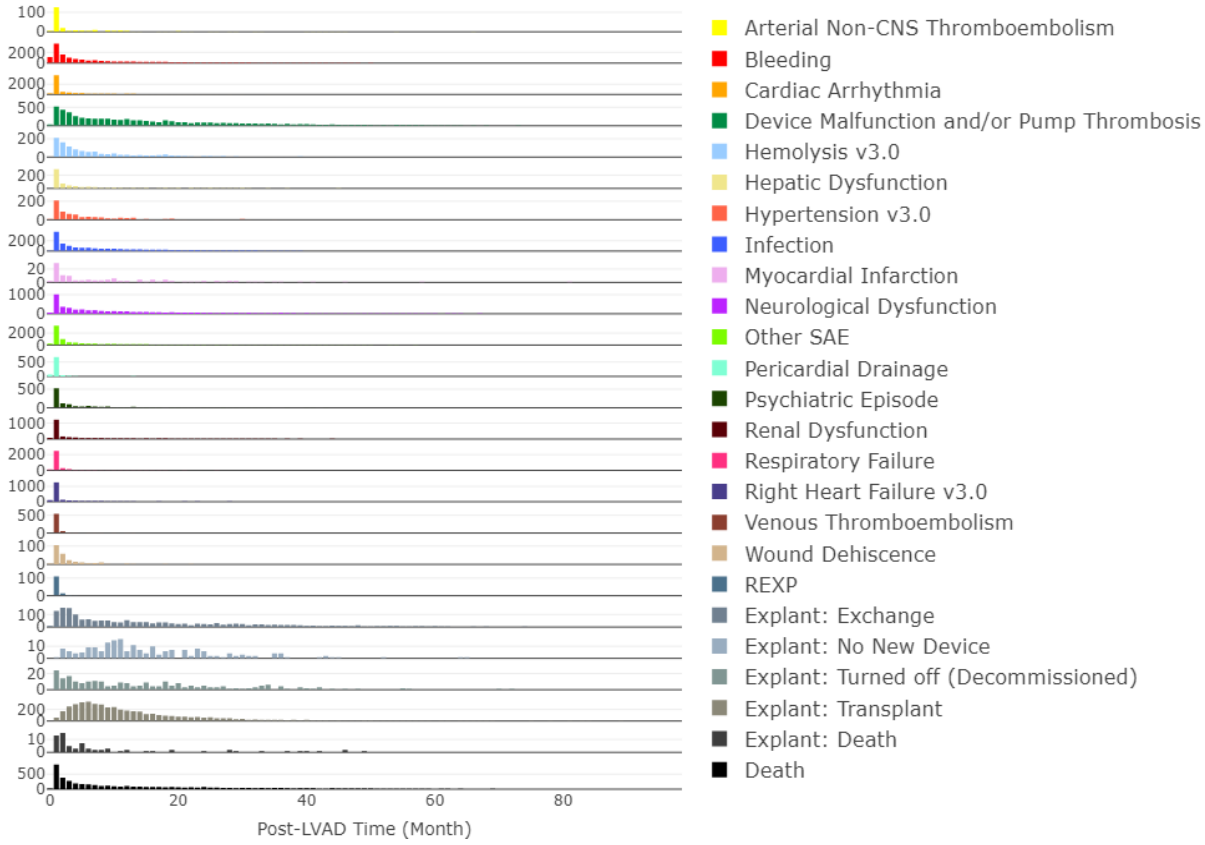


Figure 46: Bar-plots of distribution of each type of AE over the post-LVAD time, separately [190].

granularity, these five entries result in four episodes or elements of their AE sequence: one entry of "NAE" (no AE) for the day of implant (as INTERMACS starts recording AE from the day of implant), three entries in the first month, one in the second month, and one in the third month. Another time granularity selected for this study is to consolidate the AEs into 3 segments as shown in Fig. 47, in which the first two elements, the day of implant and the first month, are similar to the monthly-segment sequence, and then all AEs after the first month are aggregated into the last element. The rationale for this choice is based on the high occurrence of AEs proximate to the LVAD implant.

Each sequence of AEs can be presented in two formats, aka "channels." A single-channel includes all AE types, such as bleeding and infection, and their occurrences in a single sequence, as shown in Fig. 47. Another format, considered in this study is "multi-channel" that segregates the sequences according to the type of AE, such as bleeding or infection, and forms multiple sequences (channels) per patient as shown in Fig 48. The time intervals in which no AE was recorded is labeled NAE. In addition, the first element in each channel (sequence) corresponds to AEs on the day of implant (day 0).

6.2.2.4 Dissimilarity measurement between patients

A distance matrix comprised of dissimilarity, $\boldsymbol{d}$, between each pair of patient sequences was computed as the summation of distances across the channels, $c : 1, ...C$, as follow [191, 192]:

$$d(P_n, P_m) = \sum_{c=1}^{C} d_{CHL_c}(P_n, P_m) \tag{6.7}$$

Where the distance (dissimilarity) between each pair of sequences of channel c, d_{CHL}, was defined as follows:

$$d_{CHL_c}(P_n, P_m) = |P_{n_c}| + |P_{m_c}| - 2LCS(P_{n_c}, P_{m_c}) \tag{6.8}$$

where, $|P_{n_c}|$ and $|P_{m_c}|$ are the lengths of the sequences for patients P_n and P_m in channel c, respectively; $LCS(P_{n_c}, P_{m_c})$ is the Longest Common Subsequence (LCS) between P_n and P_m in chanle c, as formulated below:

PATIENT ID	OPERATION ID	EVENT	TIME (month)	TIME (round up)
3333	404	Bleeding	0.22998419	1
3333	404	Renal Dysfunction	0.75566233	1
3333	404	Bleeding	0.85422699	1
3333	404	Respiratory Failure	1.08421118	2
3333	404	Death	2.91932093	3

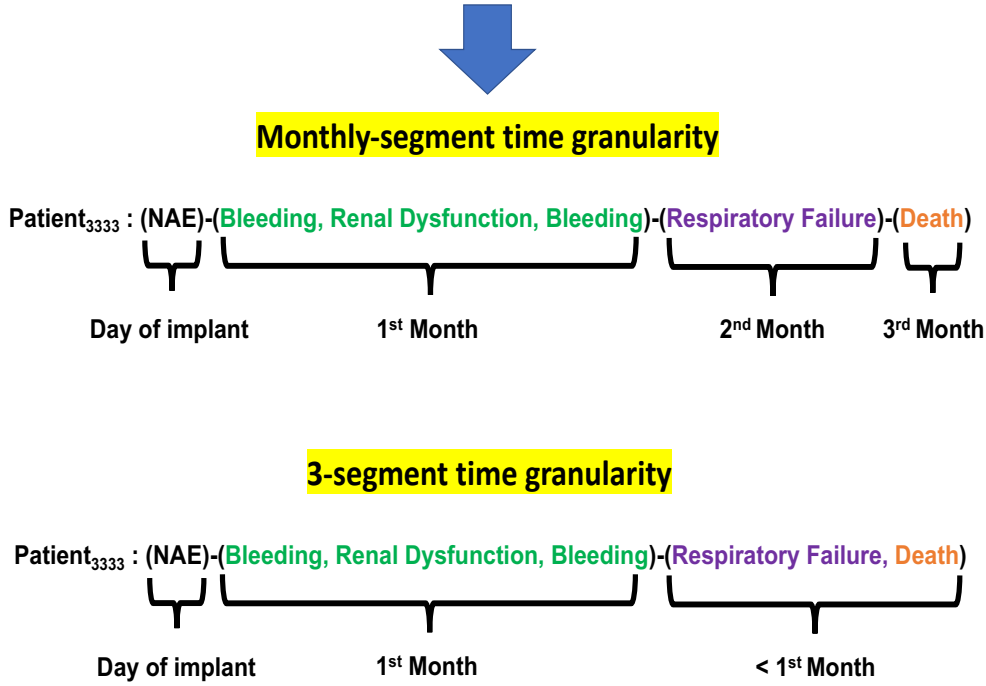


Figure 47: An example of converting INTERMACS Event data for a patient into a sequence of chronologically ordered AEs with two time-granularity choices of monthly-segment and 3-segment.

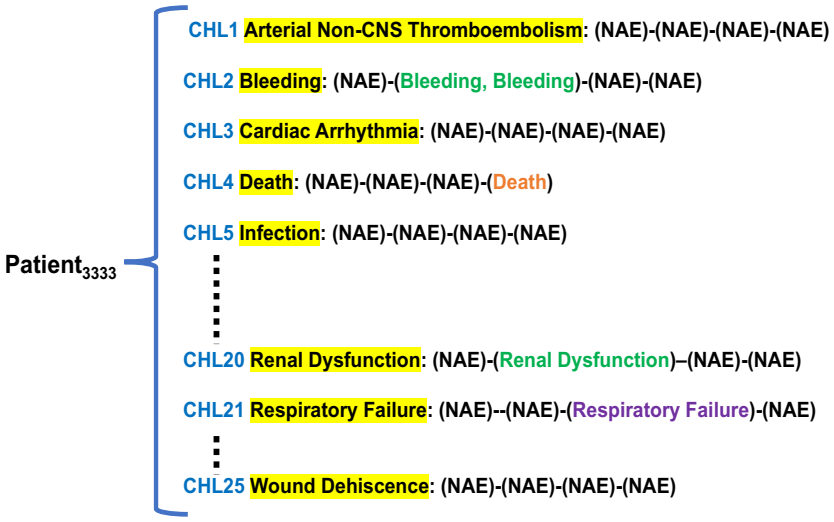


Figure 48: Multi-channel format of P_{2222}'s AE records.

$$LCS(P_{n_c}, P_{m_c}) = max\{|l| : l \in SB(P_{n_c}, P_{m_c})\} \tag{6.9}$$

Here, $SB(P_{n_c}, P_{m_c})$ is a set of all common subsequences between P_{n_c} and P_{m_c} and $|l|$ is the length of common subsequence. A subsequence is a secondary sequence derived from another (primary) sequence by deleting some (or no elements) while maintaining the same order of the remaining elements of the primary sequence. A *common subsequence* is a subsequence that is common to both P_{n_c} and P_{m_c}. As an example, the only common subsequence between the $P_{1_{c:Bleeding}}$ and $P_{2_{c:Bleeding}}$, shown in Fig. 49, is the subsequence of (Bleeding)-(NAE). Thus, in this example, the LCS is 2 and $d_{CHL:Bleeding}(P_1, P_2)$ is 2 which means that with 2 changes (deletion of NAE from the sequence of patient $P_{1_{c:Bleeding}}$ and insertion Bleeding AE, their sequence becomes similar to $P_{2_{c:Bleeding}}$'s sequence. Although the length of the sequences in this example were equal, this study includes sequences with different lengths as survival time differs among the patients. Therefore, the distance between each pair of patients was divided by the geometric mean of the two sequence lengths to normalize the dissimilarity measurement.

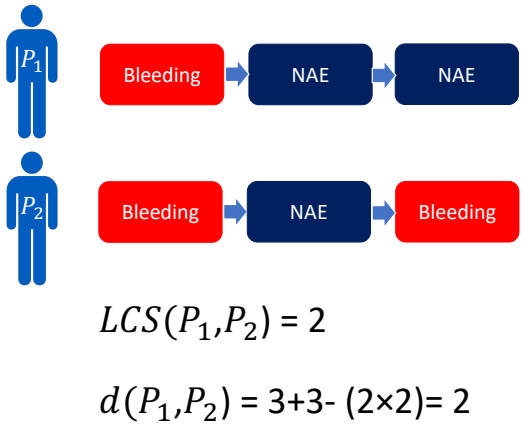


Figure 49: An example of computing dissimilarity score between two patients' sequences.

6.2.2.5 Defining linkage method for hierarchical clustering

After computing dissimilarity matrix of patients' AE sequences, they were clustered using bottom to top hierarchical clustering with Ward linkage. Hierarchical clustering merges sequences with smallest distance d into a single group, and updates the distance matrix for the newly merged group and the remaining patients. This merging process is repeated using Ward's method in which groups of patients with the smallest post-merge, in-group variance (sum of squares) are merged until all patients are in one group. The Ward linkage distance between clusters is computed using Lance–Williams recurrence algorithm [182]. Briefly, considering two clusters of C_i and C_j, the distance between new cluster C_{ij} $(= C_i \cup C_j)$ and remaining clusters such as C_k is formulated as follows [182]:

$$D_{(ij,k)} = \alpha_i D_{(ik)} + \alpha_j D_{(jk)} + \beta D_{(ij)} + \gamma |D_{(ik)} - D_{(jk)}| \tag{6.10}$$

Here, $D_{(ij,k)}$ is the distance between new cluster C_{ij} and cluster C_k. Lance–Williams coefficients α, β, and γ are different for various linkage methods and are defined for Ward linkage as follows:

$$\alpha_z = \frac{|z| + |k|}{|i| + |j| + |k|}\ ,\ z = i, j$$
$$\beta = -\frac{|k|}{|i| + |j| + |k|} \tag{6.11}$$
$$\gamma = 0$$

Here, $|.|$ indicates absolute value, and i, j, and k are numbers of patients in each cluster.

6.2.2.6 Validation of clustering and selection of the number of clusters

The goodness of clustering is often validated using three categories of criteria - internal, external and relative. Internal validation reflects the compactness of the clusters (intra-cluster similarity) and the separation between the clusters (inter-cluster dissimilarity). Two well-known internal validation indices are Silhouette [193] and Dunn [194] that consider both intra-cluster similarity and inter-cluster dissimilarity. The external validation criteria compare the clusters against some optimal grouping of the data, for example grouped by an expert (clinicians) or extracted from the literature. Lastly, the relative validation criterion evaluates the clustering results from repeated clustering of the same data with varying parameters. In many cases, the number of clusters in the data is unknown, which in fact is sometimes the main reason why clustering analysis is performed in the first place. Relative validation is performed by executing a clustering algorithm with different numbers of clusters, and then selecting those that optimize a desired criterion such as the Silhouette coefficient. In this study, internal validation using the average Silhouette width was employed over a range of clusters from 2 to 20.

A lesser-known measure for validating the clustering results is *stability* that evaluates the robustness of clusters to random noise. There are multiple methods for introducing randomness, including perturbing the variables (features) [195], randomness in the sampling (e.g. subsampling) [196, 197], bootstrapping [198, 199], and cross-validation [200]. In this study, stability validation was performed based on clustering robustness against randomness in the samples using bootstrapping. This was chosen for its efficiency in maintaining the size of the original data and providing an instability measure that is a function of number of clusters. The bootstrapping method proposed by Fang and Wang in 2012 [199] was used

to evaluate the path of instability for the number of clusters from 2 to 20.

Further details of the average silhouette width and the bootstrapping method used as the internal and instability validations, respectively, in this study are explained below.

- ***Average silhouette width:*** The silhouette coefficient is composed of the average distance of a patient (p_{i}) to all other patients in its own cluster (intra-cluster, $a(p_{\mathrm{i}})$) and the average distance of a patient to all patients in other clusters (inter-cluster, $b(p_{\mathrm{i}})$) as follows [193]:

$$s(p_{\mathrm{i}}) = \frac{b(p_{\mathrm{i}}) - a(p_{\mathrm{i}})}{\max\{a(p_{\mathrm{i}}), b(p_{\mathrm{i}})\}} \tag{6.12}$$

  When clusters are compact and well-separated from each other, $a(p_{\mathrm{i}})$ is smaller compared to $b(p_{\mathrm{i}})$ resulting in $s(p_{\mathrm{i}})$ being close to 1. Whereas, when the distance between clusters is not large, the $a(p_{\mathrm{i}})$ and $b(p_{\mathrm{i}})$ may not be dramatically different from each other resulting in $s(p_{\mathrm{i}})$ close to 0. If patients are clustered improperly, the $a(p_{\mathrm{i}})$ may be larger than $b(p_{\mathrm{i}})$ resulting in $s(p_{\mathrm{i}})$ being negative. Whereas $s(p_{\mathrm{i}})$ represents the silhouette coefficient for a *single* patient in a data, the overall performance of clustering is computed by average silhouette width (ASW):

$$ASW = \frac{1}{N}\sum_{i=1}^{N} s(p_{\mathrm{i}}) \tag{6.13}$$

  where N is the number of objects in the data set. The optimal number of clusters is the number that maximizes the value of average silhouette width.

- ***Instability:*** This study employed a measure of stability index formulated by Fang and Wang in 2012 [199].

  Briefly, B independent pairs of bootstraps were constructed from the INTERMACS data set, $\left(X_b^{n^*}\widetilde{X}_b^{n^*}\right)$ with $b = 1..., B$. Each sample consisted of n observations drawn from empirical distribution of the data with replacement (assigning a probability of $1/n$ to each of the observed values x_{i}. Hierarchical clustering was then applied to each pair for $b = 1..., B$ to form groupings of $\Psi_{X_b^{n*,k}}$ and $\Psi_{\tilde{X}_b^{n*,k}}$. Then, the empirical distance between each pair of groupings ($d_{\mathrm{groupings}}$) was calculated as:

$$d_{\text{groupings}}\left(\Psi_{X_b^{n^*,k}}, \Psi_{\tilde{X}_b^{n^*,k}}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mid I\left\{\Psi_{X_b^{n^*,k}}(x_i) = \Psi_{X_b^{n^*,k}}(x_j)\right\} - I\left\{\Psi_{\tilde{X}_b^{n^*,k}}(x_i) = \Psi_{\tilde{X}_b^{n^*,k}}(x_j)\right\} \mid \tag{6.14}$$

Finally, the instability index (INS) can be computed as:

$$INS_B = \frac{1}{B}\sum_{b=1}^{B} d_{\text{groupings}}\left(\Psi_{X_b^{n^*,k}}, \Psi_{\tilde{X}_b^{n^*,k}}\right) \tag{6.15}$$

Finally, the optimal number of clusters, k, minimizes the value of INS.

6.2.2.7 Compare Clustering results: Rand index

In this study, the similarity between two clustering results with the same number of clusters is measured using the the Rand index (R) which is calculated as follows [201]:

$$R = (\mathrm{a} + \mathrm{b}) / \begin{pmatrix} N \\ 2 \end{pmatrix} \tag{6.16}$$

where a is the number of times a pair of patients belongs to the same group across two clustering results (correct similar pairs), b is the number of times a pair of patients belong to different groups across two clustering results (*correct dissimilar pairs*), and N is the total number of patients. The denominator is the total number of possible pairs in the set of N patients. The Rand index ranges from 0 to 1. The value of 0 indicates that two clustering results do not agree on the clustering of any pair of patients, whereas the value of 1 indicates that two clustering results agree perfectly on the grouping of every pair of patients.

6.2.2.8 Hidden Markov Model

In this study, Hidden Markov Models (HMMs) were employed to analyze the multichannel sequence data. HMMs are probabilistic models where observations are related to a hidden process following a Markov chain. The hidden states cannot be directly observed but generate or "emit" observations with varying probabilities.

Let us assume we have multichannel sequence data with N individuals, T time points, C channels, and an HMM with S hidden states. We represent the hidden state sequence for individual i from time 1 to t as $z_i = (z_{i1}, z_{i2}, ..., z_{iT})$, and the observation of individual i at time t in channel c as y_{itc}.

The HMM assumes a first-order Markov assumption, where the probability of transitioning to the hidden state at time t depends only on the hidden state at the previous time point $t-1$. Additionally, the same latent structure is assumed to apply to all channels, meaning that the hidden state z_{it} emits observed states y_{itc} in all channels c. The observations $y_{it1}, ..., y_{itC}$ are assumed to be conditionally independent given the hidden state z_{it}.

A discrete first-order hidden Markov model for multichannel data is characterized by the following probabilities:

- Initial probability vector $\pi = \{\pi_s\}$ of length S, where π_s is the probability of starting from hidden state s: $\pi_s = P(z_{i1} = s)$ for $s \in \{1, ..., S\}$.
- Transition probability matrix $A = \{a_{sr}\}$ of size $S \times S$, where a_{sr} is the probability of moving from hidden state s at time $t-1$ to hidden state r at time t: $a_{sr} = P(z_{it} = r | z_{i(t-1)} = s)$ for $s, r \in \{1, ..., S\}$.
- Emission probability matrices $B_c = \{b_s(m_c)\}$ of size $S \times M_c$, where $b_s(m_c)$ is the probability of hidden state s emitting observed state m_c in channel c, and M_c is the number of observed states in channel c: $b_s(m_c) = P(y_{itc} = m_c | z_{it} = s)$ for $s \in \{1, ..., S\}$ and $m_c \in \{1, ..., M_c\}$.

To estimate these probabilities, maximum likelihood estimates are typically calculated using the Baum-Welch algorithm, which is an expectation-maximization (EM) algorithm for HMMs. The log-likelihood of the parameters $M = \{\pi, A, B_1, ..., B_C\}$ for the HMM is given by:

$$\log L = \sum_{i=1}^{N} \log P(Y_i | M) \tag{6.17}$$

where Y_i represents the observed sequences in channels $c = 1, ..., C$ for subject i. The probability of the observation sequence of subject i given the model parameters is:

$$P(Y_i|M) = \sum_{\text{all } z} P(Y_i|z, M)P(z|M) \tag{6.18}$$

$$= \sum_{\text{all } z} P(z_1|M)P(y_{i1}|z_1, M) \prod_{t=2}^{T} P(z_t|z_{t-1}, M)P(y_{it}|z_t, M)$$

$$= \sum_{\text{all } z} \pi_{z1} b_{z1}(y_{i11}) \dots b_{z1}(y_{i1C}) \prod_{t=2}^{T} \left[a_{z_{t-1}z_t} b_{z_t}(y_{it1}) \dots b_{z_t}(y_{itC})\right] \tag{6.19}$$

where z denotes the hidden state sequences that take all possible combinations of values in the hidden state space $\{1, ..., S\}$, and y_{it} represents the observations of subject i at time t in channels $1, ..., C$. Here, π_{z1} is the initial probability of the hidden state at time $t = 1$ in sequence z, a_{zt-1zt} is the transition probability from the hidden state at time $t - 1$ to the hidden state at t, and $b_{zt}(y_{itc})$ is the probability that the hidden state of subject i at time t emits the observed state at time t in channel c.

Inference on Hidden States

Given the HMM model and observed sequences, interesting inferences can be made regarding the hidden states. The forward probabilities $\alpha_{it}(s)$ [202] represent the joint probability of hidden state s at time t and the observation sequences $y_{i1}, ..., y_{it}$ given the model M, while the backward probabilities $\beta_{it}(s)$ denote the joint probability of hidden state s at time t and the observation sequences $y_{i(t+1)}, ..., y_{iT}$ given the model M.

From the forward and backward probabilities, we can compute the posterior probabilities of the states, which provide the probability of being in each hidden state at each time point, given the observed sequences of subject i. These probabilities are defined as:

$$P(z_{it} = s|Y_i, M) = \frac{\alpha_{it}\beta_{it}}{P(Y_i|M)} \tag{6.20}$$

where $P(Y_i|M)$ is the overall probability of the observation sequences for subject i. The posterior probabilities can be utilized to find the locally most probable hidden state at each time point, although the resulting sequence may not necessarily be globally optimal. To obtain the single best hidden state sequence $\hat{z}_i(Y_i) = \hat{z}_{i1}, \hat{z}_{i2}, ..., \hat{z}_{iT}$ for subject i, we

maximize $P(z|Y_i, M)$ or, equivalently, $P(z, Y_i|M)$. The Viterbi algorithm [202] is commonly employed to solve this problem.

For a more comprehensive presentation on HMMs for multichannel data, refer to *Helske* (2018) [203].

6.2.2.9 HMMs Comparison

Models with the same number of parameters can be compared using the log-likelihood. However, when choosing between models with different numbers of hidden states, we need to consider the number of parameters involved. To address this, we employ the Bayesian Information Criterion (BIC) which is defined as:

$$BIC = -2\log(L_d) + p\log\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{c=1}^{C} I(y_{itc} \text{ observed})\right) \tag{6.21}$$

where L_d is computed using Equation 6.17, p represents the number of estimated parameters, I denotes the indicator function, and the summation in the logarithm corresponds to the size of the data.

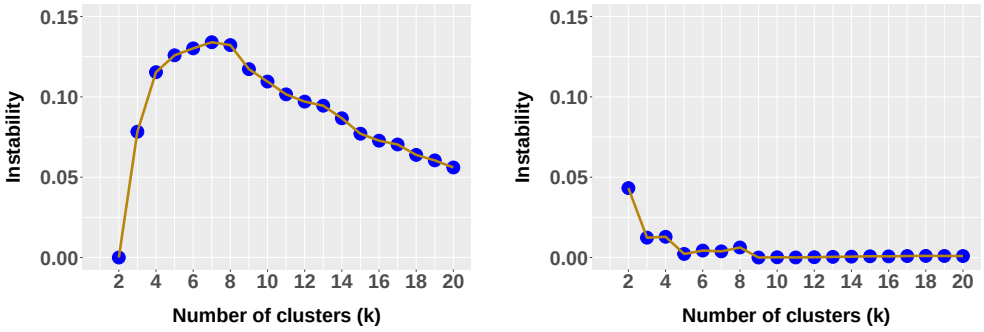
6.2.3 Results

6.2.3.1 Clustering validation

As described in the previous section, the quality of the clustering was evaluated to provide guidance for choosing the number of clusters (k). Two metrics of quality were used: stability, and average silhouette width, with the optimal number of clusters resulting from the lowest value of instability and the highest value of average silhouette width.

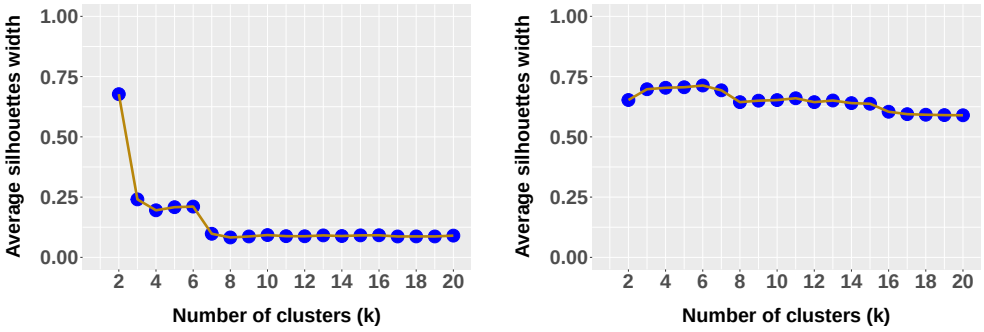
- ***Instability:*** Fig. 50 plots the bootstrap instability measure as a function of number of clusters ($k = 2, 3, ..., 20$) derived from the monthly and 3-segment aggregated time clustering. Although both choices of time granularity provide very stable clusters over the range studied (instability < 0.15), the relationship to cluster count are different. The 3-segment clustering curve expresses a unimodal relationship with a peak at approximately 7, whereas the monthly-segment curve monotonically decreases from 2 to 9 clusters and

remains near zero thereafter. Therefore the optimal number of clusters, from the perspective of instability, is 2 (2-cluster result) for the 3-segment time granularity, and 5 (or ≥ 9) for the monthly-segment clustering.



(a) 3-segment (b) Monthly-segment

Figure 50: Instability measurement for clustering results.



(a) 3-segment (b) Monthly-segment

Figure 51: Average silhouette width for clustering results.

- ***Average silhouette width:*** Fig. 51 plots the average silhouette width, which measures the intra-cluster similarity and inter-cluster dissimilarity. The average silhouette values for 3-segment range from 0.08 ($k = 7$) to 0.68 ($k = 2$). The high average silhouette values for monthly-segment, ranged from 0.64 ($k = 20$) to 0.71 ($k = 6$). These results indicate that the 2-cluster result ($k = 2$) is optimal for 3-segment time granularity and any choice

between 5 to 7 ($5 \leq k \leq 7$) is preferred for monthly-segment time granularity.

Considering both criteria of stability and silhouette values, the 2-cluster result for 3-segment time granularity and 5-cluster result for monthly-segment time granularity were chosen. For the sake of comparison, this study also investigated the 5-cluster result for 3-segment time granularity. The corresponding inter-cluster and intra-cluster distances is provided in Table 9. The average intra-cluster distances of five clusters (called groups), G1 ... G5, are low and in the same range (Table. 9a). This indicates that the AE journeys of patients within each cluster are very similar, irrespective of time granularity. In contrast, comparison of inter-cluster distances in Table. 9b reveals that the monthly-segment clusters are better separated than the 3-segment time granularity clusters.

The weak separation in 3-segment clusters could be due to high similarity among patients' AE journeys in 3-segment time granularity that caused overlapping of patients between groups due to the fact that all AEs after the first month in 3-segment time granularity are aggregated in the third segment. This is contrasted with the aggregation in monthly-segment clustering where the timing of AEs are less similar. Nevertheless, the 5-cluster result for 3-segment has a good intra-cluster similarity, despite its weak inter-cluster dissimilarity. Therefore, it is worthwhile to compare with 5-cluster result for monthly-segment clustering.

Table 9: Comparison of 5-cluster result for 3-segment vs monthly-segment time granularity

(a) Within-cluster average distances

	G1	G2	G3	G4	G5
Monthly-segment	0.26	0.05	0.19	0.24	0.17
3-segment	0.22	0.11	0.24	0.13	0.17

(b) Average distances to other clusters

	G1	G2	G3	G4	G5
Monthly-segment	0.69	0.61	0.43	0.81	0.56
3-segment	0.22	0.23	0.45	0.18	0.19

6.2.3.2 Profiling the clusters

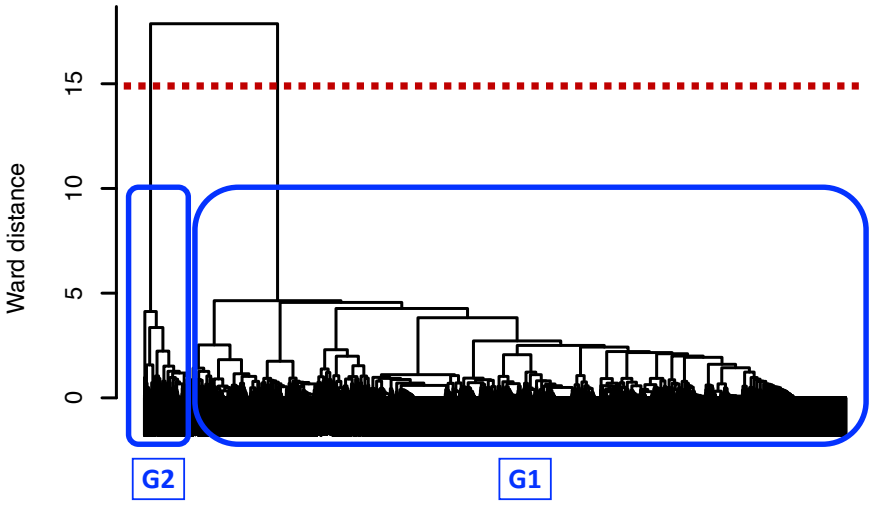
Presented below are the clustering results for the optimal number of clusters for both granularities of time: the 2-cluster result (k=2) for 3-segment time granularity and 5-cluster

result (k=5) for monthly-segment time granularity. Additionally, the 5-cluster result (k=5) for 3-segment time granularity is presented for comparison with the monthly-segment time granularity.

The result of hierarchical clustering for 3-segment time granularity is presented as a dendrogram, shown in Fig. 52a, which depicts the taxonomic relationship between clusters formed at each level of grouping. The bottom of the dendrogram represents all dissimilarity scores, which is a dense dark area because of the large number of patients. Moving from bottom to top of the dendrogram, one cluster is formed at each level of hierarchical clustering by grouping two sub-clusters until, at the top of the dendrogram, all patients are within a single group. The y axis, Ward distance, represents the dissimilarity between clusters (groups of patients) as a function of the level of the tree. Cutting the dendrogram horizontally at the Ward distance of 15 results in two groups, labeled G1 and G2. This partitioning of the dendrogram into two clusters was based on the two criteria described previously: lowest instability (see Fig. 50a) and greatest average silhouette width (see Fig. 51a).

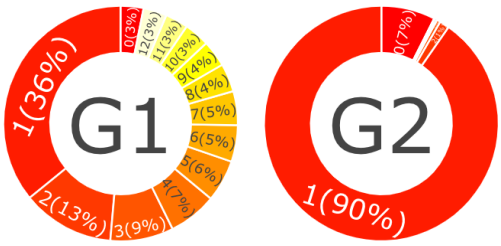
The first group, G1, represents the large majority of patients (n=14,769, 93%), with the total of 56,623 recorded AEs. G2 represents fewer patients (n=1,051, 7%) associated with a total of 4,757 recorded AEs (see Table in Fig. 52a). While patients in G1 and G2 both have an average of 4 to 5 AEs per patient, patients in G2 experienced AEs in a much shorter average time span of 0.4 month vs 5 months in G1. This contrast is also reflected in the pie charts in Fig. 52b that summarize the timing of AEs. The majority of AEs (90%) in G2 occurred during the first month and 7% occurred on the day of implant (time of ≈ 0 month). In contrast, the timing of AEs in G1 extends to 12 months with 36% in the first month. Similarly, the freedom from AE curves in Fig. 52c highlights the difference in the time course of AEs of both groups. The G2 curve goes to 0% by the first month, meaning all patients in G2 experienced at least one AE by the first month, while the curve for G1 reduces more gradually over time to the lowest value of 12% at 12^{th} month. Comparing their survival curves in Fig. 52c, the probability of survival in G2 drops to about $< 20\%$ by the first month compared to G1 with 100% survival probability by the first month and remains above 80% to the end of the first year.

To investigate the patterns of AEs for each group, Fig. 52d plots the ten most common



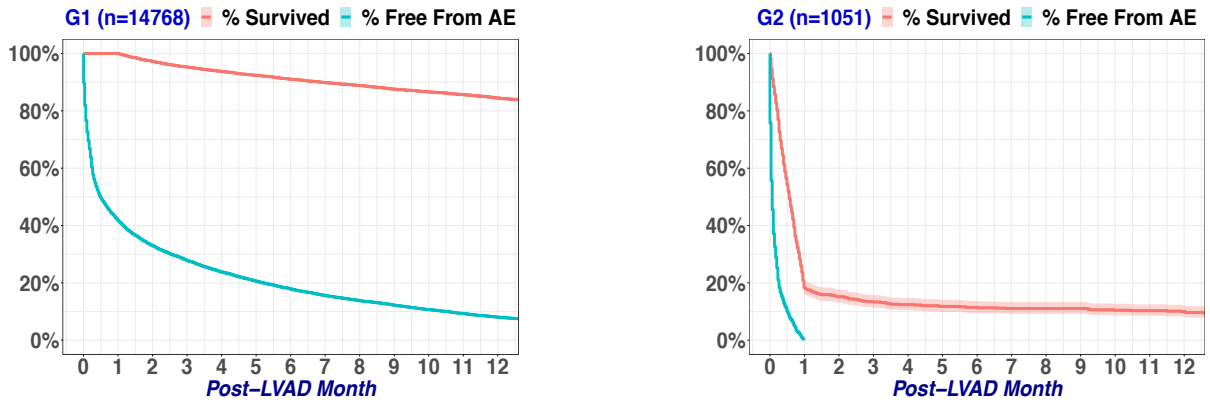
	G1	G2
Count of patients	14769 (93%)	1051 (7%)
Count of episodes of AEs	21322 (94%)	1308 (6%)
Total count of AEs	56623 (92%)	4757 (8%)
Mean count of AEs per patient	4	5
Mean of time span (month) per patient	5.1	0.4

(a) Cutting the dendrogram into the two clusters.



(b) Time distribution of AEs in G1 and G2.

Figure 52: 2-cluster result for 3-segment time granularity.

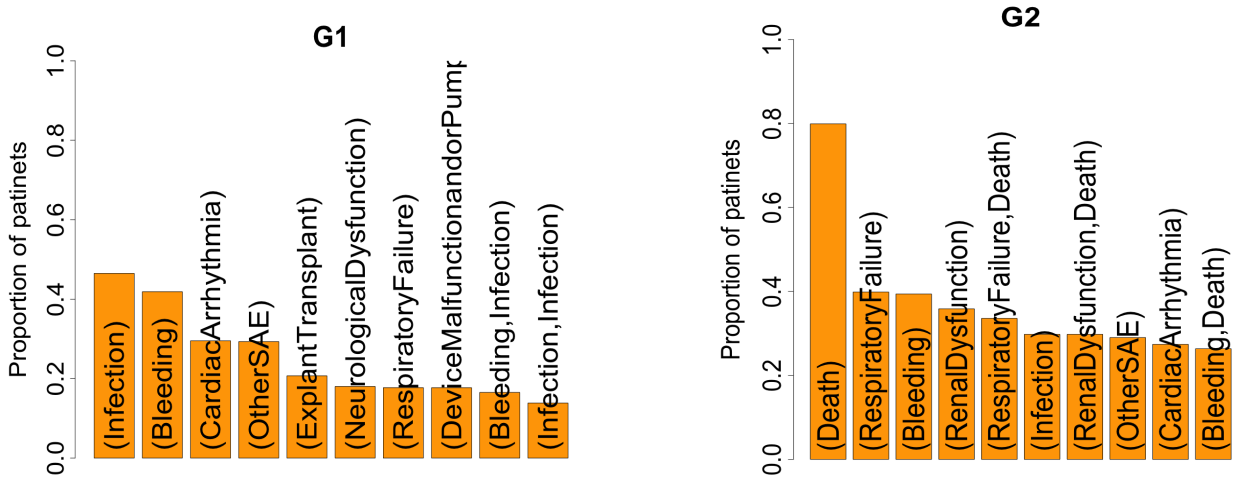


(c) Survival curves and freedom from AE curves for patients in G1 and G2.

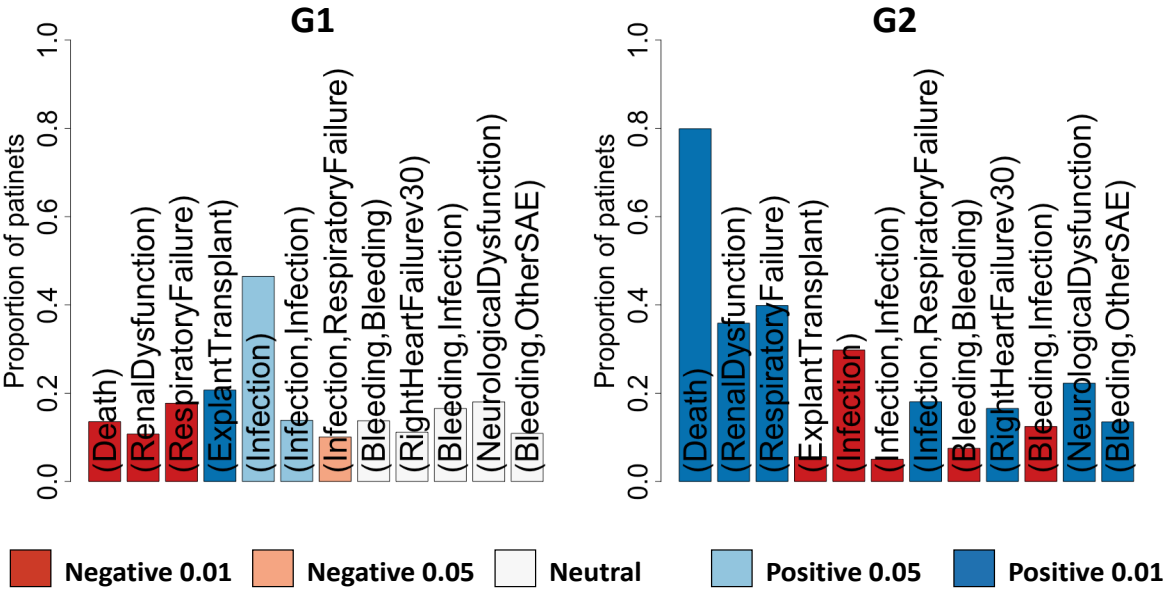
Figure 52: 2-cluster result for 3-segment time granularity.

subsequences of AEs for G1 and G2. The most common subsequences in G1 are the unary subsequences of (Infection) and (Bleeding) which are common to more than 40% of patients due to the high diversity among the large number of patients in G1. On the other hand, the most common subsequence in G2 is (Death): 80% of patients in G2 died in the first year, which corresponds to the survival curve Fig. 52c. The unary subsequences of (Respiratory Failure), (Bleeding), and (Renal Dysfunction) are the most common subsequences in G2. Most of the patients in G2 who experienced Respiratory Failure and Renal Dysfunction died in the first year; (Death, Respiratory Failure) and (Death, Renal Dysfunction).

To further investigate the dissimilarity of patterns among the patients in G1 and G2, a Pearson Chi-square test was conducted for the table that cross tabulates the presence or absence of subsequences and cluster membership. A low Chi-square value means that the distribution among clusters is independent of the presence or absence of a subsequence, and thus, the subsequence is not discriminating. Here, p-values < 0.05 were taken to indicate discrimination between groups. The subsequences that most discriminate between patients in G1 and G2 are shown in Fig. 52e, sorted by their power of discrimination. The color coding corresponds to the magnitude and sign of



(d) The most common subsequences in G1 and G2.



(e) The discriminative subsequences between G1 and G2. Ordered according to Chi-square value (decreasing left to right.) Color coded based on the sign and value of Pearson's residual.

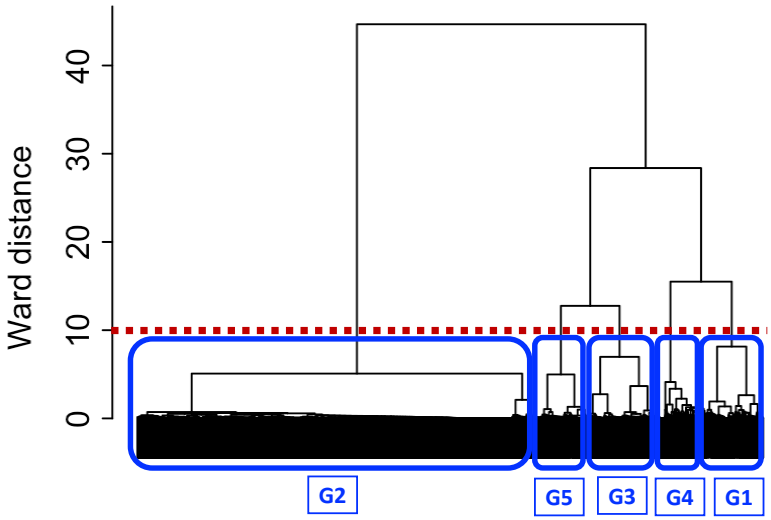
Figure 52: 2-cluster result for 3-segment time granularity.

the Pearson residuals, i.e. the contribution of each corresponding subsequence to the Pearson Chi-squared; $(observed - expected)/\sqrt{expected}$). A negative (positive) Pearson residual indicates that the observed frequencies of the subsequences are lower (greater) than equally distributed frequencies. The three most discriminating subsequences between patients in G1 and G2 are (Death), (Renal Dysfunction), and (Respiratory Failure) that were common among less than 20% of patients in G1 while 80% of patients in G2 died and about 40% of them experienced either (Renal Dysfunction) or (Respiratory Failure).

The result of hierarchical clustering for monthly-segment time granularity is presented in the dendrogram of Fig. 53a, which cuts at the Ward distance of 10 resulting in five groups G1 to G5 (5-cluster results). Similar to the previous analysis, the selected number of clusters was based on criteria of lowest instability (see Fig. 50b) and greatest average silhouette width (see Fig. 51b).

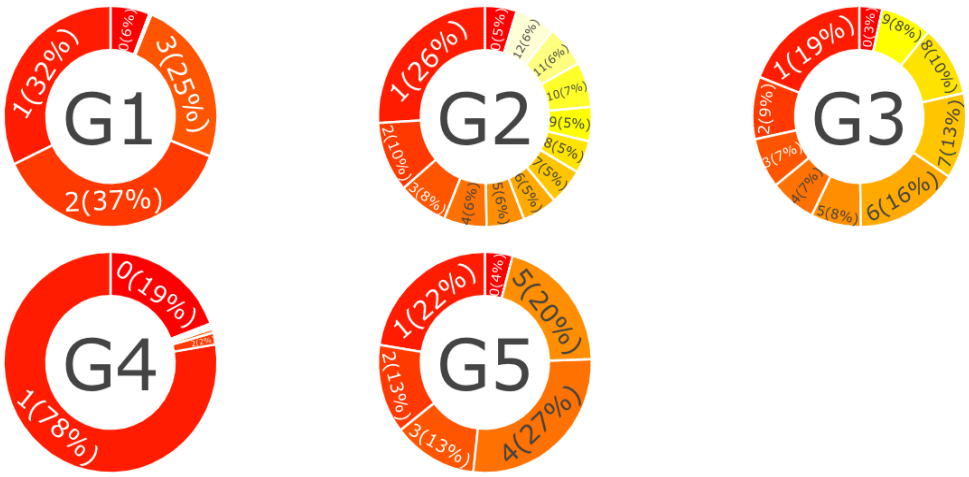
The group with the largest number of patients, G2 (n=10,229), comprising 65% of the total cohort, is associated with 56% of total recorded AEs. The remaining groups each include between 7% and 12% of total number of patients with 8% to 16% of AEs. Although, the average count of AEs per patients are similar among the groups (4 to 5 AEs per patient), the average time span of AE journeys are highly dissimilar. The Table in Fig. 53a reveals that AE journeys of patients in G1 and G4 occurred in a shorter time span (1.5 and 0.4 month, respectively) than those in G2, G3, and G5 (3 to 6 months).

Referring to the pie graphs presented in Fig. 53b, which illustrates the distribution of time over which AEs are recorded, it is observed that patients in G1, G4, and G5 experienced AEs mostly in the early months after implant (0-5 post-LVAD months), while patients in G2 and G3 had records of AEs extending into the later months of the first year; up to 9 or 12 months. The disparity of time span is reflected in the corresponding freedom from AE curves in Fig. 53c. The curves associated with G1, G3, and G5 exhibit a similar shape, decreasing to 0% between 3 and 9 months. On the other hand, the probability of freedom from AEs for G2 with the largest number of patients decreases over the first year to 12%. The most dramatic freedom from AE curve belongs to patients in G4 with 0% probability of freedom from AE by the end of the first month. The similarity in the shapes of the freedom from AE



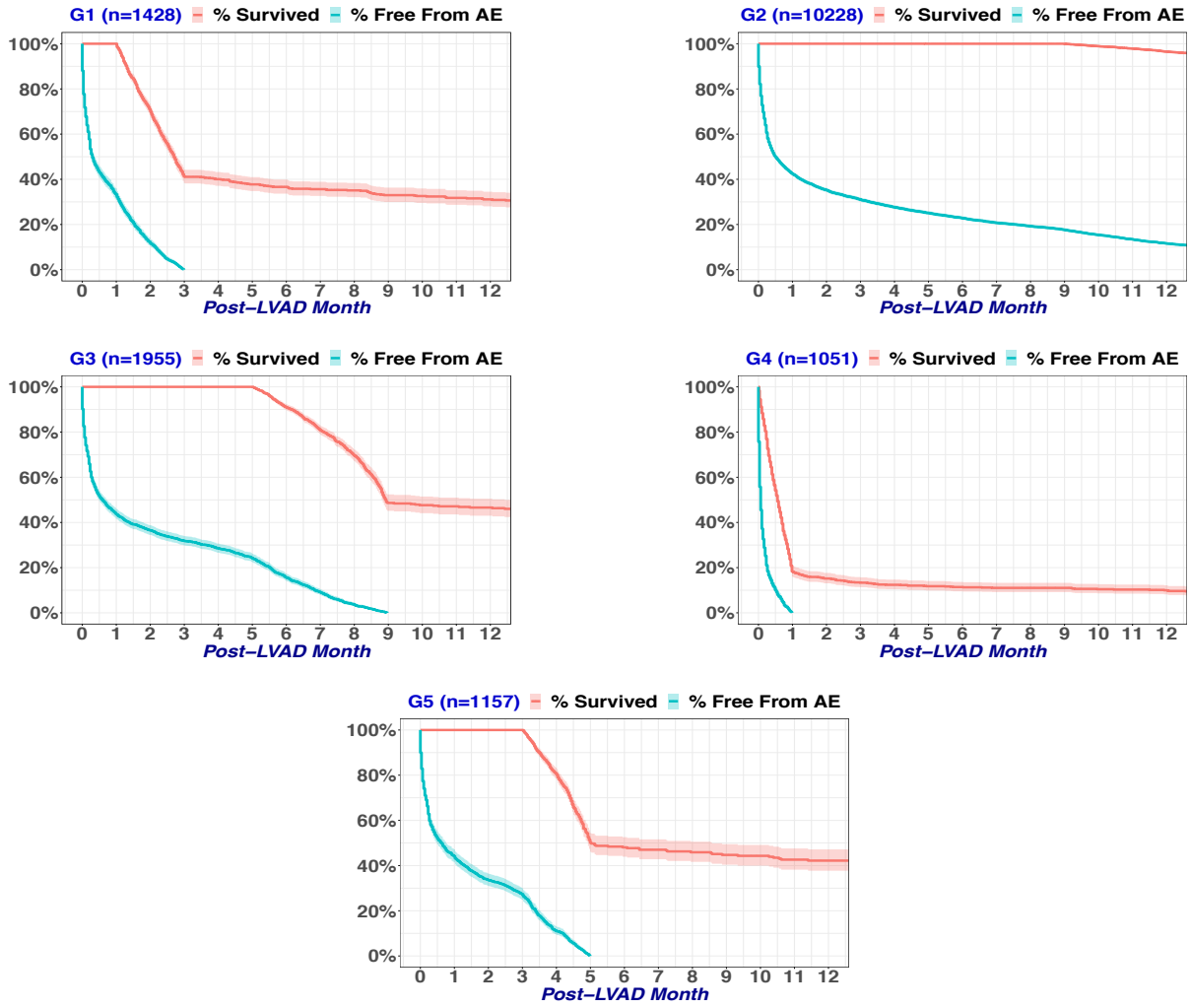
	G1	G2	G3	G4	G5
Count of patients	1428 (9%)	10229 (65%)	1955 (12%)	1051 (7%)	1157 (7%)
Count of episodes of AEs	2872 (8%)	21509 (63%)	5535 (16%)	1322 (4%)	2766 (8%)
Count of AEs	7362 (12%)	34374 (56%)	9521 (16%)	4757 (8%)	5366 (9%)
Mean count of AEs per patient	5	4	5	5	5
Mean of time span (month) per patient	1.53	6.08	5.69	0.41	3.1

(a) Cutting the dendrogram into the five clusters.



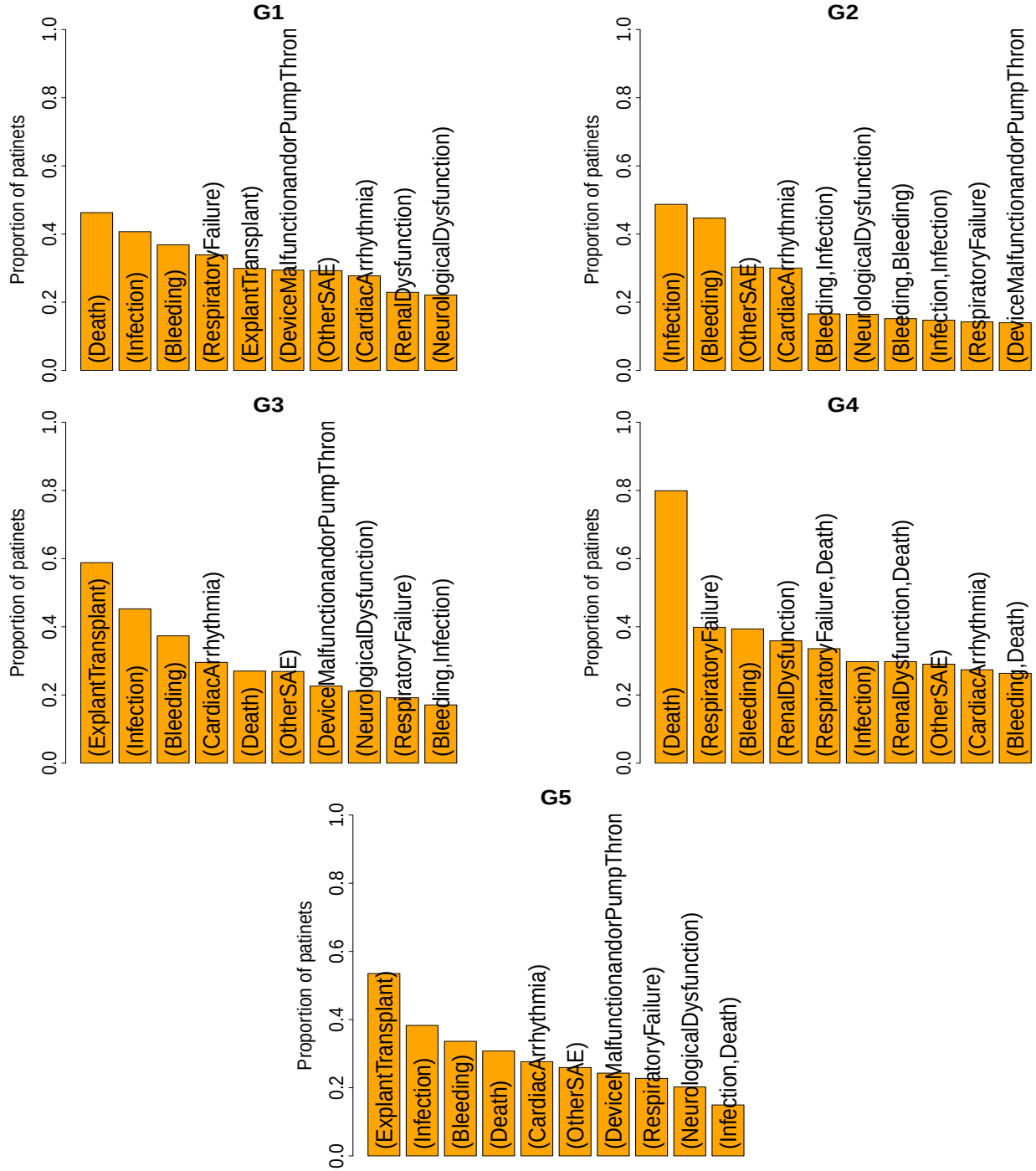
(b) Time distribution of AEs in 5-cluster results (G1,..,G5).

Figure 53: 5-cluster result for monthly-segment.



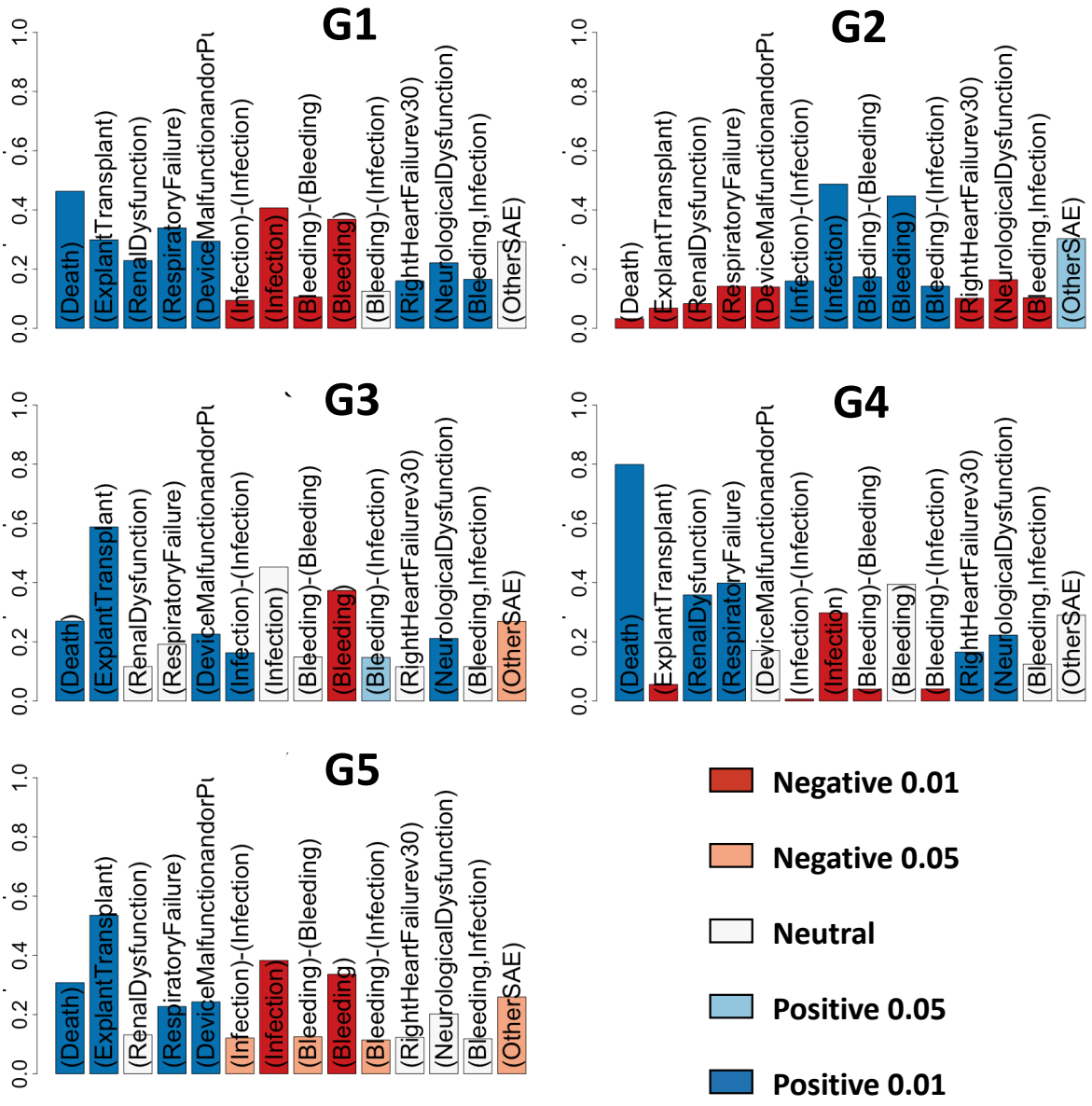
(c) Survival curves and freedom from AE curves for patients in 5-cluster results (G1,..,G5).

Figure 53: 5-cluster result for monthly-segment.



(d) The most common subsequences in 5-cluster results (G1,..,G5).

Figure 53: 5-cluster result for monthly-segment.



(e) The discriminative subsequences between groups. Ordered according to Chi-square value (decreasing left to right.) Color coded based on the sign and value of Pearson's residual.

Figure 53: 5-cluster result for monthly-segment.

curves among G1, G3, and G5 can be also seen in their survival curves (red) that indicate the probability of survival remains at 100% up to 1^{st}, 5^{th} , and 3^{rd} months, respectively, and then sharply drops to 37.5% for G1, 50% for both G3 and G5, and remains constant to the end of the first year. The best survival curve belongs to patients in G2 in which the survival probability remains nearly 100% until the end of the first year. On the other hand, the worst survival curve is associated with patients in G4 whose survival probability drops to $< 20\%$ in the first month.

The plots in Fig. 53d indicate that the most common subsequence in each cluster, except G2, is (Death) or (Explant:Transplant) with proportion of patients in the subsequence ranging from 40% to 80%. Similarly, the most discriminative subsequences (Fig. 53e) is (Death) and/or (Explant:Transplant). The remaining 12 discriminative subsequences are distributed differently among groups.

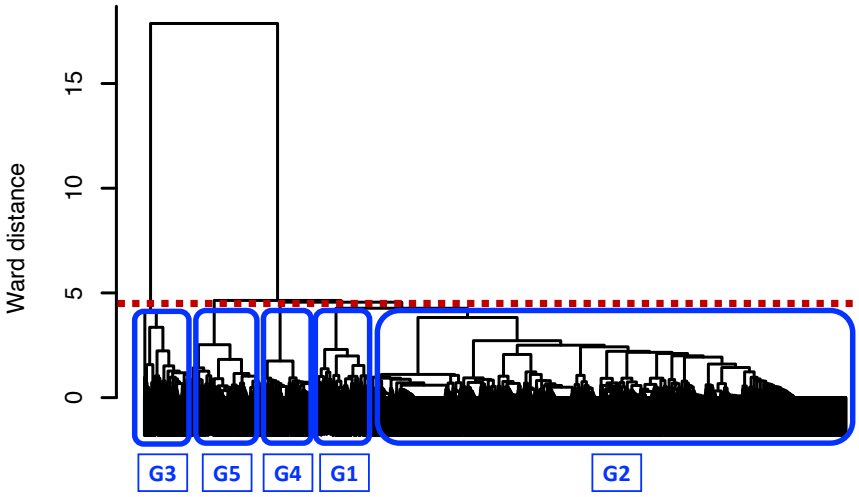
Hierarchical clustering for 3-segment time granularity is found by cutting the previous dendrogram presented in Fig. 52a) at a lower value of Ward distance, resulting in the five groups shown in Fig. 54a. The largest number of patients is found in G2 (n=10,741; 68%) which shares 8,266 patients in common with G2 from the 5-cluster result of monthly-segment time granularity. The remaining groups each include between 7% to 10% of the total number of patients. G3 in this clustering result comprises the same 1,051 patients as G4 in the previous monthly-segment clustering. The Rand index, which measures agreement between groupings, ranging from 1.0 (perfect agreement) to 0.0 (no agreement), was computed to be 0.65 for the 3-segment and monthly-segment clustering. Considering that most of this agreement is derived from the shared 8,266 patients between G2 in two clustering results and the 1,051 shared patients in G3 in 3-segment with G4 in monthly segment, it can be interpreted that the remaining groups have a poor agreement. The detailed characteristics of these five groups are summarized in Table 10, and figures 54a-54e.

Table 10: Summary of 5-cluster result for 3-segment clustering.

Freedom from AE curve (FF-AE) (See Fig. 54b and 54c)	Survival curve (SURV) (See Fig. 54c)	Frequent subsequence (*) (See Fig 54d)	Discriminitive subsequence (**) (See Fig 54e)
Group 1: Moderate survival rate despite high number of AEs in early months after implant			
FF-AE goes to 0% by the 1st month 56% of all AEs occurred within 2nd month	SURV remaining near 100% up to 1st month gradually decreasing to 60% at 12 months	79% (Respiratory Failure) 68% (Infection) 60% (Bleeding)	(Death) (Respiratory Failure) (Neurological Dysfunction) (Infection, Respiratory Failure) (Renal Dysfunction)
Group 2: High survival and high diversity among AEs journeys			
FF-AE goes to 10% by the 1st year Timing of AEs ranges 0 to 12 months 58% of AEs occurring by the 4th month	SURV remaining near 100% up to 1st year	High diversity among AEs journeys ≈ 40% (Infection) or (Bleeding)	Mostly absence of discriminative subsequences in other groups & (Explant: Transplant)
Group 3: Low survival with at least one AE by the first month			
FF-AE goes to 0% by the 1st month Timing of AEs are mostly in the first month and/or at the time of implant	SURV drops to ≈< 20% by the 1st month	80% (Death) 34% (Respiratory Failure, Death) 30% (Renal Dysfunction, Death)	(Death) (Respiratory Failure) (Neurological Dysfunction) (Infection, Respiratory Failure) (Renal Dysfunction)
Group 4: High survival with Device Malfunction and/or Pump Thrombosis			
FF-AE goes to 0% by the 1st year Timing of AEs are spread over 12 months 57% AEs occurring by the 4th month	SURV decreasing gradually to 87% by 1st year	96% (Device Malfunction and/or Pump Thrombosis) 54% (Device Malfunction and/or Pump Thrombosis, Explant:Exchange)	(Device Malfunction and/or Pump Thrombosis)
Group 5: Low survival with common AEs of Neurological Dysfunction and Respiratory Failure			
FF-AE goes to 0% by the 1st year Timing of AEs spans the full 12 months 57% occurring within the first 4th month	SURV is relatively low No deaths in the 1st month 25% by the end of the 1st year	73% (Death) 61% (Neurological Dysfunction) 54%(Infection) 36%(Respiratory Failure) 33% (Neurological Dysfunction,Death) 28% (Respiratory Failure,Death)	(Death) (Respiratory Failure) (Neurological Dysfunction) (Infection, Respiratory Failure) (Renal Dysfunction)

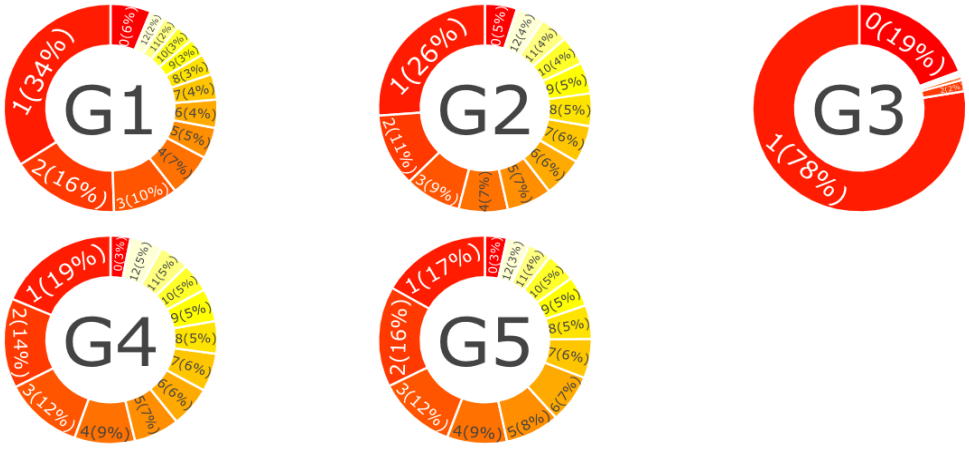
* Selected most common subsequences. See Fig. 54d for the full list of top ten most common subsequnces in each group.

** Selected most discriminitive subsequences. See Fig. 54e for the full list of discriminitive subsequnces in each group.



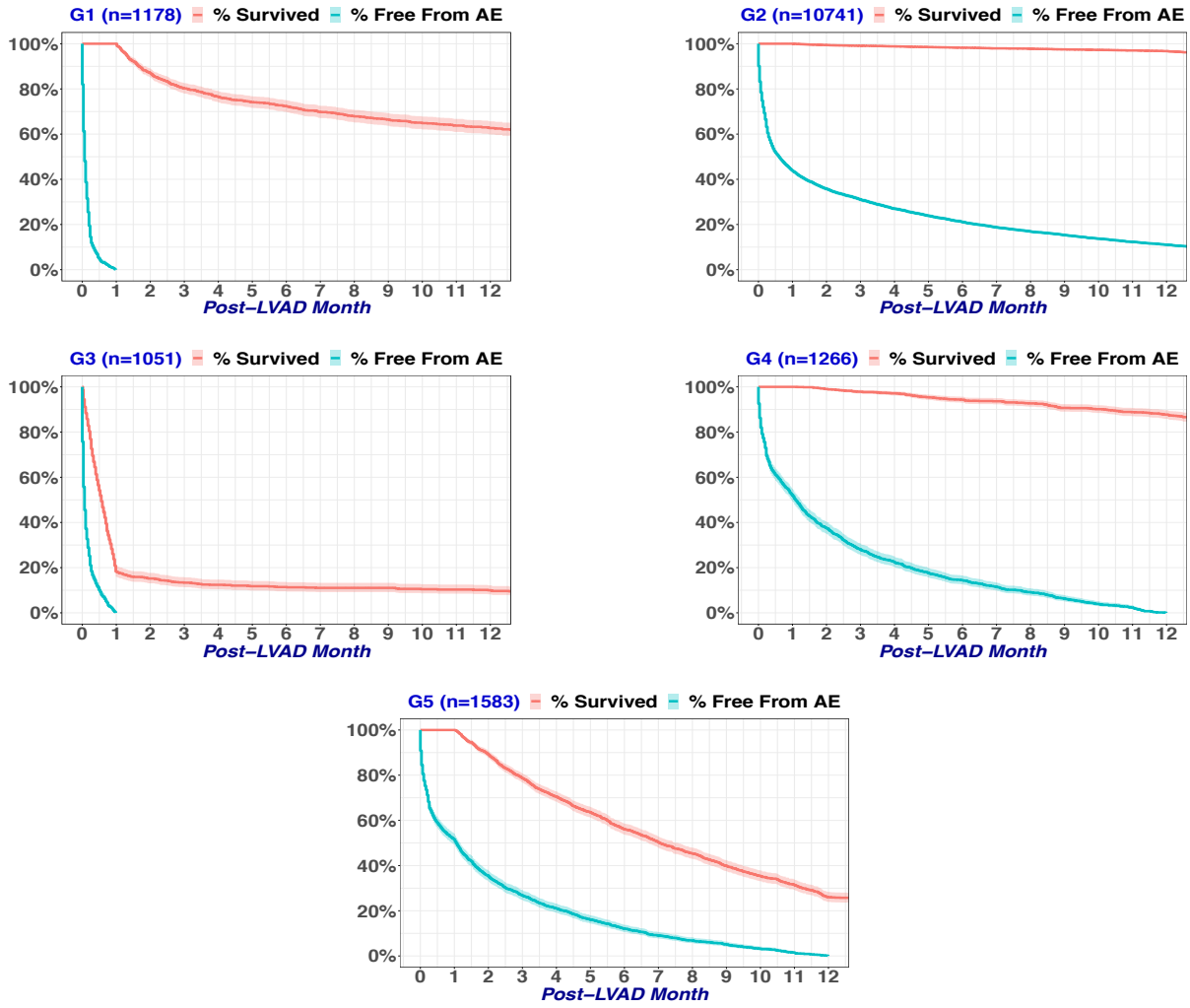
	G1	G2	G3	G4	G5
Count of patients	1179 (7%)	10741 (68%)	1051 (7%)	1266 (8%)	1583 (10%)
Count of episodes of AEs	2310 (10%)	14625 (65%)	1308 (6%)	1936 (9%)	2451 (11%)
Count of AEs	9411 (15%)	32652 (53%)	4757 (8%)	5609 (9%)	8951 (15%)
Mean count of AEs per patient	7.98	3.42	4.53	4.43	5.65
Mean of time span (month) per patient	4.9	5.33	0.41	4.78	4.94

(a) Cutting the dendrogram into the five clusters (G1,...G5).



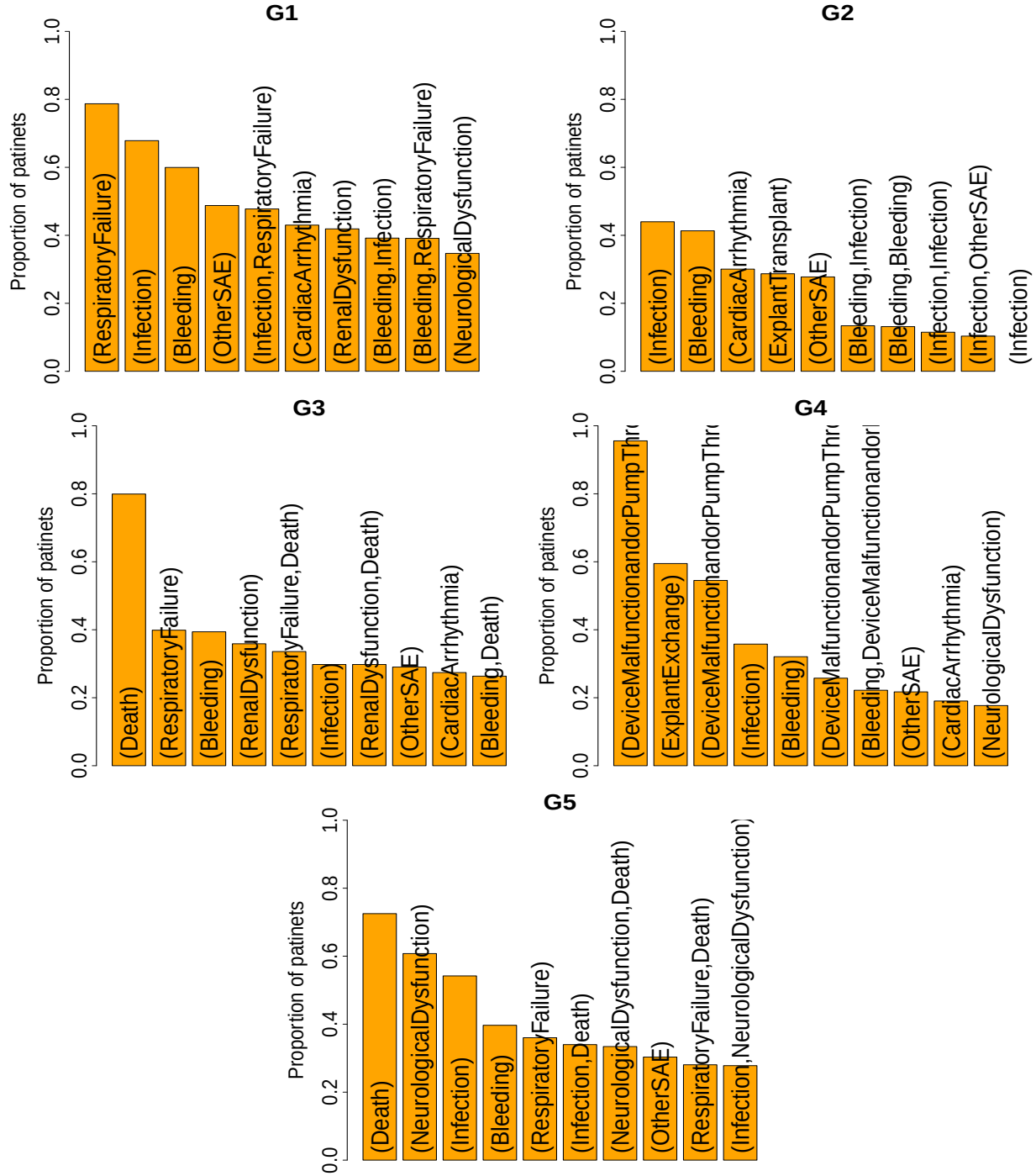
(b) Time distribution of AEs in 5-cluster results (G1,..,G5).

Figure 54: 5-cluster result for 3-segment time granularity.



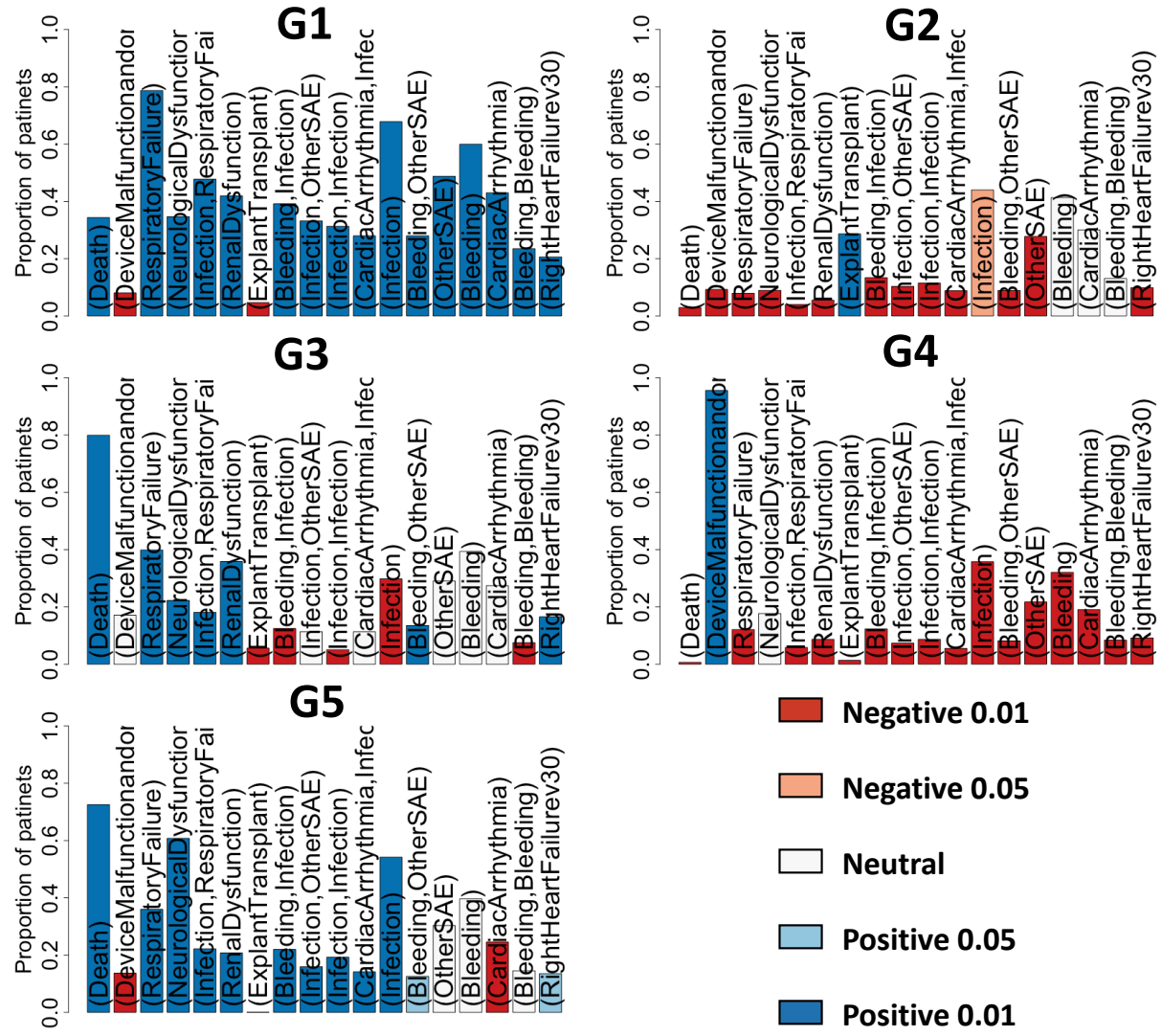
(c) Survival curves and freedom from AE curves for patients in 5-cluster results (G1,..,G5).

Figure 54: 5-cluster result for 3-segment time granularity.



(d) The most common subsequences in 5-cluster results (G1,..,G5).

Figure 54: 5-cluster result for 3-segment time granularity.



(e) The discriminative subsequences between groups. Ordered according to Chi-square value (decreasing left to right.) Color coded based on the sign and value of Pearson's residual.

Figure 54: 5-cluster result for 3-segment time granularity.

6.2.3.3 Modeling the Transition of Adverse Events within Clustered Patient

After identifying clusters of patients who have experienced similar AEs in a consistent order and timing, the transitions between AEs are now modeled using a hidden Markov model (HMM). The patients are split into train and test datasets, with the same proportion of 5 clusters. One HMM is trained for each cluster using the training dataset specific to that cluster.

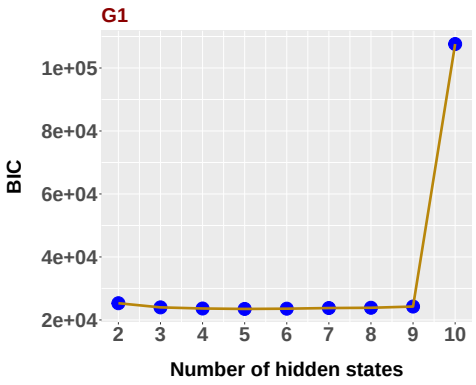
Figure 55 illustrates the Bayesian Information Criterion (BIC) values for HMM models with 2–10 hidden states in each of the five clusters (G1 to G5). The model with the lowest BIC value is chosen as the best model. The HMM models with 5 hidden states yielded the lowest BIC values for G1, G3, and G4, while the HMM model with 6 hidden states resulted in the lowest BIC values for G2 and G5.

Fig. 56 illustrates the HMM structure for each of the 5 clusters with 3-segmentation time granularity. It shows the HMMs as directed graphs, where the nodes represent hidden states. The arrows indicate transition probabilities between the hidden states (filtered to greater or equal to 0.1) , with thicker arrows representing higher probabilities.

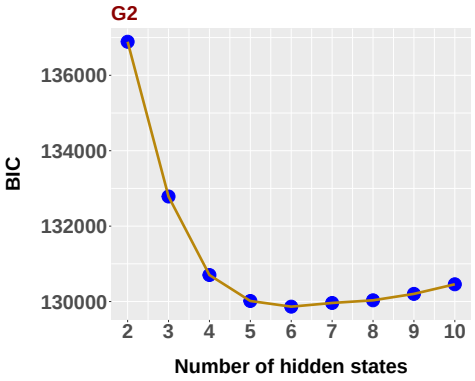
Fig. 57 illustrates the most probable hidden state paths for patients in each cluster. Similar colors are assigned to similar hidden states across clusters.

As an example of how to interpret these figures, let's examine G4, which is named "High survival with device malfunction and/or pump thrombosis' (Table 10). The HMM for G4 in Fig. 56d reveals that State 1 and State 3 have the highest probabilities of Device Malfunction, both at 0.9, and exchange of device (Explant:Exchange), with probabilities of 0.7 for State 1 and 0.5 for State 3. Thus, we can refer to States 1 and 3 as the "Device malfunction and/or Explant: Exchange' states. The other three states have either zero or low probabilities of Device Malfunction and Explant: Exchange but transition to States 1 and 3 with high probabilities ranging from 0.7 to 1.0. These states represent different paths of AEs that lead to Device Malfunction and/or Explant: Exchange.

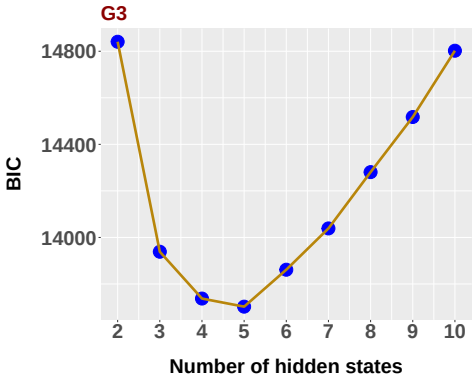
The initial probabilities of the states indicate that patients begin their journey from State 2 with no AEs (at least those listed in Table 56d). Around 80% of them then transition to State 4 with high probabilities of Bleeding and Infection, and subsequently, 90% of them



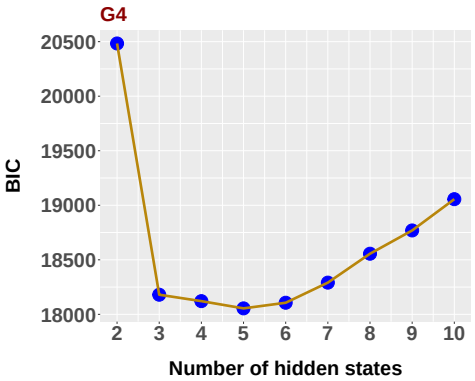
(a) Cluster 1 (G1)



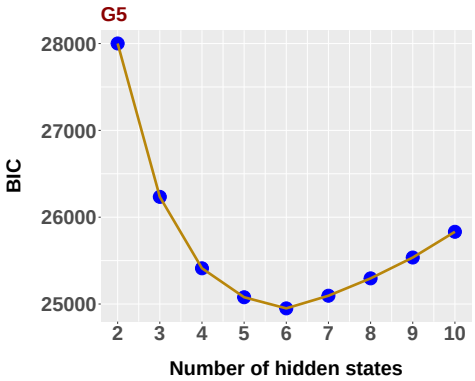
(b) Cluster 2 (G2)



(c) Cluster 3 (G3)



(d) Cluster 4 (G4)



(e) Cluster 5 (G5)

Figure 55: BIC values of 2-10 number of hidden states for 5-cluster for 3-segment time granularity.

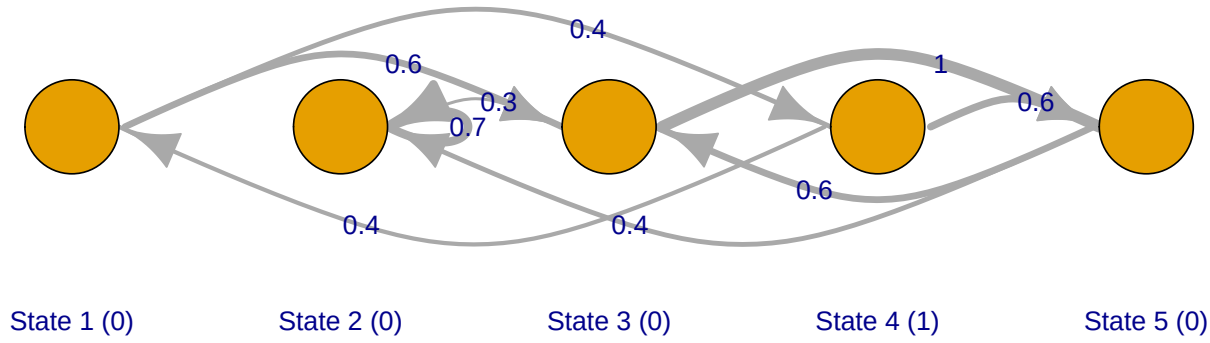
transition to State 3, from which all of them proceed to State 1. This pattern of patient journeys in G4 is also reflected in the most probable path of hidden states shown in Fig. 57. It demonstrates that the most probable state for patients in G4 at the time of implant is State 2, which has zero probability of any AEs. During the first post-implant month, the majority of patients are in States 4 and 5. After the first month until the end of the first year, the most probable states for patients in G4 are States 1 and 3, which have high probabilities of Device Malfunction and/or Explant: Exchange.

The evaluation of HMMs is based on the likelihood of the test dataset. Figure 58 displays box plots of log-likelihood values for patients in the test dataset, corresponding to each trained HMM per cluster. The box plots for HMM2, HMM3, HMM4, and HMM5 indicate higher log-likelihood values for patients in their respective clusters (1, 3, 4, and 5) compared to patients in other clusters. However, the box plots for HMM1 show an overlap between the boxes of patients in clusters 1 and 2.

Patients in the test dataset are assigned to the cluster (predicted cluster) corresponding to their maximum log-likelihood value among the HMMs of the clusters. Subsequently, the predicted clusters are compared to the real clusters to which they were assigned during the cluster analysis, resulting in the formation of a truth table as shown in Table 11. The majority of patients in the clusters are correctly assigned to their respective clusters, with an accuracy rate ranging from 90% to 97% for all clusters, except for patients in cluster 2, where the accuracy is 85%.

6.2.4 Discussion

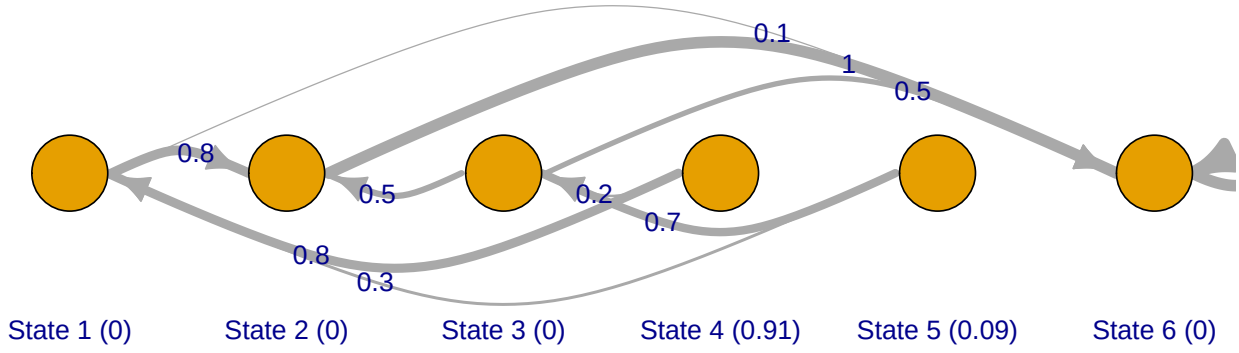
This study conducted an extensive clustering analysis of AEs in the first year after LVAD implant for a large cohort of 15,820 patients from the INTERMACS national registry. Two clustering analyses were conducted to compare the effect of choosing a relaxed (monthly-segment) vs a compact (3-segment) time granularity. The selection of the number of clusters and the goodness of clustering were evaluated based on two criteria: (1) *stability* that measures the consistency of a clustering result by comparing it with the clusters obtained from different bootstrap resampling of the full data, and (2) *silhouette coefficient* that assess the



Hidden States	Bleeding	Cardiac Arrhythmia	Infection	Neurological Dysfunction	Other SAE	Renal Dysfunction	Respiratory Failure	Right Heart Failure	Explant Transplant	Death
State 1	0.2	0.2	0.2	0.4	0.2	0.2	0.6	0.0	0.0	0.0
State 2	0.7	0.4	0.8	0.2	0.6	0.2	0.3	0.1	0.0	0.5
State 3	0.3	0.1	0.3	0.1	0.2	0.0	0.0	0.0	0.1	0.4
State 4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
State 5	0.6	0.5	0.7	0.2	0.5	0.5	0.9	0.3	0.0	0.0

(a) HMM for G1

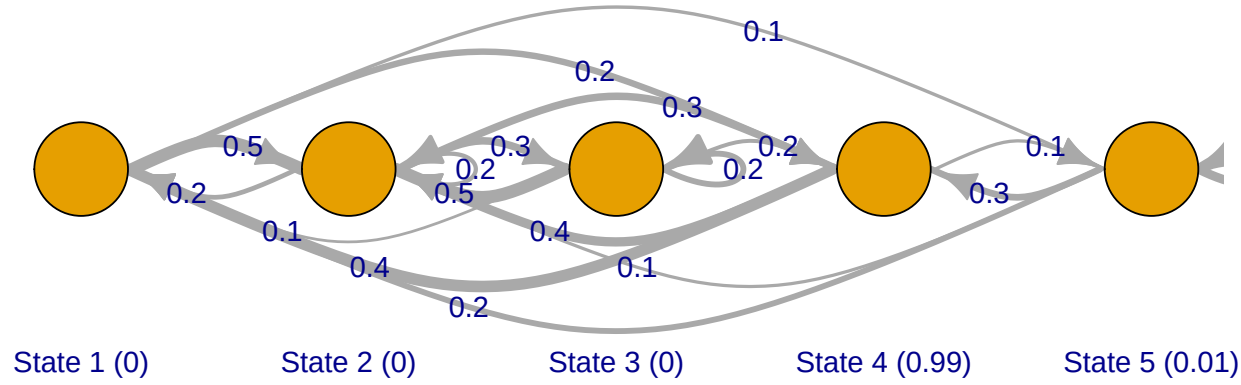
Figure 56: The plots of the HMMs for the 5 clusters are accompanied by a table displaying the emission probabilities for hidden states associated with selected AEs. The initial probabilities for each hidden state are provided in parentheses next to the states in the plot.



Hidden States	Bleeding	Cardiac Arrhythmia	Infection	Neurological Dysfunction	Other SAE	Renal Dysfunction	Respiratory Failure	Right Heart Failure	Explant Transplant	Death
State 1	0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0
State 2	0.2	0.1	0.2	0.0	0.1	0.0	0.0	0.0	0.3	0.0
State 3	0.3	0.3	0.4	0.1	0.3	0.1	0.2	0.2	0.0	0.0
State 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
State 5	0.5	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
State 6	0.5	0.2	0.6	0.2	0.4	0.1	0.1	0.1	0.1	0.1

(b) HMM for G2

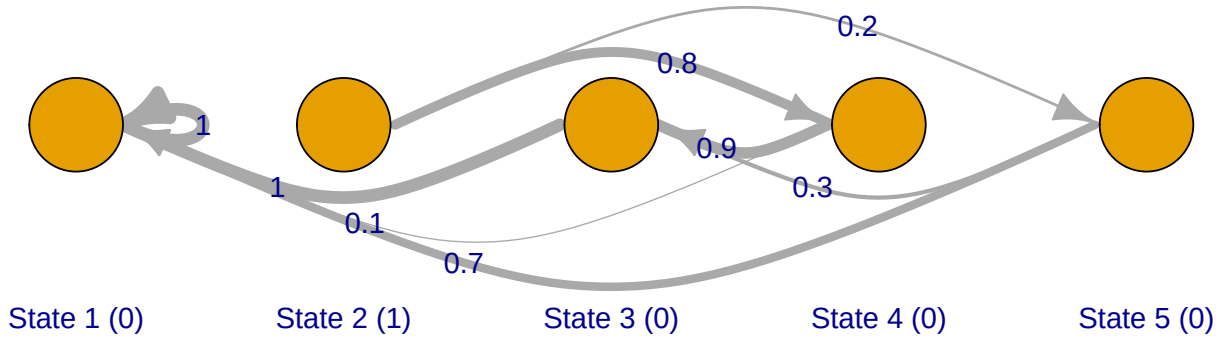
Figure 56: Plots of the HMMs for 5 clusters, accompanied by a table of emission probabilities for hidden states associated with selected AEs. The initial probabilities for each hidden state are provided in parentheses next to the states in the plot.



Hidden States	Bleeding	Cardiac Arrhythmia	Infection	Neurological Dysfunction	Other SAE	Renal Dysfunction	Respiratory Failure	Right Heart Failure	Explant Transplant	Death
State 1	0.5	0.4	0.5	0.2	0.4	0.7	0.7	0.2	0.0	1.0
State 2	0.2	0.1	0.1	0.3	0.1	0.1	0.1	0.0	0.0	1.0
State 3	0.1	0.3	0.1	0.0	0.1	0.1	0.1	0.1	0.8	0.0
State 4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
State 5	0.1	0.2	0.2	0.1	0.2	0.1	0.2	0.1	0.0	0.0

(c) HMM for G3

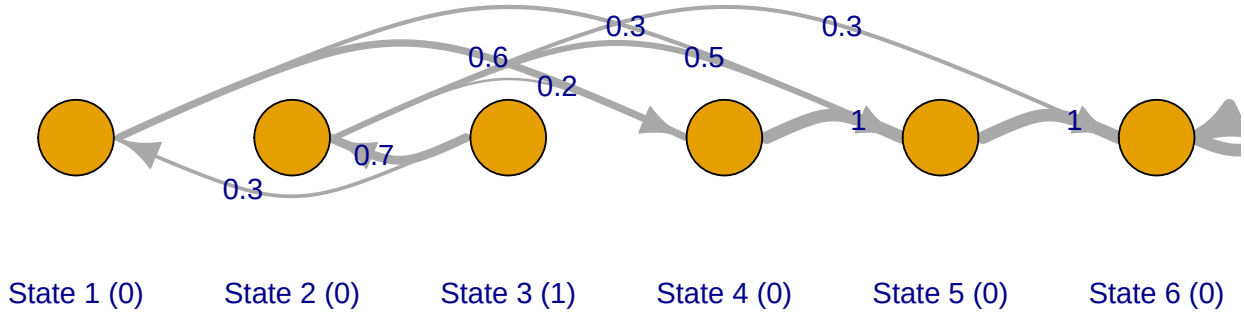
Figure 56: Plots of the HMMs for 5 clusters, accompanied by a table of emission probabilities for hidden states associated with selected AEs. The initial probabilities for each hidden state are provided in parentheses next to the states in the plot.



Hidden States	Bleeding	Cardiac Arrhythmia	Infection	Neurological Dysfunction	Other SAE	Device Malfunction	Explant Exchange	Death
State 1	0.5	0.1	0.5	0.2	0.3	0.9	0.7	0.0
State 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
State 3	0.2	0.1	0.2	0.1	0.1	0.9	0.5	0.0
State 4	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0
State 5	0.3	0.3	0.4	0.1	0.3	0.1	0.0	0.0

(d) HMM for G4

Figure 56: Plots of the HMMs for 5 clusters, accompanied by a table of emission probabilities for hidden states associated with selected AEs. The initial probabilities for each hidden state are provided in parentheses next to the states in the plot.



Hidden States	Bleeding	Cardiac Arrhythmia	Infection	Neurological Dysfunction	Other SAE	Renal Dysfunction	Respiratory Failure	Right Heart Failure	Death
State 1	0.4	0.3	0.4	0.1	0.3	0.1	0.3	0.1	0.0
State 2	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
State 3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
State 4	0.5	0.3	0.7	0.4	0.4	0.5	0.7	0.3	0.9
State 5	0.2	0.1	0.3	0.5	0.2	0.0	0.2	0.0	1.0
State 6	0.2	0.0	0.4	1.0	0.1	0.0	0.0	0.0	0.0

(e) HMM for G5

Figure 56: Plots of the HMMs for 5 clusters, accompanied by a table of emission probabilities for hidden states associated with selected AEs. The initial probabilities for each hidden state are provided in parentheses next to the states in the plot.

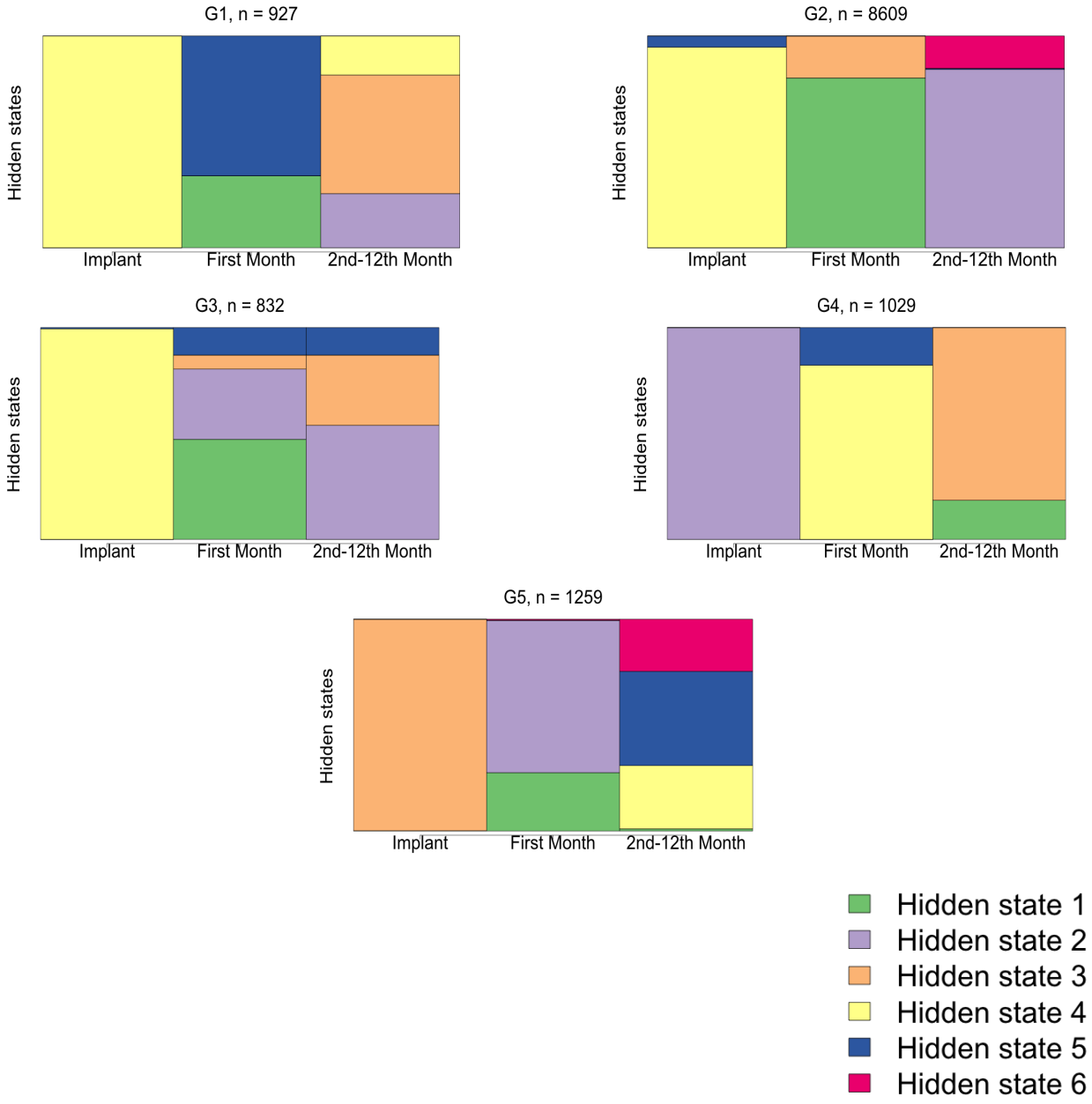


Figure 57: Most probable hidden state paths of patients in the training dataset are observed for each cluster.

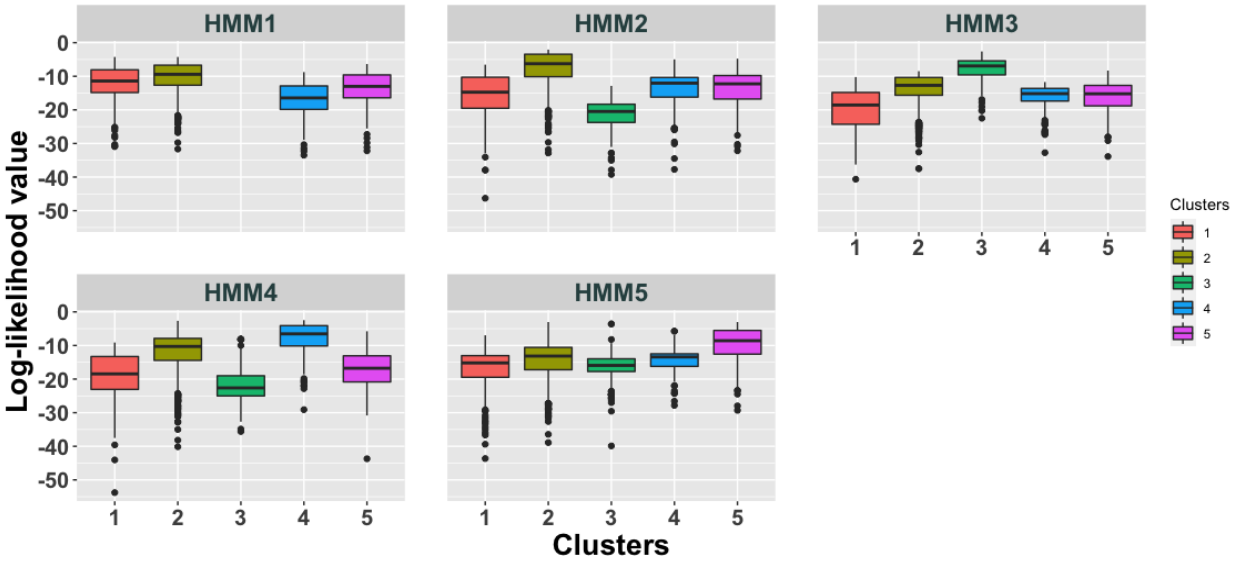


Figure 58: The log-likelihood values of patients in test dataset for each trained HMMs per cluster.

Table 11: The truth table consists of rows representing patients assigned to each cluster based on clustering analysis (real clusters), and columns representing the predicted cluster (Pred-CLS) determined by the maximum likelihood of HMMs.

Real Cluster	Pred-CLS 1	Pred-CLS 2	Pred-CLS 3	Pred-CLS 4	Pred-CLS 5
Cluster 1	229	5	0	2	16
Cluster 2	133	1822	1	85	91
Cluster 3	0	0	212	0	7
Cluster 4	2	3	0	229	3
Cluster 5	20	6	0	7	291

compactness and separation of clusters. Based on these criteria, the monthly-segment time granularity yielded 5-clusters (groups of patients) and the 3-segment time granularity yielded 2-clusters. An additional clustering for 3-segment time granularity, containing 5 clusters, was also included in this study to compare the effect of time granularity independently of the number of clusters.

The 2-cluster solution for 3-segment time granularity divided patients mostly based on the timing of their AEs. The first group (G1) includes the great majority of patients (93%) whose timing of AEs ranges from 0 (the day of implant) to 12 months, whereas the second group (G2) includes a minority of patients with poor survival probability and suffered from AEs, such as Respiratory Failure and Renal Dysfunction, within the first month post-LVAD implant.

The 5-cluster result for the monthly granularity split patients into groups mostly based on their timing of AEs, specifically the timing of final outcomes of Death or Explant: Transplant. However, the influence of timing is attenuated when adopting a 3-segment (vs monthly) time granularity - because all AEs after the 1^{st} month were aggregated into the third element of 3-segment clustering. The consequence of patients becoming more similar in timing is that the type of AEs became more discriminating between clusters. As a result, the 5-cluster result of 3-segment clustering provides more insight about the various AE journeys after LVAD implant compared to the 5-cluster result of monthly-segment clustering which mostly emphasizes the timing of AE journeys. However, this similarity of timing of AEs among the patients in the case of 3-segment time granularity would result in clusters which suffer more from overlapping (weak separation) compared to monthly-segment clustering which has patients well-separated based on their diverse timing of AEs as illustrated in the silhouette plot (Fig. 51). Therefore, clustering analysis of post-LVAD AEs involves a trade-off between uncovering informative patterns and the quality of the clusters. This trade off is governed, in turn, by the choice of time granularity.

There were two groups of patients that appeared in all the clustering results regardless of time granularity and number of clusters. One group comprised a small number of 1,051 patient with low 1-month survival probability who suffered AEs such as Respiratory Failure and Renal Dysfunction in the first month post-LVAD. In contrast, a larger group in 5-

cluster results comprising $65\% - 68\%$ of patients were associated with high 1-year survival probability (nearly 100%). This group was distinguished by rather unequally distributed frequency of final outcomes such as Death and/or Explant: Transplant. Future studies may investigate clustering analysis of post-LVAD AE journeys without considering the final outcomes to explore patterns of AE journeys that focus more closely on quality of life than manner of death. On other hand, if the goal of clustering is primarily to determine similarity or dissimilarity between two patients, then, future studies can assign a higher weight to a specific type of AE when measuring dissimilarity between AE journeys.

The results of this study can be compared with the findings of the previously reported study of *Movahedi et al.* [62] in which seven groups of patients were identified based on the order of AE occurrences without considering their timing. Some similarities can be noted between the groups found in the 5-cluster, 3-segment time granularity outcome. For example, G4, with AEs due to Device Malfunction and/or Pump Thrombosis and G3 with low survival due to Respiratory Failure and/or Renal Dysfunction AEs are common to both studies. Consequently, it can be surmised that timing of AEs do not play an important role in clustering patients in these groups.

Another difference of this study with previous studies is the treatment of combinations of AEs within the same time interval. Previous studies which used a "single-channel" format [62, 204, 205] disregarded combinations of AEs within the same time interval to avoid complexity of the analysis caused by potentially 2^{25} combinations of 25 different AE types. The current study addressed this challenge by employing a "multi-channel" format in which each sequence of AEs was segregated into multiple sequences (channels) per patient according to the type of AE. The important difference between these two formats is illustrated by considering two hypothetical patients, P_1 and P_2, P_1 who experienced Bleeding and Infection in the first month and P_2 experienced only a Bleeding episode in the first month. With the "single-channel" format, the combination of Bleeding and Infection would make a unique element in the P_2's sequence – different from Bleeding in P_1's sequence. While in the multi-channel format in which the dissimilarity between two patients is computed by summation of dissimilarities of patients for each type of AE, the Bleeding AE that is common to both patients will be reflected in their dissimilarity measurement – as explained in the example

in Fig. 49. This permits a large number of AE types to be easily combined according to the AE type-specific dissimilarities, without losing information on their complexity.

The temporal variability and treatment-specific dependence of the informational yield of diverse channels of AEs are relevant considerations in this study. Specifically, the relative informativeness of distinct channels of AEs may vary over time, and across different stages of post-LVAD treatment. This necessitates a discerning approach whereby, if a particular type of AE demonstrates superior information content at a given post-LVAD time-point, greater reliance must be placed on such informative channels. Consequently, future investigations may be well-served by focusing on the development of a weighting scheme, which could encompass an evaluation of the costs associated with substitution, while computing the dissimilarity between channels. Future investigations will also benefit from the availability of the most current version of the INTERMACS dataset that pertains to the latest technology of LVAD devices, namely the HeartMate 3 (Abbott) which has exhibited markedly improved survival and reduced incidence of complications.

6.2.5 Conclusion

In conclusion, the multi-channel approach employed in this study for mining clusters of adverse event (AE) sequences from a large diverse National registry proved to be an effective method to extract patterns of AE journeys despite the complexity created by the numerous types of AEs as well as the irregularity in the timing of AEs. The more compact temporal segmentation proved to be reasonable choice when the timing of events are skewed and when the emphasis is more on identifying patterns of types of events versus their timing. Future research based on the findings of this study could lead to improved understanding and identification of patterns of AEs in VAD patients that may inform improved clinical decision-making - potentially leading to tailored interventions at specific time points to prevent or mitigate future AEs.

7.0 Post-LVAD Mortality Risk Models

7.1 Objective

This study incorporated AE history and pre- and post-implantation data to predict mortality in the early weeks following LVAD implantation, which is the most critical period associated with the highest rate of death and all types of adverse events. Three models were investigated to predict death within specific time frames during the first post-LVAD month using data from a large International Registry for Mechanical Circulatory Support (INTERMACS), which includes records for over 20,000 patients who underwent LVAD implantation across more than 180 hospitals. By considering the post-LVAD AE history and data from the first-week follow-up visit, including lab values, medication treatment, and hemodynamics, this study aimed to overcome the issues faced by previous studies that did not account for changes in patients' health status after implantation when predicting mortality

7.2 Methods

7.2.1 Framework

The framework for this study is illustrated in Fig. 59, in which Model I focuses on prediction of death during the second to fourth week (W2-W4) based on data obtained 48 hours prior to implant, during first week of follow-up, and history of AEs during the first week; Model II predicts death during the third to forth week (W3-W4) based on data prior to implant, during the first week, and history of AEs in the first and second weeks; and Model III predicts death during the fourth post operative week (W4), based on data prior to implant, during the first week, and the history of AEs during the first three weeks post-implant.

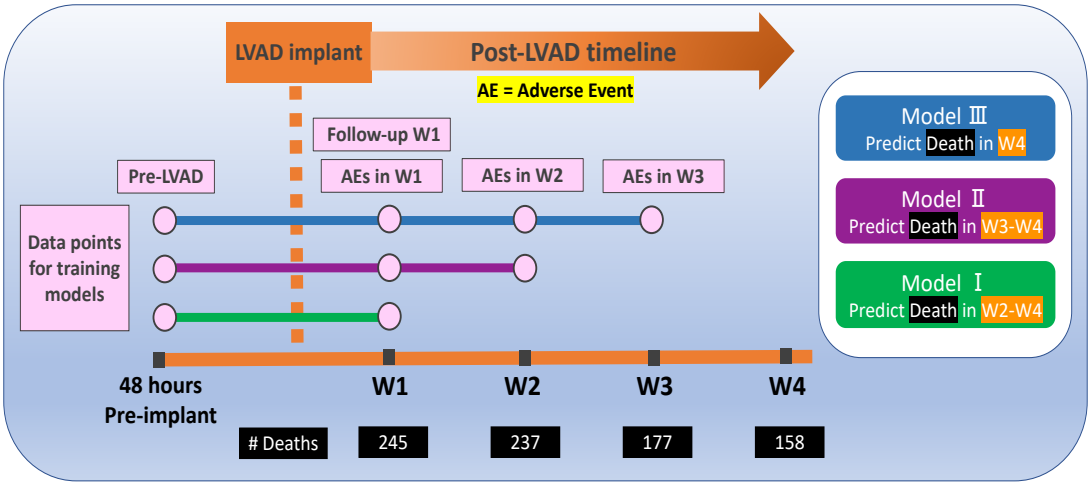


Figure 59: The framework of prediction of mortality in the first few weeks after LVAD implant.

7.2.2 Study population

The first step of the flow diagram in Fig. 60 presents the inclusion criteria used for analyzing 86,912 recorded AEs in 15,820 patients who received a continuous flow-LVAD between 2008 to 2016. These AEs were extracted from the publicly accessible INTERMACS Event dataset via BioLINCC (2017 version). The mean age of the patients was 57, with 12,429 males and 3,378 females. The study excluded the AE labeled "Rehospitalization" as it is considered a consequence of an AE rather than an AE itself and also due to its rarity during the first post operative month. Additionally, AEs that occurred after the first LVAD explant for patients with multiple device implants were also excluded from the analysis.

The patients for each model are filtered according to the time frame of interest for prediction. For example, out of 15,820 patients who survived and still had the pump implanted in the third post-operative week, 15,029 were included in the development and validation of Model III, which predicts deaths during the fourth post-operative week (W4). The percent of dead patients in the time frame of interest for prediction were 4% (572 patients) for Module I, 2% (335 patients) for Module II, and 1% (158 patients) for Module III. The percentage

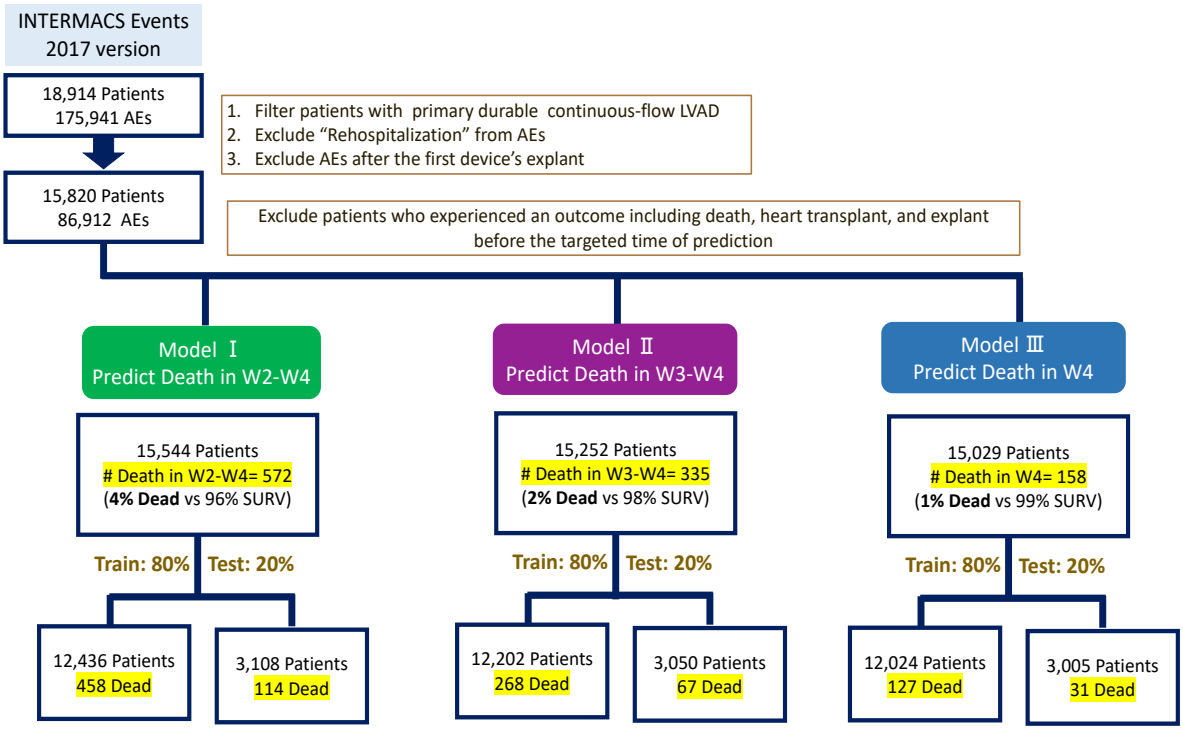


Figure 60: Data selection steps. SURV: Survived (alive).

of dead patients were kept the same after randomly dividing the data for each model into training (80%) and a testing (20%) sets.

7.2.3 Features

The pre-implant and post-implant data contain 240 and 111 features, respectively, after filtering out features with zero variance (28 features for pre-implant data and 239 features for post-implant data) and removing features with more than 30% missingness (171 features for pre-implant data and 128 features for post-implant data). The missingness was measured after filling any non-real missing data by concatenating subcategories with categories and by reviewing the reason for incompleteness in INTERMACS. The remaining missing values in the data were filled with the mode for categorical features and the mean for numeric features. The history of AEs included records of 19 types, and the number of records varied depending on the model. For example, Model III for mortality at week 4 includes bleeding events at the time of implant, and each of the subsequent three weeks.

7.2.4 Feature selection

This study utilized a feature selection method called Recursive Feature Elimination (RFE) [206, 207] on the training set of each model to reduce the > 350 features in INTERMACS to a compact set of the most relevant or important features in predicting death after LVAD implant. The RFE algorithm trains a model on the dataset and removes the least important feature(s) from the current feature set until a desired number of features is reached or the model's performance reaches a certain threshold [208] – in this study the area under the curve (AUC) of receiver operating characteristic (ROC).

7.2.5 Sampling

The data for all three models shows an imbalance in the proportion of dead patients (Dead class) containing only 1% of the data for predicting death in the forth week (versus the SURV class that accounts for 99%.) Therefore, to avoid bias towards predicting the majority

class (SURV) the ROSE (Random Over-Sampling Examples) method was employed to alter the class distribution and obtain a more balanced sample. The ROSE method involves oversampling the minority class by creating synthetic data points that closely resemble the real ones with respect to a probability distribution centered on the selected sample [209]. The ROSE method uses smoothed bootstrapping to randomly select examples from the minority class and create new copies until the number of samples in the minority class matches that of the majority class.

7.2.6 Training predictive models

The study employed a random forest (RF) algorithm to derive and tune three models from the training datasets, using 5 repeated 10-fold cross-validations. The RF algorithm is a widely-used ensemble technique that combines multiple decision trees based on bootstrap samples from the data with random feature selection [160]. Each decision tree then casts a vote for a patient and the majority vote determines the classifier's output.

The boosted decision trees are constructed by computing an importance score for each feature, which is determined by the magnitude of improvement in the performance that is measured at the corresponding split point in a tree, weighted by the number of patients in each node [172]. The overall importance scores for each of the features is then computed as the average of all scores across all decision trees.

7.2.7 Evaluation of performance of predictive models

Due to the severe imbalance in this datasets, special attention must be given to the minority (Dead) class when evaluating the performance of predictive models. Standard evaluation metrics like ROC (Receiver Operating Characteristic) that give equal weight to both majority and minority classes can lead to an overly-optimistic assessment of performance [128]. Therefore, this study also employed Precision-Recall Curve (PRC) to assess the discrimination power of the model. Specific details of ROC and PRC are provided previously in Sections 5.1.2.3 and 5.1.2.4.

7.3 Results

7.3.1 LVAD patients characteristics

Table 12 summarizes the baseline characteristics of 15,820 patients. The majority of the patients were male (78.6%), with an average age of 57 years, and of white ethnicity (67.8%). The predominant device strategies at the time of implant were destination therapy (41.1%) and possible bridge to transplant (30.9%). The severity of illness at the time of implantation was assessed using the New York Heart Association (NYHA) and INTERMACS Patient Profile stratification. Most patients were categorized as "IV: Severe limitations" (75.5%) by NYHA and as "2: Progressive decline" and "3: Stable but inotropic dependent" (68%) by the INTERMACS patient profile.

7.3.2 Survival and adverse events (AEs) of LVAD patients

The overall survival of the entire cohort during the first post-implant period is presented in red in the Fig. 61. The probability of survival to the end of the first month is 94%, and it decreased gradually to approximately 80% by the end of the first year. It is important to note that there is a 4% decrease in survival probability during the first month compared to a 14% decrease over the subsequent 11 months. Therefore, this study placed particular emphasis on predicting mortality in the first month. Multisystem organ failure (30%), neurological dysfunction (15%), withdrawal of support (8%), and right heart failure (7%) were the listed as the primary causes of the 840 counts of death that occurred during the first month. (See Fig. 62.)

The freedom from AE curve (green), indicates that 60% of all patients experienced at least one AE by the end of the first month. Thereafter, the rate incidence rate of AEs decreased, yet resulted in 90% incidence (10% freedom from) AEs by the end of the first year. The Fig. 63 also displays bar plots for the most common types of AEs for each week during the first month. Bleeding and infection were the two most frequent AEs. The first week was associated with the greatest incidence of all types of AEs, except bleeding which was evenly distributed across all weeks.

Table 12: Baseline patient characteristics on continuous-flow LVADs at time of implant

Patient Characteristics	All patients (N=15,820)
Demographic	
Gender	
Female	3,378 (21.4%)
Male	12,442 (78.6%)
Age group	
B: 19-29	656 (4.1%)
C: 30-39	1,127 (7.1%)
D: 40-49	2,210 (14.0%)
E: 50-59	4,343 (27.5%)
F: 60-69	5,353 (33.8%)
G: 70-79	2,028 (12.8%)
H: 80+	103 (0.7%)
Age	57.1 ± 12.7
Race	
Black	3,715 (23.5%)
White	10,730 (67.8%)
Other	1,375 (8.7%)
Indication	
Device strategy	
Bridge to recovery	56 (0.4%)
Bridge to transplant—listed	4,334 (27.4%)
Possible bridge to transplant	4,891 (30.9%)
Destination therapy	6,499 (41.1%)
Rescue therapy	40 (0.3%)
Severity of illness	
New York Heart Association (NYHA)	
I: No limitation	12 (0.1%)
II: Slight limitations	98 (0.6%)
III: Marked limitations	2,777 (17.6%)
IV: Severe limitations	11,977 (75.7%)
Unknown	956 (6.0%)
Patient profile	
1: Critical cardiogenic shock	2,377 (15.0%)
2: Progressive decline	5,838 (36.9%)
3: Stable but inotropic dependent	4,896 (30.9%)
4: Resting symptoms	2,104 (13.3%)
5: Exertion intolerant	371 (2.3%)
6: Exertion limited	152 (1.0%)
7: Advanced NYHA class III	82 (0.5%)

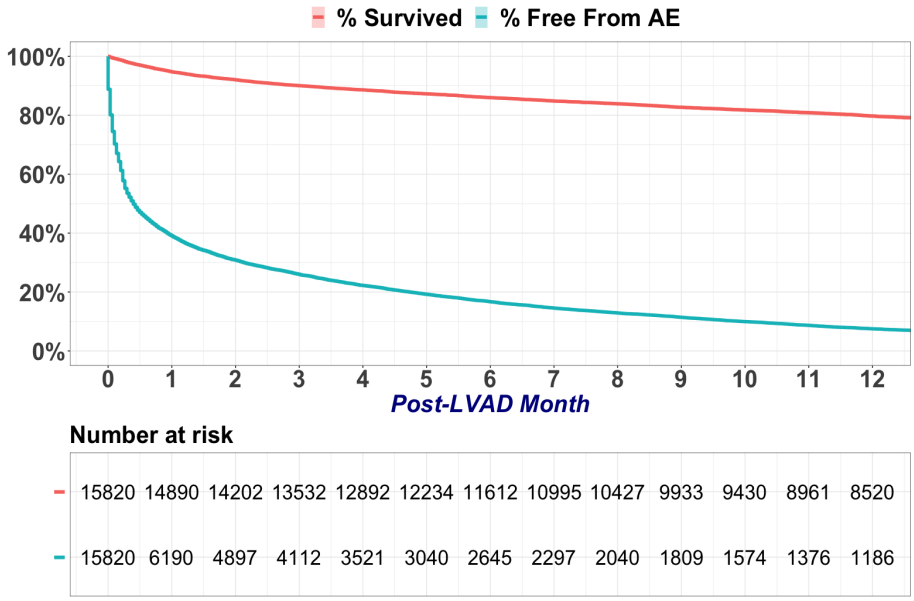


Figure 61: Survival for LVAD patients during the first post-LVAD year.

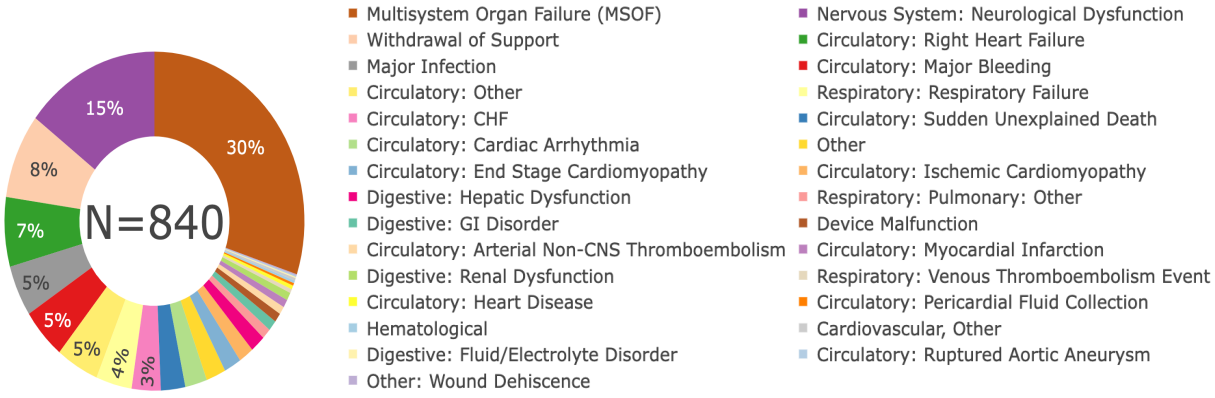


Figure 62: The primary cause of death during the first post-LVAD month.

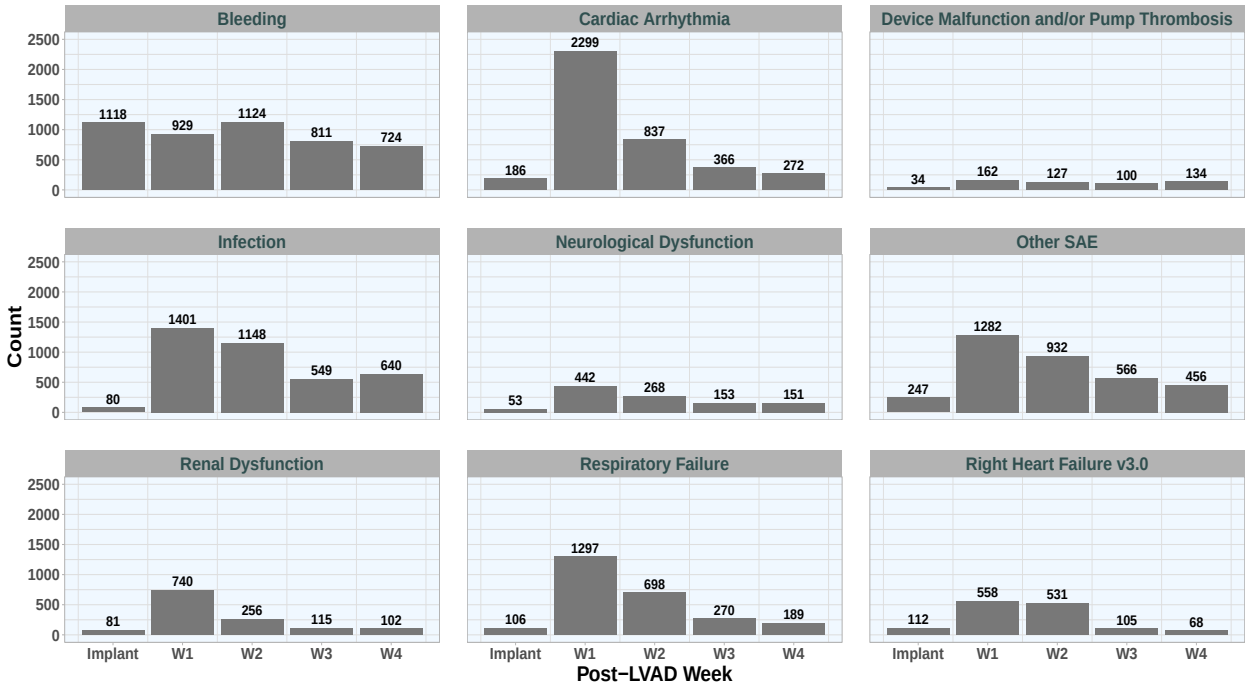


Figure 63: Bar-plots of distribution of common types of AE per week over the first post-LVAD month. Other SAE: Other serious AE.

7.3.3 Feature selection

The number of features fed to RFE for Models I, II, and III, were 376, 394, and 413, respectively. The RFE results indicated that the maximum AUC-ROC can be achieved by including all features; however, a limited subset of features can also achieve an AUC-ROC within 0.1, namely 15, 16, and 16 for developing Models I, II, III, respectively. For Model I and Model II, the RFE algorithm selected age at the time of implant as the sole pre-implant feature. However, in the case of Model III, the RFE algorithm also identified hemoglobin, serum glutamic-oxaloacetic transaminase-aspartate aminotransferase (SGOT-AST), and blood urea nitrogen as relevant features. Considering the history of AEs, respiratory failure, renal dysfunction, and hepatic dysfunctions were selected for all three models. Neurological dysfunction was an additional feature selected for Model III. Several features from first post-LVAD followup were selected for each model such as blood urea nitrogen and platelets that were common among three models. The complete list of features for each model are presented and discussed in the "feature importance" section, below.

7.3.4 Performance of Three Models

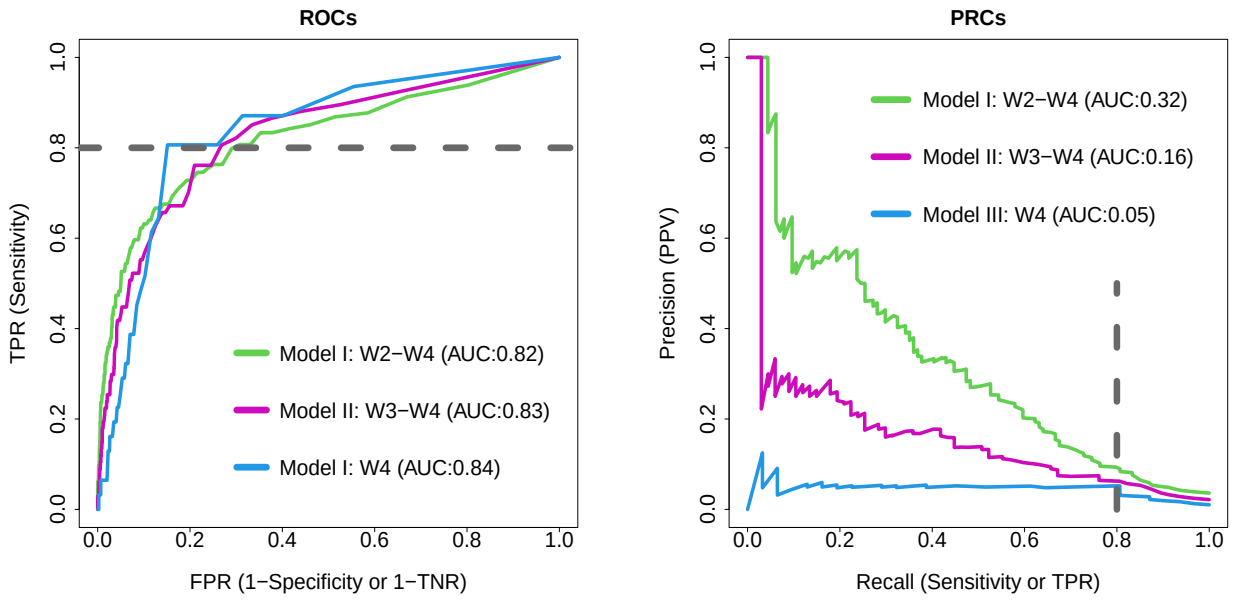
All three models demonstrated promising performance in predicting early post-LVAD mortality in the test datasets, with ROC-AUCs ranging from 0.82 to 0.84, as depicted in Fig. 64a. Their ROC curves exhibited steep initial slopes, indicating high sensitivity and accurate identification of most mortality cases. Additionally, the curves demonstrated a balance between sensitivity and specificity, achieving a sensitivity of 80% to 81% at specificity of 71% to 85% (dashed gray line in Fig. 64a). Fig. 64b shows that the AUC-PRCs ranged from 0.05 to 0.32, with moderate precision of $100\% - 60\%$ for Model I and $100\% - 30\%$ in Model II for a small subset of patients with a high predicted probability of mortality ($> 80\%$) located in the upper left side of the curves. However, the precision drops dramatically for predicted probabilities of mortality less than 80%. The ROCs and PRCs indicate that the three models are highly sensitive and specific, but not precise. This means that selecting a threshold that provides high sensitivity – more accurately predicting true deaths – comes at the cost of low precision – predicting a large number of false deaths. A specific illustration

can be provided by comparing the corresponding confusion matrices, which are tables of true and false predictions for all patients in the test datasets, are shown in Fig. 65 for a chosen sensitivity of 80% to 81% (gray dashed lines in 8a and 8b), and corresponding specificity between 71% and 85% and precision between 5% and 9%. The true predictions (green elements) in the confusion matrix for Model I indicate that 91 out of 114 dead patients were correctly predicted as dead by Model I (corresponding to in a sensitivity of 80%) and 2,160 out of 2,994 survived patients were correctly predicted (corresponding to a specificity of 71%.) However, the false predictions in the confusion matrix indicate that 23 out of 114 dead patients and 834 out of 2,994 survived patients in the test dataset were incorrectly classified, corresponding to a low precision of 9% for the prediction of death.

7.3.5 Feature importance

Fig. 66 shows the ranking of features in each model based on their relative contribution to the prediction of mortality, as determined by the RF algorithm. Among the top six features common to all three models are: age at the time of implant, platelet count, and blood urea nitrogen (BUN) in the first week after implantation (W1). For Models I and II, additional top features at W1 included total bilirubin, white blood cell count, and SGOT-AST. For Model III, other top features were BUN, hemoglobin, and SGOT-AST, all measured before implantation.

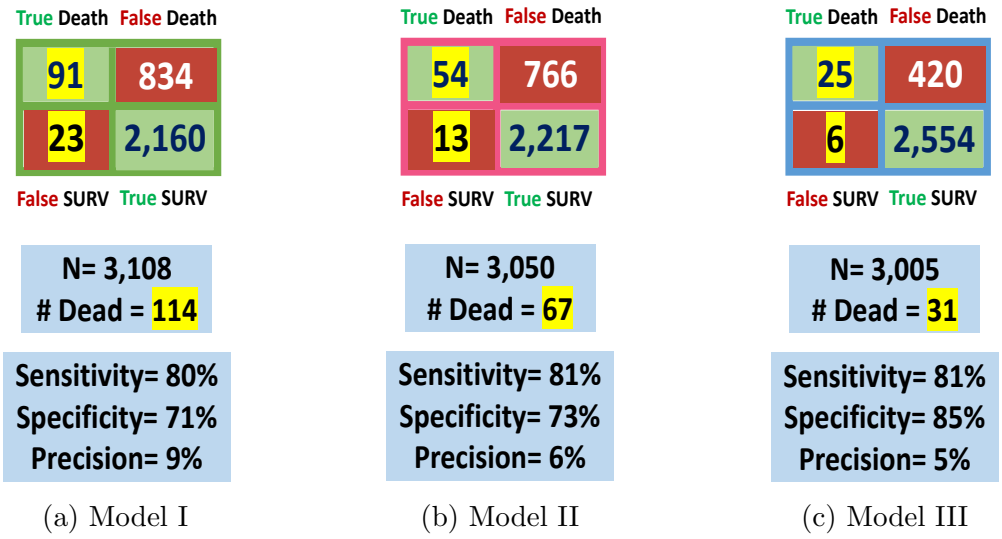
To further investigate the top features for each model, descriptive statistics and box plots are presented in Fig. 67, Fig. 68, and Fig. 69 for the entire dataset (including both the training and testing sets). The figures categorize the features by the outcome (class of prediction: Death vs. SURV) and also show the results of the Wilcoxon test, which tests the null hypothesis that the feature distribution is the same between survived and dead patients. For all three models, all features were significantly different between survived and dead patients, save one: pre-implant SGOT-AST and hemoglobin at W1 for Model III, which also exhibit similar box plots. The box plots of other features for the three models show varying degrees of difference between dead and survived patients. For instance, the distributions of total bilirubin in the datasets for Models I and II are highly right-skewed for



(a) ROCs

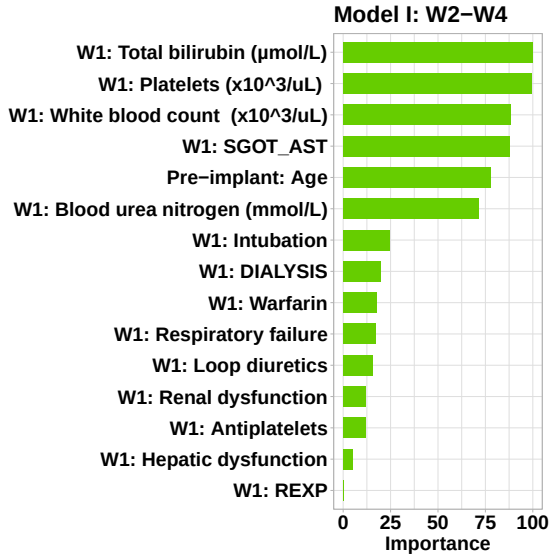
(b) PRCs.

Figure 64: Discrimination power of models in prediction of early death for patients in test datasets evaluated by ROC and PRC.

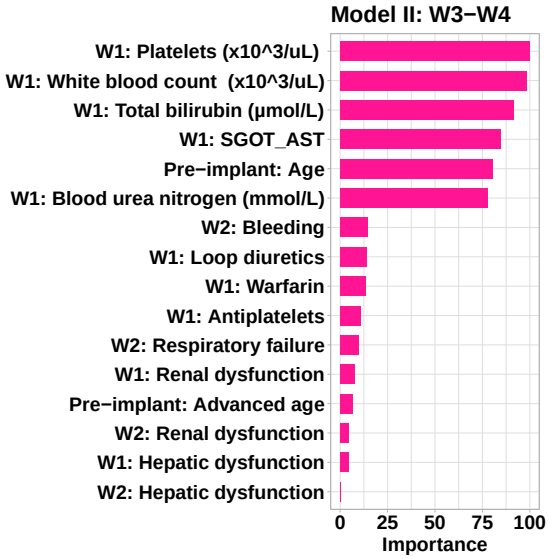


(a) Model I (b) Model II (c) Model III

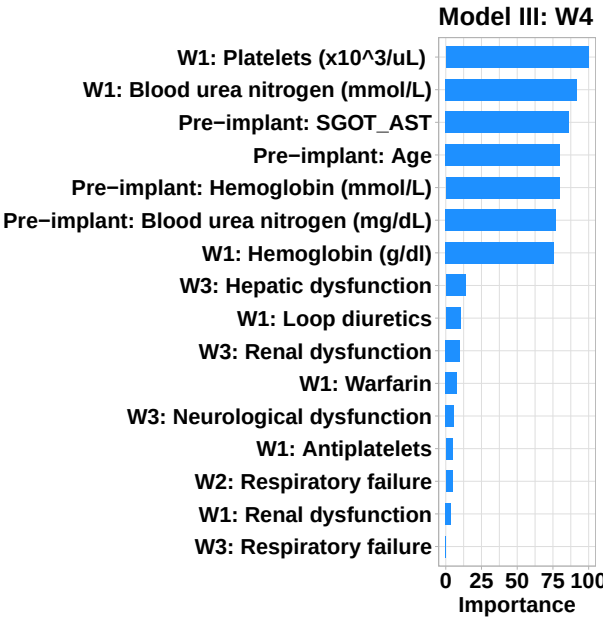
Figure 65: Confusion matrices corresponding to sensitivity of $80-81\%$ (gray dashed lines in ROCs and PRCs in Fig. 64). The green elements are true predictions and the red elements are false predictions. The sum of the elements in confusion matrices are the total number of patients in the test datasets for three models.



(a) Model I. REXP:explant of right ventricular assist device.



(b) Model II



(c) Model III

Figure 66: Plot of the relative importance of features for prediction of death for each of the three models.

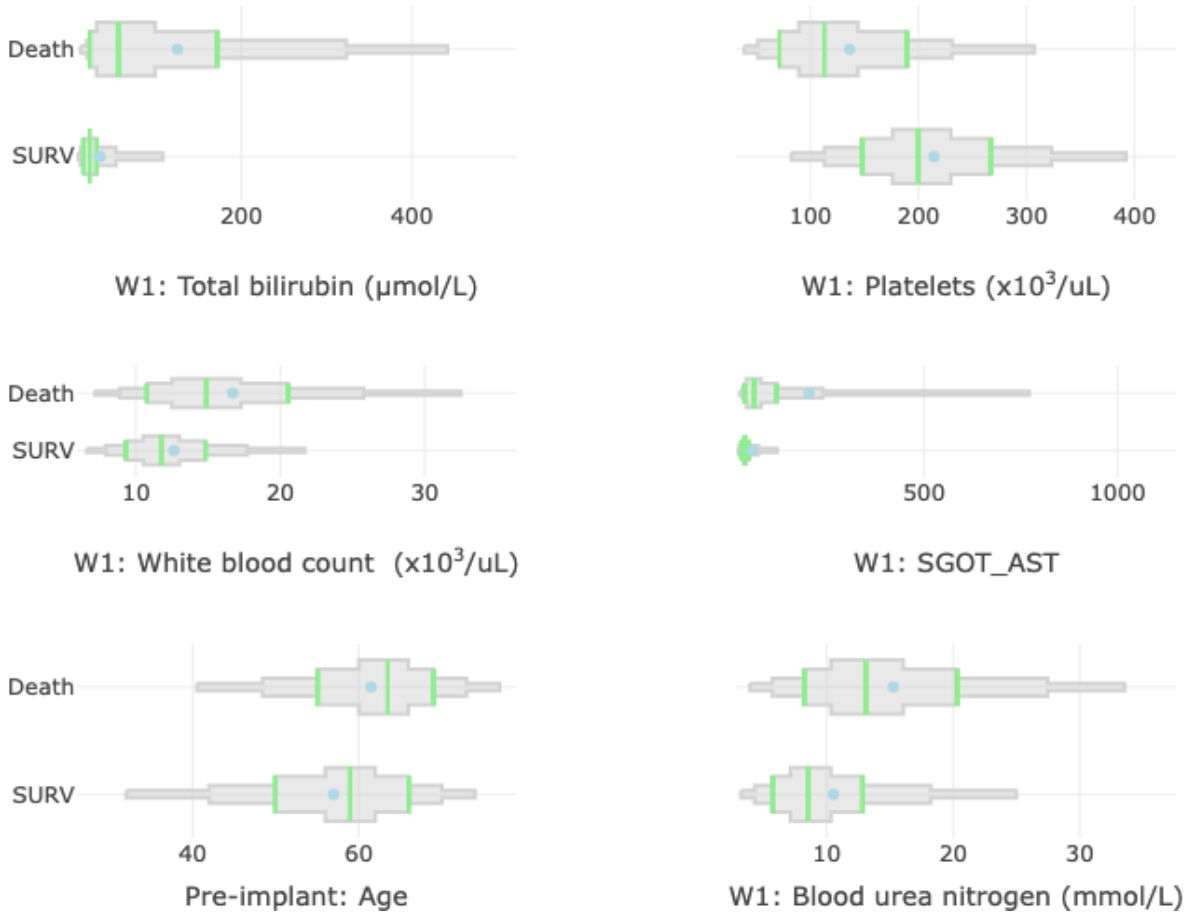
dead patients compared to survived patients, with average total bilirubin levels (µmol/L) of 56 vs. 22 and 46 vs. 22 for dead patients vs. survived patients in Model I and II, respectively. Another example is the average platelet value ($\times 10^3$/L) of 112-121 vs. 200-203 for dead vs. survived patients in three models.

7.4 Discussion

This study was conducted to explore the hypothesis that early mortality after LVAD implantation is affected by both pre-implant risk factors as well as post-implant follow-up data and adverse events. Three models were developed based on the INTERMACS National registry that demonstrated good sensitivity ($80\% - 81\%$) and specificity ($71\% - 85\%$) for predicting mortality within three time frames: second to fourth week (W2-W4), second to third week (W2-W3), and fourth week (W4). The corresponding AUC-ROCs were in the range 0.82-0.84. Consequently these models offer a reliable tool to identify patients at risk of mortality in the early weeks following LVAD implantation, hence guide timely intervention and improved patient outcomes.

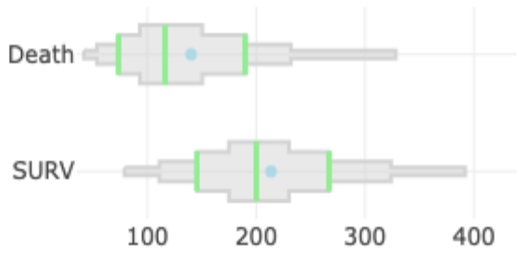
Contrasted with the high sensitivity of the models, their precision was poor, with AUC-PRC ranging from 0.05 to 0.32. This outcome is not surprising in light of the highly imbalanced dataset. It indicates a significant number of deaths falsely predicted by the model. Therefore, patients with predicted mortality require further evaluation through medical diagnostic tests - as the cost of testing is far less than the cost of erroneously predicting survival. Nevertheless, the potential cost-effectiveness of these models in clinical management of LVAD patients during the critical post-operative period is worthy of future investigation.

Overall, the three models confirmed previously identified risk factors for LVAD mortality, such as age, platelet count, total bilirubin, blood urea nitrogen, SGOT-AST, white blood cell count, as well as AEs such as renal failure, respiratory failure, liver dysfunction, stroke, and post-LVAD treatments such as dialysis [51, 52, 57, 59, 61–65, 210–213]. The top three features in terms of importance for Models I and II were platelet count, white blood count and total bilirubin recorded during the first week. The former is consistent with recent studies

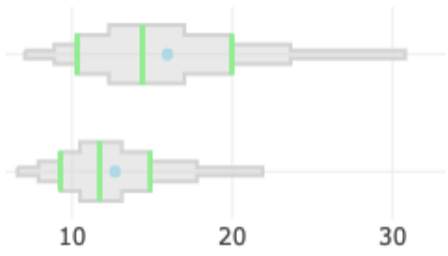


	Death $N = 572$	SURV $N = 14972$	Test Statistic
BILI_TOTAL_UMOL_LPost	22 56 172	15 22 31	$F_{1,15542}$=410, P<0.001
PLATELET_X10_3_ULPost	71 112 189	148 200 267	$F_{1,15542}$=422, P<0.001
SGOT_ASTPost	38 61 120	30 38 48	$F_{1,15542}$=296, P<0.001
WBC_X10_3_ULPost	11 15 21	9 12 15	$F_{1,15542}$=163, P<0.001
AGE_DEIDENT	55 64 69	50 59 66	$F_{1,15542}$=79, P<0.001
BUN_MMOL_LPost	8 13 20	6 9 13	$F_{1,15542}$=185, P<0.001

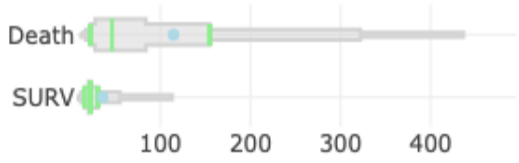
Figure 67: Descriptive statistics for top features in Model I predicting death during W2-W4 ($N = 15,544$). Table: a b c represent the lower quartile a, the median b, and the upper quartile c for continuous variables. Test used: Wilcoxon test.



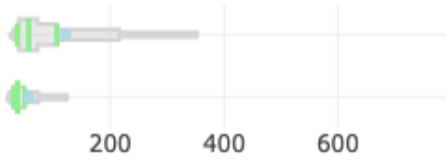
W1: Platelets (x10^3/uL)



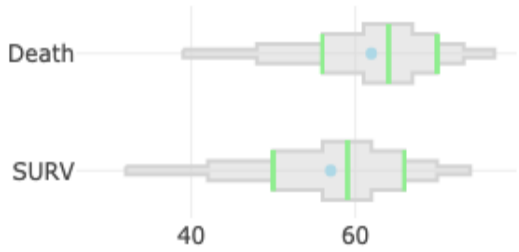
W1: White blood count (x10^3/uL)



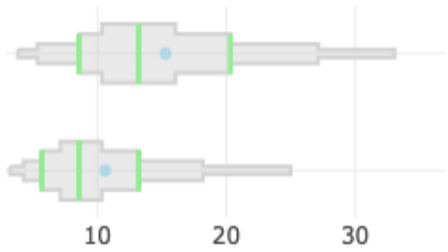
W1: Total bilirubin (µmol/L)



W1: SGOT_AST



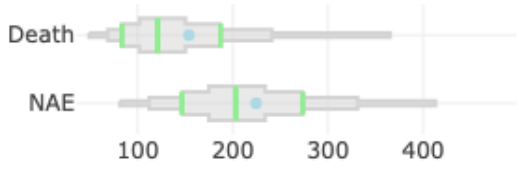
Pre-implant: Age



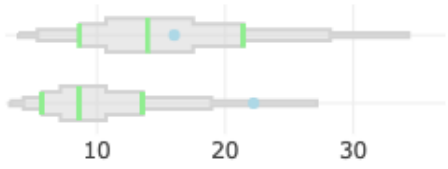
W1: Blood urea nitrogen (mmol/L)

	Death $N = 335$	SURV $N = 14917$	Test Statistic
W1: Platelets (x10^3/uL)	74 116 190	146 200 267	$F_{1,15250}$=221, P<0.001
W1: White blood count (x10^3/uL)	10 14 20	9 12 15	$F_{1,15250}$=74, P<0.001
W1: Total bilirubin (µmol/L)	22 46 155	15 22 31	$F_{1,15250}$=190, P<0.001
W1: SGOT_AST	38 56 107	30 38 49	$F_{1,15250}$=137, P<0.001
Pre-implant: Age	56 64 70	50 59 66	$F_{1,15250}$=60, P<0.001
W1: Blood urea nitrogen (mmol/L)	9 13 20	6 9 13	$F_{1,15250}$=111, P<0.001

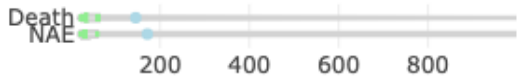
Figure 68: Descriptive statistics for top features in Model II predicting death during W3-W4 ($N = 15,252$). a b c represent the lower quartile a, the median b, and the upper quartile c for continuous variables. Test used: Wilcoxon test.



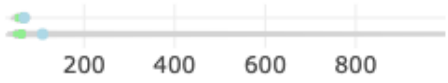
W1: Platelets (x10^3/uL)



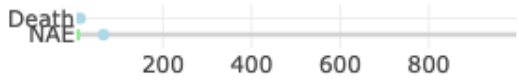
W1: Blood urea nitrogen (mmol/L)



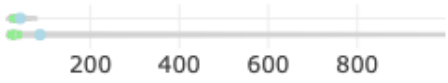
Pre-implant: SGOT_AST



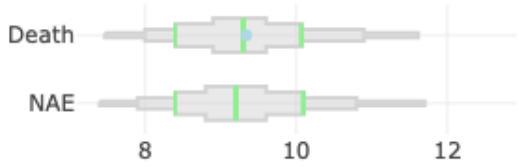
Pre-implant: Age



Pre-implant: Hemoglobin (mmol/L)



Pre-implant: Blood urea nitrogen (mg/



W1: Hemoglobin (g/dL)

	Death $N = 158$	NAE $N = 14871$	Test Statistic
W1: Platelets (x10^3/uL)	83 121 187	147 203 273	$F_{1,15027}$=94, P<0.001
W1: Blood urea nitrogen (mmol/L)	9 14 21	6 9 14	$F_{1,15027}$=56, P<0.001
Pre-implant: SGOT_AST	25 34 62	22 32 59	$F_{1,15027}$=2, P=0.2
Pre-implant: Age	54 64 70	50 60 67	$F_{1,15027}$=13, P<0.001
Pre-implant: Hemoglobin (mmol/L)	6 7 8	6 7 8	$F_{1,15027}$=13, P<0.001
Pre-implant: Blood urea nitrogen (mg/dL)	22 32 45	18 26 39	$F_{1,15027}$=12, P<0.001
W1: Hemoglobin (g/dL)	8 9 10	8 9 10	$F_{1,15027}$=0.03, P=0.9

Figure 69: Descriptive Statistics for top features in Model III predicting death during W4 ($N = 15,029$). $_a$ b $_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. Test used: Wilcoxon test.

suggesting that platelet dysfunction may be a strong predictor of adverse outcomes [214,215]. Additionally, a 2021 study by Shah et al. found that total bilirubin after LVAD was a predictor of 60-day mortality [216].

Renal failure and respiratory failure were the the only AEs that were common to three models among the 19 recorded types. This finding is consistent with multi-system organ failure being the primary cause of death within the first month (as shown in Fig. 62). The only other AEs that were identified as risk factors were bleeding (in Model II) and neurological dysfunction (in Model III.)

When considering the timing of risk factors, it is worth noting that all risk factors for mortality in W2-W4 (Model I) and W3-W4 (Model II) except for age were related to the first week post-implant data and the patient's AE history. Additionally, the top features that influenced the predictions in all three models, especially the first two, were related to the first-week follow-up data in INTERMACS, rather than the patient's AE history.

It is worth noting that INTERMACS collects post-implant follow-up data at one week and then at one month, and there is a missingness rate of over 30% in more than 200 columns of data at one week. The importance of post-LVAD data, specifically lab values, in the early weeks after implant highlights the significance of monitoring patients closely in the early post-implant period. Future model developments could benefit from access to more frequent, and more complete (less missing) data with in the early weeks.

A limitation of this study is that it was restricted to the set of clinical variables that were present in the INTERMACS registry, which, as previously mentioned, was further hampered by the high degree of missingness. Future addition of *unstructured* data, such as echocardiographic or radiographic data could potentially improve the performance of the models. A further limitation was that the majority of patients included in the publicly available version of the INTERMACS registry (2017) were implanted with older generation LVADs, such as Heartmate II. Therefore, the generalizability of the findings needs to be examined for new technology, as patients implanted with newer devices, such as Heartmate 3 LVAD, were not well represented in this cohort. However, most of the risk factors in all three models have been identified in very recent studies with the latest generation of LVADs.

7.5 Conclusion

This study used machine learning models to predict mortality within three different time frames in first month after LVAD implantation using data obtained prior to implant, during the first week of follow-up, and history of AEs from the INTERMACS National registry. The models performed well in terms of high sensitivity (AUC-ROC of greater than 0.8) suggesting they have the potential to be valuable screening tools for identifying patients who may benefit from closer monitoring and interventions during the critical post-operative period. Age, platelet count, and blood urea nitrogen levels during the first week following LVAD implantation were among the most relevant features in all three models. This study confirmed the importance of follow-up data, treatment and adverse events for predicting mortality in the early weeks after LVAD implantation.

8.0 Concluding Remarks and Future Work

This study addressed the potential risks associated with left ventricular assist devices (LVADs), which have been shown to improve survival rates and quality of life while carrying a high risk of more than twenty various types of adverse events (AEs) such as bleeding, infection, neurological dysfunction, and respiratory failure. These AEs could lead to morbidity, mortality, and adversely affect the quality of life, ultimately increasing the cost of LVAD. As a result, there was a pressing need to gain a deep understanding of AE journeys in LVAD patients, including the assessment of future AE risks during patient selection before implantation and the early detection of AEs after implantation. This study addressed these needs, some of them for the first time and others for improvement, by encompassing three machine learning applications: (1) post-LVAD sequential AE pattern mining (Chapter 4) and clustering (Chapter 6), (2) pre-LVAD predictive risk modeling (Chapter 5) , and (3) post-LVAD risk modeling (Chapter 7).

For the first time, this study presented patients' post-LVAD AEs as a temporal sequence, where each AE was considered a unique element connected to other elements in a sequence to identify transitions between AEs. Data from a large number of patients' histories of temporal AEs were summarized and visualized as a set of temporal AE sequences. Sequential pattern mining techniques were applied to extract patterns within the AE sequences, answering questions such as the types of AEs experienced by patients and the order in which they occurred. Additionally, this study explored the temporal aspects of the AE journey by identifying critical time points following LVAD implantation, common time spans, and time gaps between AEs. The post-LVAD clustering analysis provided a comprehensive overview of AEs in the LVAD population by identifying subgroups of patients with specific patterns of AEs. The results of this study has the potential to inform personalized care strategies, resource allocation, and the scheduling of follow-up appointments.

Furthermore, this study improved pre-implant predictive risk models for post-LVAD mortality and right heart failure (RHF) and addressed, for the first time, the imbalance issue and the choice of proper evaluation metrics for model performance. The pre-LVAD risk

models can enhance patient selection for LVAD implantation, enabling the identification and stabilization of high-risk patients before the procedure.

Moreover, this study developed a post-LVAD mortality risk model for the early weeks after LVAD implantation, based on pre- and post-LVAD data as well as the history of AEs. The post-LVAD risk model enable the anticipation of future AEs, empowering healthcare providers to implement preventive measures or make necessary adjustments to treatment plans. This enhancement improved patient outcomes and reduced potential complications.

8.1 Pre- and Post-LVAD Risk Models as Screening Tools

Both pre-LVAD predictive risk models and post-LVAD models have demonstrated high sensitivity but low precision. The high sensitivity indicates that these risk models can effectively serve as screening tools, aiding clinicians in the early detection of AEs by accurately identifying a majority of patients with low risk. However, due to the considerable false positive rate (low precision), they cannot be solely relied upon as diagnostic tools.

Screening tools have proven effective in various diseases, such as breast cancer, detecting conditions at early stages. Similarly, pre- and post-LVAD risk models can assist in the early identification of AEs in patients with low risk. Nevertheless, a small percentage of patients identified as high risk by these models require further diagnostic tests or procedures to confirm the presence of actual high risk of AEs, as these predictions may be false positives. Future studies should explore the existence of appropriate clinical tests or procedures to validate the risk predictions.

Additionally, a thorough investigation into the cost implications of false predictions is essential. Estimating the costs associated with false high vs. low risk predictions, along with the consequences of subsequent decisions based on these predictions, would provide valuable insights. Evaluating the cost-effectiveness from financial, clinical, and emotional perspectives will aid in determining the overall benefit of the screening tools and their applicability in specific LVAD populations. This cost analysis could also establish a threshold for acceptable false predictions, acting as a benchmark for future predictive models.

Furthermore, optimizing the usage of pre-LVAD models requires considering the entire patient journey, including the path leading to LVAD implantation. Currently, pre-implant models rely on data from a single time point shortly before surgery, whereas LVAD candidates are typically under clinical supervision for an extended period. Incorporating risk models at earlier stages of patient care, taking into account longitudinal data, may enhance their effectiveness in guiding therapy selection and decision-making.

8.2 Post-LVAD Clustering

In this research, patients' AE sequences were initially grouped into clusters based on similarity, and individual Markov models or hidden Markov models (HMMs) were fitted for each cluster. However, this approach is sensitive to the original clustering and may not adequately handle borderline cases. As a suggestion for future studies, an alternative approach is to fit a single model for the entire dataset and determine clustering during the modeling process by employing mixture hidden Markov models (MHMMs). MHMMs assume that the data consists of latent sub-populations with varying patterns, allowing for the clustering of AE sequences during the modeling process. MHMMs offer several advantages, including the ability to incorporate covariates to predict cluster memberships or transition probabilities. This enables a more comprehensive exploration of post-LVAD journey phenotypes by considering factors such as age, lab values, medication treatment, and demographic information alongside post-implant AEs.

8.3 Toward A clinical decision support system

The integration of the three modules in this research presents promising opportunities for laying the groundwork for a comprehensive toolkit or software that aids decision-making throughout the entire process of LVAD therapy. This toolkit can assist in candidate selection for VAD implantation, designing personalized care strategies, allocating resources effectively,

and implementing preventive measures for potential future AEs. By encompassing the full spectrum of VAD therapy, this integrated toolkit can support clinicians in making informed decisions and improving patient care at every stage of the treatment journey. However, the development of a clinical decision support system (CDSS) for LVAD is not without challenges. Previous studies have shown that CDSSs often fail to improve patient outcomes due to a lack of user-centered human-computer interaction considerations in their design. Additionally, there are attitudinal, informational, social, and environmental barriers that impact the adoption and perceived value of prognostic decision support tools in clinical practice. These barriers include a perceived lack of need, a lack of trust in intelligent systems, mismatched information flow, and restrictions posed by hospital environments. Future research is needed to address these challenges and develop effective CDSSs that embrace clinical context, integrate with the decision process, and blend human and machine intelligence to enhance clinical decision-making in LVAD therapy.

8.4 Beyond LVAD Therapy

This thesis primarily explores the applications of machine learning in optimizing a specific therapy option for patients with advanced heart failure. However, it is important to acknowledge that there are alternative treatment options available, including cardiac rehabilitation (a.k.a. cardiac recovery.) Future studies in the field of advanced heart failure should adopt a holistic approach to therapy selection, taking into account not only the efficacy of the treatments but also societal factors and cost effectiveness. By incorporating a comprehensive analysis of cost effectiveness, healthcare professionals can make more informed decisions about the most suitable therapy for each patient, considering both clinical outcomes and financial implications. Such research endeavors have the potential to greatly enhance the management of advanced heart failure and improve patient care in a more holistic and economically efficient manner.

Bibliography

[1] J. J. McMurray, S. Adamopoulos, S. D. Anker, A. Auricchio, M. Böhm, K. Dickstein, V. Falk, G. Filippatos, C. Fonseca, M. A. Gomez-Sanchez *et al.*, "ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012," *Eur. J. Heart Fail.*, vol. 14, no. 8, pp. 803–869, 2012.

[2] C. Norton, V. V. Georgiopoulou, A. P. Kalogeropoulos, and J. Butler, "Epidemiology and cost of advanced heart failure," *Prog. Cardiovasc. Dis.*, vol. 54, no. 2, pp. 78–85, 2011.

[3] New York Heart Association. Criteria Committee and New York Heart Association, *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels.* Little, Brown Medical Division, 1979.

[4] S. P. Chaudhry and G. C. Stewart, "Advanced heart failure: prevalence, natural history, and prognosis," *Heart Fail. Clin.*, vol. 12, no. 3, pp. 323–333, 2016.

[5] J. C. Fang, G. A. Ewald, L. A. Allen, J. Butler, C. A. W. Canary, M. Colvin-Adams, M. G. Dickinson, P. Levy, W. G. Stough, N. K. Sweitzer *et al.*, "Advanced (stage d) heart failure: a statement from the heart failure society of america guidelines committee," *J. Card. Fail.*, vol. 21, no. 6, pp. 519–534, 2015.

[6] D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, S. R. Das, S. de Ferranti, J.-P. Després, H. J. Fullerton *et al.*, "Heart disease and stroke statistics—2016 update: a report from the american heart association," *Circ.*, vol. 133, no. 4, pp. e38–e360, 2016.

[7] A. P. Ambrosy, G. C. Fonarow, J. Butler, O. Chioncel, S. J. Greene, M. Vaduganathan, S. Nodari, C. S. Lam, N. Sato, A. N. Shah *et al.*, "The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries," *J Am Coll Cardiol*, vol. 63, no. 12, pp. 1123–1133, 2014.

[8] V. L. Roger, A. S. Go, D. M. Lloyd-Jones, E. J. Benjamin, J. D. Berry, W. B. Borden, D. M. Bravata, S. Dai, E. S. Ford, C. S. Fox *et al.*, "Heart disease and stroke statistics—2012 update: a report from the american heart association," *Circulation*, pp. CIR–0b013e31 823ac046, 2011.

[9] J. J. McMurray and S. Stewart, "Epidemiology, aetiology, and prognosis of heart failure," *Heart*, vol. 83, no. 5, pp. 596–602, 2000.

[10] A. Mosterd and A. W. Hoes, "Clinical epidemiology of heart failure," *Heart*, vol. 93, no. 9, pp. 1137–1146, 2007.

[11] J. G. Rogers, R. R. Bostic, K. B. Tong, R. Adamson, M. Russo, and M. S. Slaughter, "Cost-effectiveness analysis of continuous flow left ventricular assist devices as destination therapy," *Circ. Heart Fail.*, pp. CIRCHEARTFAILURE–111, 2011.

[12] A. J. Moskowitz, E. A. Rose, and A. C. Gelijns, "The cost of long-term lvad implantation," *Ann. Thorac. Surg.*, vol. 71, no. 3, pp. S195–S198, 2001.

[13] L. K. Truby and J. G. Rogers, "Advanced heart failure: epidemiology, diagnosis, and therapeutic approaches," *HF*, vol. 8, no. 7, pp. 523–536, 2020.

[14] L. H. Lund, L. B. Edwards, A. I. Dipchand, S. Goldfarb, A. Y. Kucheryavaya, B. J. Levvey, B. Meiser, J. W. Rossano, R. D. Yusen, and J. Stehlik, "The registry of the international society for heart and lung transplantation: thirty-third adult heart transplantation report—2016; focus theme: primary diagnostic indications for transplant," *J. Heart Lung Transplant.*, vol. 35, no. 10, pp. 1158–1169, 2016.

[15] K. Khush, J. Zaroff, J. Nguyen, R. Menza, and B. Goldstein, "National decline in donor heart utilization with regional variability: 1995–2010," *Am. J. Transplantn.*, vol. 15, no. 3, pp. 642–649, 2015.

[16] A. Kilic, "The future of left ventricular assist devices," *J. Thorac. Dis.*, vol. 7, no. 12, p. 2188, 2015.

[17] G. Rigatelli, F. Santini, and G. Faggian, "Past and present of cardiocirculatory assist devices: a comprehensive critical review," *J. Geriatr. Cardiol.*, vol. 9, no. 4, p. 389, 2012.

[18] M. Yuzefpolskaya, S. E. Schroeder, B. A. Houston, M. R. Robinson, I. Gosev, A. Reyentovich, D. Koehl, R. Cantor, U. P. Jorde, J. K. Kirklin *et al.*, "The society of thoracic surgeons intermacs 2022 annual report: focus on the 2018 heart transplant allocation system," *Ann. Thorac. Surg.*, vol. 115, no. 2, pp. 311–327, 2023.

[19] H. J. Eisen, "Left ventricular assist devices (lvads): history, clinical application and complications," *Korean Circ. J.*, vol. 49, no. 7, pp. 568–585, 2019.

[20] J. K. Kirklin, F. D. Pagani, R. L. Kormos, L. W. Stevenson, E. D. Blume, S. L. Myers, M. A. Miller, J. T. Baldwin, J. B. Young, and D. C. Naftel, "Eighth annual intermacs report: Special focus on framing the impact of adverse events," *J. Heart Lung Transplant.*, vol. 36, no. 10, pp. 1080–1086, 2017.

[21] D. Emerson, J. Chikwe, P. Catarino, M. Hassanein, L. Deng, R. S. Cantor, A. Roach, R. Cole, F. Esmailian, J. Kobashigawa *et al.*, "Contemporary left ventricular assist device outcomes in an aging population: an sts intermacs analysis," *J. Am. Coll. Cardiol.*, vol. 78, no. 9, pp. 883–894, 2021.

[22] I. M. Hariri, T. Dardas, M. Kanwar, R. Cogswell, I. Gosev, E. Molina, S. L. Myers, J. K. Kirklin, P. Shah, F. D. Pagani *et al.*, "Long-term survival on lvad support: Device complications and end-organ dysfunction limit long-term success," *J. Heart Lung Transplant.*, vol. 41, no. 2, pp. 161–170, 2022.

[23] M. R. Mehra, J. C. Cleveland Jr, N. Uriel, J. A. Cowger, S. Hall, D. Horstmanshof, Y. Naka, C. T. Salerno, J. Chuang, C. Williams *et al.*, "Primary results of long-term outcomes in the momentum 3 pivotal trial and continued access protocol study phase: a study of 2200 heartmate 3 left ventricular assist device implants," *Eur. J. Heart Fail.*, vol. 23, no. 8, pp. 1392–1400, 2021.

[24] R. L. Kormos, J. Cowger, F. D. Pagani, J. J. Teuteberg, D. J. Goldstein, J. P. Jacobs, R. S. Higgins, L. W. Stevenson, J. Stehlik, P. Atluri *et al.*, "The society of thoracic surgeons intermacs database annual report: evolving indications, outcomes, and scientific partnerships," *J. Heart Lung Transplant.*, vol. 38, no. 2, pp. 114–126, 2019.

[25] J. G. Rogers, K. D. Aaronson, A. J. Boyle, S. D. Russell, C. A. Milano, F. D. Pagani, B. S. Edwards, S. Park, R. John, J. V. Conte *et al.*, "Continuous flow left ventricular assist device improves functional capacity and quality of life of advanced heart failure patients," *JACC*, vol. 55, no. 17, pp. 1826–1834, 2010.

[26] K. L. Grady, D. Naftel, L. Stevenson, M. A. Dew, G. Weidner, F. D. Pagani, J. K. Kirklin, S. Myers, T. Baldwin, and J. Young, "Overall quality of life improves to similar levels after mechanical circulatory support regardless of severity of heart failure before implantation," *J. Heart Lung Transplant.*, vol. 33, no. 4, pp. 412–421, 2014.

[27] T. Ben Gal, B. Ben Avraham, D. Milicic, M. G. Crespo-Leiro, A. J. Coats, G. Rosano, P. Seferovic, F. Ruschitzka, M. Metra, S. Anker *et al.*, "Guidance on the management of left ventricular assist device (lvad) supported patients for the non-lvad specialist

healthcare provider: executive summary," *Eur. J. Heart Fail.*, vol. 23, no. 10, pp. 1597–1609, 2021.

[28] F. Gustafsson, B. Ben Avraham, O. Chioncel, T. Hasin, A. Grupper, A. Shaul, S. Nalbantgil, Y. Hammer, W. Mullens, L. F. Tops *et al.*, "Hfa of the esc position paper on the management of lvad-supported patients for the non-lvad specialist healthcare provider part 3: at the hospital and discharge," *ESC Heart Fail.*, vol. 8, no. 6, pp. 4425–4443, 2021.

[29] M. Cameli, M. C. Pastore, G. E. Mandoli, F. Landra, M. Lisi, L. Cavigli, F. D'Ascenzi, M. Focardi, C. Carrucola, A. Dokollari *et al.*, "A multidisciplinary approach for the emergency care of patients with left ventricular assist devices: A practical guide," *Front. Cardiovasc. Med*, vol. 9, p. 923544, 2022.

[30] K. Toda, Y. Yonemoto, T. Fujita, Y. Shimahara, S. Sato, T. Nakatani, and J. Kobayashi, "Risk analysis of bloodstream infection during long-term left ventricular assist device support," *Ann. Thorac. Surg.*, vol. 94, no. 5, pp. 1387–1393, 2012.

[31] D. J. Goldstein, D. Naftel, W. Holman, L. Bellumkonda, S. V. Pamboukian, F. D. Pagani, and J. Kirklin, "Continuous-flow devices and percutaneous site infections: clinical outcomes," *J. Heart Lung Transplant.*, vol. 31, no. 11, pp. 1151–1157, 2012.

[32] N. Makki, O. Mesubi, C. Steyers, B. Olshansky, and W. T. Abraham, "Meta-analysis of the relation of ventricular arrhythmias to all-cause mortality after implantation of a left ventricular assist device," *Am. J. Card.*, vol. 116, no. 9, pp. 1385–1390, 2015.

[33] D. Acharya, R. Loyaga-Rendon, C. J. Morgan, K. A. Sands, S. V. Pamboukian, I. Rajapreyar, W. L. Holman, J. K. Kirklin, and J. A. Tallaj, "Intermacs analysis of stroke during support with continuous-flow left ventricular assist devices: risk factors and outcomes," *JACC: Heart Fail.*, vol. 5, no. 10, pp. 703–711, 2017.

[34] S. A. Akhter, A. Badami, M. Murray, T. Kohmoto, L. Lozonschi, S. Osaki, and E. B. Lushaj, "Hospital readmissions after continuous-flow left ventricular assist device implantation: incidence, causes, and cost analysis," *Ann. Thorac. Surg.*, vol. 100, no. 3, pp. 884–889, 2015.

[35] T. Hasin, Y. Marmor, W. Kremers, Y. Topilsky, C. J. Severson, J. A. Schirger, B. A. Boilson, A. L. Clavell, R. J. Rodeheffer, R. P. Frantz *et al.*, "Readmissions after implantation of axial flow left ventricular assist device," *J. Am. Coll. Cardiol.*, vol. 61, no. 2, pp. 153–163, 2013.

[36] N. McNamara, H. Narroway, M. Williams, J. Brookes, J. Farag, D. Cistulli, P. Bannon, S. Marasco, E. Potapov, and A. Loforte, "Contemporary outcomes of continuous-flow left ventricular assist devices—a systematic review," *Ann. cardiothoracic surg.*, vol. 10, no. 2, p. 186, 2021.

[37] C. Mahr, E. McGee Jr, A. Cheung, N. A. Mokadam, M. Strueber, M. S. Slaughter, M. R. Danter, W. C. Levy, R. K. Cheng, J. A. Beckman *et al.*, "Cost-effectiveness of thoracotomy approach for the implantation of a centrifugal left ventricular assist device," *Am. Soc. Artif. Intern. Organs J.*, vol. 66, no. 8, p. 855, 2020.

[38] P. E. Miller, S. G. Bromfield, Q. Ma, G. Crawford, J. Whitney, A. DeVries, and N. R. Desai, "Clinical outcomes and cost associated with an intravascular microaxial left ventricular assist device vs intra-aortic balloon pump in patients presenting with acute myocardial infarction complicated by cardiogenic shock," *JAMA Intern. Med.*, vol. 182, no. 9, pp. 926–933, 2022.

[39] M. S. Slaughter, R. Bostic, K. Tong, M. Russo, and J. G. Rogers, "Temporal changes in hospital costs for left ventricular assist device implantation," *J Card. Surg.*, vol. 26, no. 5, pp. 535–541, 2011.

[40] A. Kilic, M. A. Acker, and P. Atluri, "Dealing with surgical left ventricular assist device complications," *J Thorac. Dis.*, vol. 7, no. 12, p. 2158, 2015.

[41] A. H. Healy, S. H. McKellar, S. G. Drakos, A. Koliopoulou, J. Stehlik, and C. H. Selzman, "Physiologic effects of continuous-flow left ventricular assist devices," *J surg Res*, vol. 202, no. 2, pp. 363–371, 2016.

[42] K. V. Draper, R. J. Huang, and L. B. Gerson, "GI bleeding in patients with continuous-flow left ventricular assist devices: a systematic review and meta-analysis," *Gastrointest. Endosc.*, vol. 80, no. 3, pp. 435–446, 2014.

[43] M. S. Slaughter, J. G. Rogers, C. A. Milano, S. D. Russell, J. V. Conte, D. Feldman, B. Sun, A. J. Tatooles, R. M. Delgado III, J. W. Long *et al.*, "Advanced heart failure treated with continuous-flow left ventricular assist device," *N. Engl. J. Med.*, vol. 361, no. 23, pp. 2241–2251, 2009.

[44] J. M. Stulak, M. E. Davis, N. Haglund, S. Dunlay, J. Cowger, P. Shah, F. D. Pagani, K. D. Aaronson, and S. Maltais, "Adverse events in contemporary continuous-flow left ventricular assist devices: A multi-institutional comparison shows significant differences," *J. Thorac. Cardiovasc. Surg.*, vol. 151, no. 1, pp. 177–189, 2016.

[45] M. C. Bunte, E. H. Blackstone, L. Thuita, J. Fowler, L. Joseph, A. Ozaki, R. C. Starling, N. G. Smedira, and M. M. Mountis, "Major bleeding during HeartMate II support," *JACC*, vol. 62, no. 23, pp. 2188–2196, 2013.

[46] V. K. Topkara, S. Kondareddy, F. Malik, I.-W. Wang, D. L. Mann, G. A. Ewald, and N. Moazami, "Infectious complications in patients with left ventricular assist device: etiology and outcomes in the continuous-flow era," *Ann. Thorac. Surg.*, vol. 90, no. 4, pp. 1270–1277, 2010.

[47] L. Harvey, C. Holley, S. S. Roy, P. Eckman, R. Cogswell, K. Liao, and R. John, "Stroke after left ventricular assist device implantation: outcomes in the continuous-flow era," *Ann. Thorac. Surg.*, vol. 100, no. 2, pp. 535–541, 2015.

[48] J. Z. Willey, M. V. Gavalas, P. N. Trinh, M. Yuzefpolskaya, A. R. Garan, A. P. Levin, K. Takeda, H. Takayama, J. Fried, Y. Naka *et al.*, "Outcomes after stroke complicating left ventricular assist device," *J. Heart Lung Transplant.*, vol. 35, no. 8, pp. 1003–1009, 2016.

[49] P. S. Joy, G. Kumar, A. K. Guddati, J. K. Bhama, and L. M. Cadaret, "Risk factors and outcomes of gastrointestinal bleeding in left ventricular assist device recipients," *Am. J. Cardiol.*, vol. 117, no. 2, pp. 240–244, 2016.

[50] A. J. Boyle, U. P. Jorde, B. Sun, S. J. Park, C. A. Milano, O. H. Frazier, K. S. Sundareswaran, D. J. Farrar, S. D. Russell, H. I. C. Investigators *et al.*, "Pre-operative risk factors of bleeding and stroke during left ventricular assist device support: an analysis of more than 900 heartmate ii outpatients," *JACC*, vol. 63, no. 9, pp. 880–888, 2014.

[51] J. Cowger, K. Sundareswaran, J. G. Rogers *et al.*, "Predicting survival in patients receiving continuous flow left ventricular assist devices: the heartmate II risk score," *J. Am. Coll. Cardiol.*, vol. 61, no. 3, pp. 313–321, 2013.

[52] K. Lietz, J. Long, A. Kfoury *et al.*, "Outcomes of left ventricular assist device implantation as destination therapy in the post-rematch era," *Circulation*, vol. 2007, p. 116.

[53] E. Y. Birati, T. C. Hanff, D. Maldonado *et al.*, "Predicting long term outcome in patients treated with continuous flow left ventricular assist device-the penn-columbia risk score," *Circulation*, vol. 132, no. suppl_3, pp. A18 674–A18 674, 2015.

[54] N. A. Loghmanpour, M. K. Kanwar, M. J. Druzdzel, R. L. Benza, S. Murali, and J. F. Antaki, "A new bayesian network-based risk stratification model for prediction of short-term and long-term lvad mortality," *ASAIO journal*, vol. 61, no. 3, p. 313, 2015.

[55] M. K. Kanwar, L. C. Lohmueller, R. L. Kormos, J. J. Teuteberg, J. G. Rogers, J. Lindenfeld, S. H. Bailey, C. K. McIlvennan, R. Benza, S. Murali *et al.*, "A bayesian model to predict survival after left ventricular assist device implantation," *JACC*, 2018.

[56] C. Frankfurter, M. Molinero, J. K. Vishram-Nielsen, F. Foroutan, S. Mak, V. Rao, F. Billia, A. Orchanian-Cheff, and A. C. Alba, "Predicting the risk of right ventricular failure in patients undergoing left ventricular assist device implantation: a systematic review," *Circ. Heart Fail.*, vol. 13, no. 10, p. e006994, 2020.

[57] A. K. Ravichandran and J. Cowger, "Left ventricular assist device patient selection: do risk scores help?" *J. Thorac. Dis.*, vol. 7, no. 12, p. 2080, 2015.

[58] A. Kilic, D. Dochtermann, R. Padman, J. K. Miller, and A. Dubrawski, "Using machine learning to improve risk prediction in durable left ventricular assist devices," *Plos one*, vol. 16, no. 3, p. e0247866, 2021.

[59] E. Y. Birati, T. C. Hanff, D. Maldonado, E. W. Grandin, P. J. Kennel, J. A. Mazurek, E. Vorovich, M. Seigerman, J. L. Howard, M. A. Acker *et al.*, "Predicting long term outcome in patients treated with continuous flow left ventricular assist device: the penn—columbia risk score," *J. Am. Heart Assoc.*, vol. 7, no. 6, p. e006408, 2018.

[60] J. A. Cowger, L. Castle, K. D. Aaronson, M. S. Slaughter, S. Moainie, M. Walsh, and C. Salerno, "The heartmate ii risk score: an adjusted score for evaluation of all continuous-flow left ventricular assist devices," *ASAIO J.*, vol. 62, no. 3, pp. 281–285, 2016.

[61] A. Kilic, L. Seese, F. Pagani, and R. Kormos, "Identifying temporal relationships between in-hospital adverse events after implantation of durable left ventricular assist devices," *J. Am. Heart Assoc.*, vol. 9, no. 8, p. e015449, 2020.

[62] F. Movahedi, R. L. Kormos, L. Lohmueller, L. Seese, M. Kanwar, S. Murali, Y. Zhang, R. Padman, and J. F. Antaki, "Sequential pattern mining of longitudinal adverse events after left ventricular assist device implant," *IEEE J Biomed Health Inform*, vol. 24, no. 8, pp. 2347–2358, 2019.

[63] E. A. Genovese, M. A. Dew, J. J. Teuteberg, M. A. Simon, J. K. Bhama, C. A. Bermudez, K. L. Lockard, S. Winowich, and R. L. Kormos, "Early adverse events as predictors of 1-year mortality during mechanical circulatory support," *J. Heart Lung Transplant.*, vol. 29, no. 9, pp. 981–988, 2010.

[64] E. A. Genovese, M. A. Dew, J. J. Teuteberg, M. A. Simon, J. Kay, M. P. Siegenthaler, J. K. Bhama, C. A. Bermudez, K. L. Lockard, S. Winowich *et al.*, "Incidence and patterns of adverse event onset during the first 60 days after ventricular assist device implantation," *Ann. Thorac. Surg.*, vol. 88, no. 4, pp. 1162–1170, 2009.

[65] K. Majumder, J. R. Spratt, C. T. Holley, S. S. Roy, R. J. Cogswell, K. Liao, and R. John, "Impact of postoperative liver dysfunction on survival after left ventricular assist device implantation," *Ann. Thorac. Surg.*, vol. 104, no. 5, pp. 1556–1562, 2017.

[66] A. Tibrewala, R. M. Wehbe, T. Wu, R. Harap, K. Ghafourian, J. E. Wilcox, I. S. Okwuosa, E. E. Vorovich, F. S. Ahmad, C. Yancy *et al.*, "Hyponatremia is a powerful predictor of poor prognosis in left ventricular assist device patients," *ASAIO Journal*, vol. 68, no. 12, pp. 1475–1482, 2022.

[67] G. P. Macaluso, F. D. Pagani, M. S. Slaughter, C. A. Milano, E. D. Feller, A. J. Tatooles, J. G. Rogers, and G. M. Wieselthaler, "Time in therapeutic range significantly impacts survival and adverse events in destination therapy patients," *ASAIO Journal*, vol. 68, no. 1, p. 14, 2022.

[68] B. K. Martinez, B. Yik, R. Tran, S. Ilham, C. I. Coleman, D. L. Jennings, and W. L. Baker, "Meta-analysis of time in therapeutic range in continuous-flow left ventricular assist device patients receiving warfarin," *Artif. Organs*, vol. 42, no. 7, pp. 700–704, 2018.

[69] K. Kido, B. George, R. J. Charnigo, T. E. Macaulay, S. D. Brouse, and M. Guglin, "Chronologic changes and correlates of loop diuretic dose in patients with left ventricular assist device," *ASAIO Journal*, vol. 63, no. 6, pp. 774–780, 2017.

[70] P. Khazanie, B. G. Hammill, C. B. Patel, M. S. Kiernan, L. B. Cooper, S. V. Arnold, T. J. Fendler, J. A. Spertus, L. H. Curtis, and A. F. Hernandez, "Use of heart failure medical therapies among patients with left ventricular assist devices: insights from intermacs," *J. Card. Fail.*, vol. 22, no. 9, pp. 672–679, 2016.

[71] R. Krishnamani, D. DeNofrio, and M. A. Konstam, "Emerging ventricular assist devices for long-term cardiac support," *Nat. Rev. Cardiol.*, vol. 7, no. 2, p. 71, 2010.

[72] L. E. Rodriguez, E. E. Suarez, M. Loebe, and B. A. Bruckner, "Ventricular assist devices (vad) therapy: New technology, new hope?" *Methodist Debakey Cardiovasc J*, vol. 9, no. 1, p. 32, 2013.

[73] Thoratec, [Online; accessed 11-July-2018]. [Online]. Available: http://www.thoratec.com/about-us/media-room/library.aspx

[74] F. D. Pagani, L. W. Miller, S. D. Russell, K. D. Aaronson, R. John, A. J. Boyle, J. V. Conte, R. C. Bogaev, T. E. MacGillivray, Y. Naka *et al.*, "Extended mechanical circulatory support with a continuous-flow rotary left ventricular assist device," *J. Am. Coll. Cardiol.*, vol. 54, no. 4, pp. 312–321, 2009.

[75] D. Feldman, S. V. Pamboukian, J. J. Teuteberg, E. Birks, K. Lietz, S. A. Moore, J. A. Morgan, F. Arabia, M. E. Bauman, H. W. Buchholz *et al.*, "The 2013 international society for heart and lung transplantation guidelines for mechanical circulatory support: executive summary," *J. Heart Lung Transplant.*, vol. 32, no. 2, pp. 157–187, 2013.

[76] M. C. Deng, L. B. Edwards, M. I. Hertz, A. W. Rowe, and R. L. Kormos, "Mechanical circulatory support device database of the international society for heart and lung transplantation: first annual report—2003," *J. Heart Lung Transplant.*, vol. 22, no. 6, pp. 653–662, 2003.

[77] J. K. Kirklin, D. C. Naftel, L. W. Stevenson, R. L. Kormos, F. D. Pagani, M. A. Miller, K. Ulisney, and J. B. Young, "Intermacs database for durable devices for circulatory support: first annual report," *J. Heart Lung Transplant.*, vol. 27, no. 10, pp. 1065–1072, 2008.

[78] U. Food and D. A. R. world evidence. [Online]. Available: https://www.fda.gov/science-research/science-and-research-spe-cial-topics/real-world-evidence

[79] R. L. Kormos, C. F. Antonides, D. J. Goldstein, J. A. Cowger, R. C. Starling, J. K. Kirklin, J. E. Rame, D. Rosenthal, M. L. Mooney, K. Caliskan *et al.*, "Updated definitions of adverse events for trials and registries of mechanical circulatory support: A consensus statement of the mechanical circulatory support academic research consortium," *The Journal of Heart and Lung Transplantation*, vol. 39, no. 8, pp. 735–750, 2020.

[80] A. R. Consortium. [Online]. Available: https://www.cardialy-sis.com/events/clinical-endpoint-definitions-consensus

[81] M. A. Al-Ani, C. Bai, A. Hashky, A. M. Parker, J. R. Vilaro, J. M. Aranda Jr, B. Shickel, P. Rashidi, A. Bihorac, M. M. Ahmed *et al.*, "Artificial intelligence guidance of advanced heart failure therapies: A systematic scoping review," *Front. Cardiovas. Med.*, vol. 10, p. 1127716, 2023.

[82] N. Gautam, S. N. Ghanta, A. Clausen, P. Saluja, K. Sivakumar, G. Dhar, Q. Chang, D. DeMazumder, M. G. Rabbat, S. J. Greene *et al.*, "Contemporary applications of machine learning for device therapy in heart failure," *Heart Fail.*, vol. 10, no. 9, pp. 603–622, 2022.

[83] A. Montisci, V. Palmieri, M. T. Vietri, S. Sala, C. Maiello, F. Donatelli, and C. Napoli, "Big data in cardiac surgery: real world and perspectives," *J. Card. Surg.*, vol. 17, no. 1, p. 277, 2022.

[84] J. M. Graves, C. Pruett, K. Stephenson, E. Deych, B. Q. Yang, J. M. Vader, and B. F. Gage, "Established clinical prediction rules for bleeding had mediocre discrimination in left ventricular assist device recipients," *Am. Soc. Artif. Intern. Organs J.*, pp. 10–1097, 2022.

[85] A. Peivandi, H. Welp, M. Scherer, J. R. Sindermann, N.-M. Wagner, and A. M. Dell'Aquila, "An external validation study of the utah bleeding risk score," *Eur. J. Cardio-Thorac.*, vol. 62, no. 1, p. ezab572, 2022.

[86] M. K. Gangwani, M. Aziz, A. Nawras, F. Priyanka, Z. Ahmed, R. S. Khan, M. A. Qamar, F. Haroon, A. Aziz, W. L. Smith *et al.*, "Predictors of gastrointestinal bleeding in patients following left ventricular assist device implantation: a systematic review and meta-analysis," *Future Cardiol.*, vol. 18, no. 12, pp. 957–967, 2022.

[87] A.-K. Köhler, H. Körperich, M. Morshuis, C. C. Freytag, J. Gummert, W. Burchert, R. Preuss, and J. Körfer, "Pre-operative risk factors for driveline infection in left ventricular-assist device patients," *ESC Heart Fail.*, 2022.

[88] P. Tattevin, E. Flécher, V. Auffret, C. Leclercq, S. Boulé, A. Vincentelli, C. Dambrin, C. Delmas, L. Barandon, V. Veniard *et al.*, "Risk factors and prognostic impact of left ventricular assist device–associated infections," *American Heart Journal*, vol. 214, pp. 69–76, 2019.

[89] L. Hubbert, J. Baranowski, B. Delshad, and H. Ahn, "Left atrial pressure monitoring with an implantable wireless pressure sensor after implantation of a left ventricular assist device," *Am. Soc. Artif. Intern. Organs J.*, vol. 63, no. 5, p. e60, 2017.

[90] M. Fetanat, M. Stevens, C. Hayward, and N. Lovell, "Adaptive sensorless control of lvad using deep convolutional neural network," *J. Heart Lung Transplant.*, vol. 40, no. 4, p. S172, 2021.

[91] R. Cogswell, M. Masotti, J. Schultz, V. Maharaj, A. El Rafei, M. Fraser, M. Mutschler, C. Martin, T. Thenappan, M. Pritzker *et al.*, "A new hemodynamic profile signaling early death on left ventricular assist device," *J. Heart Lung Transplant.*, vol. 40, no. 4, p. S172, 2021.

[92] J. K. Kirklin, F. D. Pagani, D. J. Goldstein, R. John, J. G. Rogers, P. Atluri, F. A. Arabia, A. Cheung, W. Holman, C. Hoopes *et al.*, "American association for thoracic surgery/international society for heart and lung transplantation guidelines on selected topics in mechanical circulatory support," *J. Thorac. Cardiovasc. Surg.*, vol. 159, no. 3, pp. 865–896, 2020.

[93] J.-N. Trochu, P. Leprince, M. Bielefeld-Gomez, O. Bastien, F. Beauvais, J.-P. Gueffet, D. Logeart, R. Isnard, M.-C. Iliou, C. Leclercq *et al.*, "Left ventricle assist device: When and which patients should we refer?" *Arch. Cardiovasc. Dis.*, vol. 105, no. 2, pp. 114–121, 2012.

[94] M. K. Kanwar, A. Kilic, and M. R. Mehra, "Machine learning, artificial intelligence and mechanical circulatory support: A primer for clinicians," pp. 414–425, 2021.

[95] A. Fernández, S. García, M. Galar *et al.*, *Learning from imbalanced data sets*. Springer, 2018.

[96] V. López, A. Fernández, S. García *et al.*, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Inf. Sci.*, vol. 250, pp. 113–141, 2013.

[97] B. Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016.

[98] C. G. Weng and J. Poon, "A new evaluation measure for imbalanced datasets," in *Proc. 7th AusDM*, 2008, pp. 27–32.

[99] D. Berrar and P. Flach, "Caveats and pitfalls of ROC analysis in clinical microarray research (and how to avoid them)," *Brief. bioinformatics*, vol. 13, no. 1, pp. 83–97, 2012.

[100] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.

[101] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. the 23rd ICML*, 2006, pp. 233–240.

[102] E. J. Molina, P. Shah, M. S. Kiernan, W. K. Cornwell III, H. Copeland, K. Takeda, F. G. Fernandez, V. Badhwar, R. H. Habib, J. P. Jacobs *et al.*, "Sts INTERMACS 2020 annu. rep." *Ann. Thorac. Surg.*, vol. 111, no. 3, pp. 778–792, 2021.

[103] S. E. Felix, L. Numan, M. I. Oerlemans, E. Aarts, F. Z. Ramjankhan, M. Gianoli, F. W. Asselbergs, N. De Jonge, and L. W. Van Laake, "Incidence and risk factors of late right heart failure in chronic mechanical circulatory support," *Artif. Organs*, 2023.

[104] G. Del Rio-Pertuz and N. Nair, "Gastrointestinal bleeding in patients with continuous-flow left ventricular assist devices: A comprehensive review," *Artif. Organs.*, vol. 47, no. 1, pp. 12–23, 2023.

[105] G. Spano, E. Buffle, L. N. Walti, M. Mihalj, D. R. Cameron, M. Martinelli, M. Fürholz, Y.-A. Que, C. Hayward, D. Reineke *et al.*, "Ten-year retrospective cohort analysis of ventricular assist device infections," *Artif. Organs.*, 2022.

[106] J. M. Schaffer, J. G. Allen, E. S. Weiss, G. J. Arnaoutakis, N. D. Patel, S. D. Russell, A. S. Shah, and J. V. Conte, "Infectious complications after pulsatile-flow and continuous-flow left ventricular assist device implantation," *J. Heart Lung Transplant.*, vol. 30, no. 2, pp. 164–174, 2011.

[107] V. Sharma, S. V. Deo, J. M. Stulak, L. A. Durham, R. C. Daly, S. J. Park, L. M. Baddour, K. Mehra, and L. D. Joyce, "Driveline infections in left ventricular assist devices: implications for destination therapy," *Ann. Thorac. Surg.*, vol. 94, no. 5, pp. 1381–1386, 2012.

[108] F. Al-Mufti, A. Bauerschmidt, J. Claassen, P. M. Meyers, P. C. Colombo, and J. Z. Willey, "Neuroendovascular interventions for acute ischemic strokes in patients supported with left ventricular assist devices: A single-center case series and review of the literature," *World Neurosurg*, vol. 88, pp. 199–204, 2016.

[109] T. S. Kato, P. C. Schulze, J. Yang, E. Chan, K. Shahzad, H. Takayama, N. Uriel, U. Jorde, M. Farr, Y. Naka *et al.*, "Pre-operative and post-operative risk factors associated with neurologic complications in patients with advanced heart failure supported by a left ventricular assist device," *J Heart Lung Transplant*, vol. 31, no. 1, pp. 1–8, 2012.

[110] S. Sami, S. Al-Araji, and K. Ragunath, "Review article: gastrointestinal angiodysplasia–pathogenesis, diagnosis and management," *Aliment. pharmacol. & Ther.*, vol. 39, no. 1, pp. 15–34, 2014.

[111] P. Shah, U. S. Tantry, K. P. Bliden, and P. A. Gurbel, "Bleeding and thrombosis associated with ventricular assist device therapy," *J. Heart Lung Transplant*, vol. 36, no. 11, pp. 1164–1173, 2017.

[112] B. H. Trachtenberg, A. M. Cordero-Reyes, M. Aldeiri, P. Alvarez, A. Bhimaraj, G. Ashrith, B. Elias, E. E. Suarez, B. Bruckner, M. Loebe *et al.*, "Persistent blood stream infection in patients supported with a continuous-flow left ventricular assist device is associated with an increased risk of cerebrovascular accidents," *J. Card. Fail.*, vol. 21, no. 2, pp. 119–125, 2015.

[113] A. Aggarwal, A. Gupta, S. Kumar, J. A. Baumblatt, S. Pauwaa, C. Gallagher, A. Treitman, P. Pappas, A. Tatooles, and G. Bhat, "Are blood stream infections associated with an increased risk of hemorrhagic stroke in patients with a left ventricular assist device?" *ASAIO J.*, vol. 58, no. 5, pp. 509–513, 2012.

[114] M. J. Zaki and W. Meira Jr, *Data mining and machine learning: Fundamental concepts and algorithms*. Cambridge University Press, 2020.

[115] T. F. Liao, D. Bolano, C. Brzinsky-Fay, B. Cornwell, A. E. Fasang, S. Helske, R. Piccarreta, M. Raab, G. Ritschard, E. Struffolino *et al.*, "Sequence analysis: Its past, present, and future," *Soc. Sci. Res.*, vol. 107, p. 102772, 2022.

[116] P. Fournier-Viger, J. C.-W. Lin, R. U. Kiran, Y. S. Koh, and R. Thomas, "A survey of sequential pattern mining," *Data Sci. Pattern Recognit.*, vol. 1, no. 1, pp. 54–77, 2017.

[117] P. Yadav, M. Steinbach, V. Kumar, and G. Simon, "Mining electronic health records (ehrs) a survey," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 1–40, 2018.

[118] M. S. Islam, M. M. Hasan, X. Wang, H. D. Germack *et al.*, "A systematic review on healthcare analytics: application and theoretical perspective of data mining," in *Healthcare*, vol. 6, no. 2. Multidisciplinary Digital Publishing Institute, 2018, p. 54.

[119] T. Mitsa, *Temporal data mining*. CRC press, 2010.

[120] J. J. Teuteberg, J. C. Cleveland Jr, J. Cowger, R. S. Higgins, D. J. Goldstein, M. Keebler, J. K. Kirklin, S. L. Myers, C. T. Salerno, J. Stehlik *et al.*, "The society of thoracic surgeons intermacs 2019 annual report: the changing landscape of devices and indications," *Ann. Thorac. Surg.*, vol. 109, no. 3, pp. 649–660, 2020.

[121] F. Moscato, C. Gross, M. Maw, T. Schlöglhofer, M. Granegger, D. Zimpfer, and H. Schima, "The left ventricular assist device as a patient monitoring system," *Ann. Thorac. Cardiovasc. Surg.*, vol. 10, no. 2, p. 221, 2021.

[122] M. Khaledi, M. Dehghani, M. Mohammadi, and R. Abolpour, "Controller design for left ventricular assist devices in patients with heart failure," in *27th national and 5th Int. Iranian conf. Biomed. Eng.* IEEE, 2020, pp. 326–332.

[123] M. Maw, C. Gross, T. Schloeglhofer, K. Dimitrov, D. Zimpfer, F. Moscato, and H. Schima, "Development of suction detection algorithms for a left ventricular assist device from patient data," *Biomed. Signal Process. Control*, vol. 69, p. 102910, 2021.

[124] S. Sadatieh, M. Dehghani, M. Mohammadi, and R. Boostani, "Extremum-seeking control of left ventricular assist device to maximize the cardiac output and prevent suction," *Chaos, Solitons & Fractals*, vol. 148, p. 111013, 2021.

[125] S. Aras, T. Johnson, C. Gniady, R. Skaria, and Z. Khalpey, "Indetector–automatic detection of infected driveline regions," *Smart Health*, vol. 9, pp. 170–178, 2018.

[126] K. Kourou, G. Rigas, K. P. Exarchos, Y. Goletsis, T. P. Exarchos, S. Jacobs, B. Meyns, M.-G. Trivella, and D. I. Fotiadis, "Prediction of time dependent survival in hf patients after vad implantation using pre-and post-operative data," *Comput. Biol. Med.*, vol. 70, pp. 99–105, 2016.

[127] S. E. Felix, A. Bagheri, F. R. Ramjankhan, M. R. Spruit, D. Oberski, N. De Jonge, L. W. Van Laake, W. J. Suyker, and F. W. Asselbergs, "A data mining-based cross-industry process for predicting major bleeding in mechanical circulatory support," *Eur. heart J. Digit. health*, vol. 2, no. 4, pp. 635–642, 2021.

[128] F. Movahedi, R. Padman, and J. F. Antaki, "Limitations of receiver operating characteristic curve on imbalanced data: assist device mortality risk scores," *The Journal of Thoracic and Cardiovascular Surgery*, 2021.

[129] J. A. Osheroff, J. M. Teich, D. Levick, L. Saldana, F. Velasco, D. F. Sittig, K. M. Rogers, and R. A. Jenders, *Improving outcomes with clinical decision support: an implementer's guide*. CRC Press, 2012.

[130] J. A. Osheroff, J. M. Teich, B. Middleton, E. B. Steen, A. Wright, and D. E. Detmer, "A roadmap for national action on clinical decision support," *J. Am. Med. Inform. Assoc.*, vol. 14, no. 2, pp. 141–145, 2007.

[131] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: benefits, risks, and strategies for success," *NPJ Digit. Med.*, vol. 3, no. 1, p. 17, 2020.

[132] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: a systematic review," *Appl. Sci.*, vol. 11, no. 11, p. 5088, 2021.

[133] A. Belard, T. Buchman, J. Forsberg, B. K. Potter, C. J. Dente, A. Kirk, and E. Elster, "Precision diagnosis: a view of the clinical decision support systems (cdss) landscape through the lens of critical care," *J. Clin. Monit. Comput.*, vol. 31, pp. 261–271, 2017.

[134] M. Abbasi and S. Kashiyarndi, "Clinical decision support systems: A discussion on different methodologies used in health care," *Marlaedalen University Sweden*, 2006.

[135] E. S. Berner, *Clinical decision support systems*. Springer, 2007, vol. 233.

[136] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, and P. C. Tang, "Clinical decision support systems for the practice of evidence-based medicine," *J. Am. Med. Inform. Assoc.*, vol. 8, no. 6, pp. 527–534, 2001.

[137] R. C. Deo, "Machine learning in medicine," *Circ.*, vol. 132, no. 20, pp. 1920–1930, 2015.

[138] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *N. Engl. J. Med.*, vol. 375, no. 13, p. 1216, 2016.

[139] J. Schaaf, M. Sedlmayr, J. Schaefer, and H. Storf, "Diagnosis of rare diseases: a scoping review of clinical decision support systems," *Orphanet J. Rare Dis.*, vol. 15, no. 1, pp. 1–14, 2020.

[140] S. Walsh, E. E. de Jong, J. E. van Timmeren, A. Ibrahim, I. Compter, J. Peerlings, S. Sanduleanu, T. Refaee, S. Keek, R. T. Larue *et al.*, "Decision support systems in oncology," *JCO Clin. Cancer Inform.*, vol. 3, pp. 1–9, 2019.

[141] C. Mazo, C. Kearns, C. Mooney, and W. M. Gallagher, "Clinical decision support systems in breast cancer: a systematic review," *Cancers*, vol. 12, no. 2, p. 369, 2020.

[142] F. Velickovski, L. Ceccaroni, J. Roca, F. Burgos, J. B. Galdiz, N. Marina, and M. Lluch-Ariet, "Clinical decision support systems (CDSS) for preventive management of COPD patients," *J. Transl. Med.*, vol. 12, pp. 1–10, 2014.

[143] P. Sherimon, V. Sherimon, S. Preethii, R. V. Nair, and R. Mathew, "A systematic review of clinical decision support systems in alzheimer's disease domain," *Int. j. online biomed.*, vol. 17, no. 08, p. 75, 2021.

[144] P. Jia, P. Zhao, J. Chen, and M. Zhang, "Evaluation of clinical decision support systems for diabetes care: an overview of current evidence," *J Eval. Clin. Pract.*, vol. 25, no. 1, pp. 66–77, 2019.

[145] R. Jeffery, E. Iserman, R. Haynes, and C. S. R. Team, "Can computerized clinical decision support systems improve diabetes management? a systematic review and meta-analysis," *Diabet. Med.*, vol. 30, no. 6, pp. 739–745, 2013.

[146] S. M. Ali, R. Giordano, S. Lakhani, and D. M. Walker, "A review of randomized controlled trials of medical record powered clinical decision support system to improve quality of diabetes care," *Int. J. Med. Inform.*, vol. 87, pp. 91–100, 2016.

[147] Y. Lu, E. R. Melnick, and H. M. Krumholz, "Clinical decision support in cardiovascular medicine," *bmj*, vol. 377, 2022.

[148] R. Anchala, M. P. Pinto, A. Shroufi, R. Chowdhury, J. Sanderson, L. Johnson, P. Blanco, D. Prabhakaran, and O. H. Franco, "The role of decision support system (dss) in prevention of cardiovascular disease: a systematic review and meta-analysis," vol. 7, no. 10, p. e47064, 2012.

[149] E. Karvounis, M. Tsipouras, A. Tzallas, N. Katertsidis, K. Stefanou, Y. Goletsis, M. Frigerio, A. Verde, R. Caruso, B. Meyns *et al.*, "A decision support system for the treatment of patients with ventricular assist device support," *Methods Inf. Med.*, vol. 53, no. 02, pp. 121–136, 2014.

[150] A. T. Tzallas, N. S. Katertsidis, E. C. Karvounis, M. G. Tsipouras, G. Rigas, Y. Goletsis, K. Zielinski, L. Fresiello, A. Di Molfetta, G. Ferrari *et al.*, "Modeling and simulation of speed selection on left ventricular assist devices," *Comput. Biol. Med.*, vol. 51, pp. 128–139, 2014.

[151] M. G. Tsipouras, E. C. Karvounis, A. T. Tzallas, N. S. Katertsidis, Y. Goletsis, M. Frigerio, A. Verde, M. G. Trivella, and D. I. Fotiadis, "Adverse event prediction in patients with left ventricular assist devices," in *35th Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.* IEEE, 2013, pp. 1314–1317.

[152] P. Valdastri, N. Taccini, A. Pinciaroli, M. Nannizzi, and P. Dario, "Wearable and implanted sensors platform to monitor and control left ventricular assist devices," in *5th European Conference of the International Federation for Medical and Biological Engineering: 14–18 September 2011, Budapest, Hungary.* Springer, 2012, pp. 964–967.

[153] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. Stat.*, pp. 1189–1232, 2001.

[154] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd acm sigkdd*, 2016, pp. 785–794.

[155] B. Ghojogh and M. Crowley, "The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial," *arXiv preprint arXiv:1905.12787*, 2019.

[156] M. J. Zaki, "Spade: An efficient algorithm for mining frequent sequences," *Mach. Learn.*, vol. 42, no. 1, pp. 31–60, 2001.

[157] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *JASA*, vol. 53, no. 282, pp. 457–481, 1958.

[158] K. Lietz, J. W. Long, A. G. Kfoury, M. S. Slaughter, M. A. Silver, C. A. Milano, J. G. Rogers, L. W. Miller, M. Deng, Y. Naka *et al.*, "Impact of center volume on outcomes of left ventricular assist device implantation as destination therapy: analysis of the

thoratec heartmate registry, 1998 to 2005," *Circ. Heart Fail.*, vol. 2, no. 1, pp. 3–10, 2009.

[159] J. A. Cowger, J. M. Stulak, P. Shah, T. F. Dardas, F. D. Pagani, S. M. Dunlay, S. Maltais, K. D. Aaronson, R. Singh, N. A. Mokadam *et al.*, "Impact of center left ventricular assist device volume on outcomes after implantation: an intermacs analysis," *JACC: Heart Fail.*, vol. 5, no. 10, pp. 691–699, 2017.

[160] L. Breiman, "Random forests," *Mach. learning*, vol. 45, no. 1, pp. 5–32, 2001.

[161] J. Cook and V. Ramadas, "When to consult precision-recall curves," *SJ*, vol. 20, no. 1, pp. 131–148, 2020.

[162] M. A. Mazurowski, P. A. Habas, J. M. Zurada *et al.*, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," *Neural Netw.*, vol. 21, no. 2-3, pp. 427–436, 2008.

[163] L. Zhang, H. Yang, and Z. Jiang, "Imbalanced biomedical data classification using self-adaptive multilayer elm combined with dynamic gan," *Biomed. Eng. Online*, vol. 17, no. 1, p. 181, 2018.

[164] T. Gao, Y. Hao, H. Zhang *et al.*, "Predicting pathological response to neoadjuvant chemotherapy in breast cancer patients based on imbalanced clinical data," *Pers. Ubiquit. Comput.*, vol. 22, no. 5-6, pp. 1039–1047, 2018.

[165] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," *J. Biomed. Inform*, vol. 90, p. 103089, 2019.

[166] H. Ishwaran and E. H. Blackstone, "Commentary: Dabblers: Beware of hidden dangers in machine-learning comparisons," 2020.

[167] H. Ishwaran and R. O'Brien, "Editorial commentary: the problem of class imbalance in biomedical data," *J Thorac Cardiovasc Surg*, vol. 1, p. 2, 2020.

[168] ——, "Letter to the editor: the standardization and automation of machine learning for biomedical data," *J Thorac Cardiovasc Surg*, 2020.

[169] H. Guo and H. L. Viktor, "Learning from imbalanced data sets with boosting and data generation: the databoost-im approach," *SIGKDD Explor.*, vol. 6, no. 1, pp. 30–39, 2004.

[170] R. O'Brien and H. Ishwaran, "A random forests quantile classifier for class imbalanced data," *Pattern Recognit.*, vol. 90, pp. 232–249, 2019.

[171] S. J. Swamidass, C.-A. Azencott, K. Daily, and P. Baldi, "A croc stronger than roc: measuring, visualizing and optimizing early retrieval," *Bioinformatics*, vol. 26, no. 10, pp. 1348–1356, 2010.

[172] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction.* Springer, 2009, vol. 2.

[173] J. F. Yates, "External correspondence: Decompositions of the mean probability score," *Organ. Behav. Hum. Decis. Process.*, vol. 30, no. 1, pp. 132–156, 1982.

[174] Y. Ochiai, P. M. McCarthy, N. G. Smedira, M. K. Banbury, J. L. Navia, J. Feng, A. P. Hsu, M. L. Yeager, T. Buda, K. J. Hoercher *et al.*, "Predictors of severe right ventricular failure after implantable left ventricular assist device insertion: analysis of 245 patients," *Circulation*, vol. 106, no. 12_suppl_1, pp. I–198, 2002.

[175] T. S. Wang, A. F. Hernandez, G. M. Felker, C. A. Milano, J. G. Rogers, and C. B. Patel, "Valvular heart disease in patients supported with left ventricular assist devices," *Circ. Heart Fail.*, vol. 7, no. 1, pp. 215–222, 2014.

[176] M. Çelik, J. M. Stulak, and S. Maltais, "The importance of coronary artery disease and special considerations for left ventricular assist device implantation," *Ann. Cardiothorac. Surg.*, vol. 10, no. 2, p. 268, 2021.

[177] N. S. Rajliwall, R. Davey, and G. Chetty, "Cardiovascular risk prediction based on xgboost," in *5th APWC on CSE.* IEEE, 2018, pp. 246–252.

[178] M. Athanasiou, K. Sfrintzeri, K. Zarkogianni, A. C. Thanopoulou, and K. S. Nikita, "An explainable xgboost–based approach towards assessing the risk of cardiovascular disease in patients with type 2 diabetes mellitus," in *IEEE 20th Int Conf Bioinform Biomed Eng.* IEEE, 2020, pp. 859–864.

[179] S. A. Hall, H. Copeland, A. Alam, and S. M. Joseph, "The "right" definition for post–left ventricular assist device right heart failure: The more we learn, the less we know," *Frontiers Cardiovas. Med.*, p. 962, 2022.

[180] Y. Zhang, R. Padman, and N. Patel, "Paving the cowpath: Learning and visualizing clinical pathways from electronic health record data," *J. Biomed. Inform.*, vol. 58, pp. 186–197, 2015.

[181] Y. Zhang, R. Padman, L. Wasserman, N. Patel, P. Teredesai, and Q. Xie, "On clinical pathway discovery from electronic health record data," *IEEE Intell. Syst.*, vol. 30, no. 1, pp. 70–75, 2015.

[182] F. Murtagh and P. Legendre, "Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion?" *J. classification*, vol. 31, no. 3, pp. 274–295, 2014.

[183] F. Masseglia, "Algorithmes et applications pour l'extraction de motifs séquentiels dans le domaine de la fouille de données: de l'incrémental au temps réel," *These de doctorat, Université de Versailles Saint-Quentin-en-Yvelines*, 2002.

[184] G. Ritschard, R. Bürgin, and M. Studer, "Exploratory mining of life event histories," in *Contemporary issues in exploratory data mining in the behavioral sciences*. Routledge New York, NY, 2013, pp. 221–253.

[185] F. A. Sonnenberg and J. R. Beck, "Markov models in medical decision making: a practical guide," *Med. Decis. Making*, vol. 13, no. 4, pp. 322–338, 1993.

[186] W. P. Dembitsky, A. J. Tector, S. Park, A. J. Moskowitz, A. C. Gelijns, N. S. Ronan, W. Piccione, W. L. Holman, S. Furukawa, A. D. Weinberg *et al.*, "Left ventricular assist device performance with long-term circulatory support: lessons from the rematch trial," *Ann Thorac Surg*, vol. 78, no. 6, pp. 2123–2130, 2004.

[187] A. G. Rizzieri, J. L. Verheijde, M. Y. Rady, and J. L. McGregor, "Ethical challenges with the left ventricular assist device as a destination therapy," *Philos Ethics Humanit Med*, vol. 3, no. 1, p. 20, 2008.

[188] J. K. Kirklin, D. C. Naftel, F. D. Pagani, R. L. Kormos *et al.*, "Seventh INTERMACS annual report: 15,000 patients and counting," *J. Heart Lung Transplant.*, vol. 34, no. 12, pp. 1495–1504, 2015.

[189] J. Ayres, J. Flannick, J. Gehrke *et al.*, "Sequential pattern mining using a bitmap representation," in *8th ACM SIGKDD*, 2002, pp. 429–435.

[190] F. Movahedi, M. K. Kanwar, and J. F. Antaki, "Timelines of adverse event journeys of lvad patients," *J. Thorac. Cardiovasc. Surg.*, in press.

[191] G. Pollock, "Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis," *J R Stat Soc Ser A Stat Soc)*, vol. 170, no. 1, pp. 167–183, 2007.

[192] J.-A. Gauthier, E. D. Widmer, P. Bucher, and C. Notredame, "Multichannel sequence analysis applied to social science data," *Sociol. Methodol.*, vol. 40, no. 1, pp. 1–38, 2010.

[193] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comp. Appl. Math.*, vol. 20, pp. 53–65, 1987.

[194] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," 1973.

[195] K. Y. Yeung, D. R. Haynor, and W. L. Ruzzo, "Validating clustering for gene expression data," *Bioinformatics*, vol. 17, no. 4, pp. 309–318, 2001.

[196] A. Ben-Hur, A. Elisseeff, and I. Guyon, "A stability based method for discovering structure in clustered data," in *Biocomputing 2002*. World Scientific, 2001, pp. 6–17.

[197] E. Levine and E. Domany, "Resampling method for unsupervised estimation of cluster validity," *Neural computation*, vol. 13, no. 11, pp. 2573–2593, 2001.

[198] C. Hennig, "Cluster-wise assessment of cluster stability," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 258–271, 2007.

[199] Y. Fang and J. Wang, "Selection of the number of clusters via the bootstrap method," *Computational Statistics & Data Analysis*, vol. 56, no. 3, pp. 468–477, 2012.

[200] J. Wang, "Consistent selection of the number of clusters via crossvalidation," *Biometrika*, vol. 97, no. 4, pp. 893–904, 2010.

[201] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J Am Stat Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.

[202] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[203] S. Helske and J. Helske, "Mixture hidden markov models for sequence data: The seqhmm package in r," *arXiv preprint arXiv:1704.00543*, 2017.

[204] A. Kilic, J. Macickova, L. Duan, F. Movahedi, L. Seese, Y. Zhang, M. V. Jacoski, and R. Padman, "Machine learning approaches to analyzing adverse events following durable lvad implantation," *Ann. Thorac. Surg.*, vol. 112, no. 3, pp. 770–777, 2021.

[205] L. Seese, F. Movahedi, J. Antaki, A. Kilic, R. Padman, Y. Zhang, M. Kanwar, S. Burki, C. Sciortino, M. Keebler *et al.*, "Delineating pathways to death by multisystem organ failure in patients with a left ventricular assist device," *Ann. Thorac. Surg.*, vol. 111, no. 3, pp. 881–888, 2021.

[206] B. Gregorutti, B. Michel, and P. Saint-Pierre, "Correlation and variable importance in random forests," *Stat. Comput.*, vol. 27, pp. 659–678, 2017.

[207] V. Svetnik, A. Liaw, C. Tong, and T. Wang, "Application of breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules," in *Multiple Classifier Systems: 5th International Workshop, MCS 2004, Cagliari, Italy, June 9-11, 2004. Proceedings 5.* Springer, 2004, pp. 334–343.

[208] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, 2002.

[209] G. Menardi and N. Torelli, "Training and assessing classification rules with imbalanced data," *Data Min. Knowl. Discov.*, vol. 28, pp. 92–122, 2014.

[210] J. M. Schaffer, J. G. Allen, E. S. Weiss, N. D. Patel, S. D. Russell, A. S. Shah, and J. V. Conte, "Evaluation of risk indices in continuous-flow left ventricular assist device patients," *Ann. Thorac. Surg.*, vol. 88, no. 6, pp. 1889–1896, 2009.

[211] H. A. Cooper, D. V. Exner, M. A. Waclawiw, and M. J. Domanski, "White blood cell count and mortality in patients with ischemic and nonischemic left ventricular systolic dysfunction (an analysis of the studies of left ventricular dysfunction [solvd])," *Am J Cardiol*, vol. 84, no. 3, pp. 252–257, 1999.

[212] S. Lakhdar, M. Nassar, C. Buttar, L. M. G. Perez, S. Akbar, A. Zafar, M. Munira, and A. Z. Gondal, "Outcomes with left ventricular assist device in end-stage renal disease: A systematic review," *Cureus*, vol. 14, no. 4, 2022.

[213] S. A. Silver, J. Long, Y. Zheng, A. B. Goldstone, D. Franz, T. I. Chang, and G. M. Chertow, "Outcomes after left ventricular assist device implantation in patients with acute kidney injury," *J. Thorac. Cardiovasc. Surg.*, vol. 159, no. 2, pp. 477–486, 2020.

[214] K. Klaeske, M.-T. Dieterlen, S. Eifert, U. Scholz, J. Garbade, K. Jawad, F. Sieg, M. A. Borger, and A. L. Meyer, "Device-induced platelet dysfunction in patients after left ventricular assist device implantation," *J. Thromb. Haemost.*, vol. 19, no. 5, pp. 1331–1341, 2021.

[215] T. Granja, H. Magunia, P. Schüssel, C. Fischer, T. Prüfer, D. Schibilsky, L. Serna-Higuita, H. P. Wendel, C. Schlensak, H. Häberle *et al.*, "Left ventricular assist device implantation causes platelet dysfunction and proinflammatory platelet-neutrophil interaction," *Platelets*, vol. 33, no. 1, pp. 132–140, 2022.

[216] H. Shah, A. Shaffer, R. John, M. Masotti, J. Schultz, V. Maharaj, and R. Cogswell, "Rising total bilirubin after cf-lvad implantation is highly associated with post left ventricular assist device mortality," *J. Card. Fail.*, vol. 26, no. 10, p. S47, 2020.