Project Report

Predictive Modeling Tool to More Effectively Transition Patients from Hospital

AUTHORS

Gulshan Sankhyan
sankhyag@mcmaster.ca

Minfei Wang
wangm32@mcmaster.ca

Premal Bhatt
bhattp23@mcmaster.ca

Acknowledgements

We would like to express our profound gratitude to our supervisor, Dr. Andrea Hemmerich, for her insightful and helpful suggestions while our research was being developed and written. We appreciate you taking the time and making the effort to ensure that we were making the correct choices and that our project was progressing in the right direction, as well as for your encouragement and regular checks on our work.

We are particularly grateful to Debbie Thibeau who was our only resource for understanding and learning the ALC process from the perspective of a social worker as well as a patient. Thank you so much for being so kind and patient in helping us understand the ALC process as a whole and for providing valuable feedback and suggestions for improving the Flowchart. Also thank you, Debbie, for connecting us with other social workers, the IT director, patients, and family members to interview and dedicate time to contribute to the development of our research work including the development of our Flowchart while also promptly replying to our many emails.

Additionally, we would like to express our gratitude to Dr. Robert Fleisig for helping us design our prototype, for his outstanding feedback and constructive criticism, and for guiding us in the right direction as we worked to complete the MEng Design program.

Our special thanks are to the staff at the SJHH for their valuable contributions. We are grateful to Jane Loncke for giving McMaster University students the opportunity to work with the Hospital and We extend our thanks to the many other SJHH staff whose feedback helped steer us in the right direction with our project. Finally, we would like to express our gratitude to the patients and their families for agreeing to be interviewed and for offering valuable concrete input on our prototype.

ENGINEERING
W Booth School of Engineering
Practice and Technology

McMaster University

St. Joseph's Healthcare Hamilton

Abstract

At St. Joseph's Healthcare Hamilton (SJHH), as of July 25th, 2022, 1 in 6 patients are Alternative Level of Care patients. Those who are clinically stable but awaiting discharge from the hospital because they require additional services and support are occupying 15% of the 887 funded beds at SJHH (Thibeau, D., Social Worker, 2022). From the interviews with SJHH staff, we realized that despite the patient being at the center of the process, the ALC(Alternate Level of Care) discharge process as a whole is not very patient-centric. Social workers and hospital professionals must constantly consider how much information to give patients. According to one social worker, "There is always a fine balance. If we give too much information they become paralyzed and don't know what to do." Patients feel frustrated because their treatment is finished but the Hospital is not letting them go home because they need additional care.

It becomes especially frustrating for patients when they are waiting to be discharged after they are given an ALC designation. They have just spent weeks talking with their Social Worker about selecting an ALC destination and now they have to wait for an unknown period of time until they can be discharged to that outside care destination. As Launi Greenspan, a Social Worker at SJHH, bluntly put, *"it is the fear of the unknown that stresses people out."* And true to her words, the longer the waiting time, the more stressful and agitated the patient becomes.

The Predictive Tool's main purpose is to provide more information, using the power of data and technologies like Machine Learning, to the patients so as to make their overall situation less uncertain. The tool's main features include:

- Provide a list of possible ALC destinations to a patient based on their situation.
- For each destination, estimate how long they have to wait in the Hospital until they can be discharged.
- Approximate cost for each destination so that they can compare the ALC destinations.
- Integration of the tool with MyDovetail in order for seamless sharing of information with patients.

We interviewed two Social workers, Sarah and Launi from SJHH to test the usability, desirability, and feasibility of the tool and the proposed solution received positive feedback. Due to barriers like the unavailability of information from third-party organizations like HCCSS (Home and Community Care Support Services) and privacy concerns from SJHH, we only have limited mock data to train our model instead of actual patient data. This solution has great potential and it is anticipated that the accuracy and effectiveness of the Machine Learning Model will increase with better input data. Although this project is still in the initial stage, we think we have put the project's trajectory in the appropriate direction for the following group of students to work on.

Glossary

- Algorithm: In computer programming terms, an algorithm is a set of well-defined instructions to solve a particular problem. It takes a set of input(s) and produces the desired output.
- Algorithms our model used: AdaBoost, LDA (Linear discriminant analysis), SVM (Support vector machine).
- ADL: Activities of Daily Living
- ALC: Alternate Level of Care
- CIHI: Canadian Institute for Health Information website
- ELOS: Expected Length of Stay
- Flowchart: Simplified translation of ALC journey/process in the form of a chart
- HCCSS: Home and Community Care Support Services
- IT: Information Technology
- KNN (K-Nearest Neighbors)
- LDA: Linear discriminant analysis
- LTC: Long Term Care
- LucidChart: Software used to create the Flowchart
- MatLab: Software used to make mathematical model
- MatLab model: Mathematical model that can generate predictions
- MRDx: Most Responsive Diagnosis
- PCA (Principal component analysis)
- SDM: Secondary Decision Maker
- SJHH: Saint Joseph Healthcare Hamilton
- SVM: Support Vector Machine
- SW: Social Worker

Table of Contents

Introduction

Marco(a persona based on our design reserch) is an 80-year-old man who lives with his wife. He has two sons who live in the same city as him. Marco recently started having trouble breathing so he went to a hospital. The doctor told him that he needs a tracheostomy where they will insert a tube in the trachea and he will breathe through the trach tube instead of through the nose or mouth. Marco and his family decided to go through it. Marco's eldest son Paul was a decision-maker for his father.

Since Marco needs additional care after the tracheostomy, he can't be discharged home. In the family meeting, Marcos's case is discussed with the physician, social worker, other care team members, care coordinator, patient, and family members in attendance. The Social Worker recommended a place in 'Ableliving' (https://ableliving.org/) which is a particular ALC destination which can handle patients like Marco. Marco's son Paul was not happy with this recommendation.

It is a misconception in the community that Nursing Homes and Long Term Care can handle any kind of Patient. But that is not true in this case because the staff doesn't have the training to handle cases like this. So, it is very likely that the Long-Term care facilities will reject Marco's application.

Paul started calling Long-Term Care facilities and their director of care by himself. And most of the time the directors will just tell any patient to apply to their facilities. Marco and his family will think that the LTC facilities can handle patients like him. This contrasting information creates an environment of distrust between the patient's family and the Social Worker.

Marco and his family now think that their Social Worker doesn't have their best interest in their heart. And now social worker not only have to regain their trust back but they also have to explain that the LTC facilities will accept his application but they will reject it after waiting for months, maybe even years as the LTC facilities won't review Marco's file until they have a bed available.

Begrudgingly, Macro and his family decided to accept their Social Worker's recommendation and now they are waiting in the Hospital until their file is forwarded to Home and Community Care Support Services (HCCSS). After HCCSS review their file and accept it the patient again has to wait until a bed is available in the Care facility where they are to be discharged.

The patient and their family members have no idea how long they usually have to wait in the hospital which leads to anxiety, frustration and deterioration of trust in the Hospital Institution.

Project Overview

The Predictive Modeling Tool is a software application to help Patients and Family Members during their ALC(Alternate Level of Care) journey. The tool achieves this by providing information like the possible discharge destinations, the Expected length of stay in the hospital before they are discharged and the cost associated with each destination.

After a patient is treated they might need additional care that a hospital can't provide. But the journey to getting discharged to these outside care facilities is very long and full of uncertainties for patients and family members. It is quite common for them to feel left out which not only leads to anxiety and frustration but also makes them distrustful of the authorities.

The social workers also face a dilemma of how much information they can give to the patients. Too much information might overwhelm them but if they are not given enough information the patient might feel that they are kept in the dark. Most of the time social workers shy away from making any promises because the ultimate decision of the discharge destination might come out differently.

This tool will help bridge this gap by providing a list of possible discharge destinations personalized to each patient. With this patients will not be bogged down by the overwhelming information and will easily be able to compare their options. The patient will be able to make better decisions during the ALC determination phase because they will be better informed and their waiting experience will also be less uncertain.

Alternate Level of Care

Patients like Marco are designated Alternate Level of Care (ALC) when they are medically stable but cannot be discharged to home as they require additional care.

According to the Canadian Institute for Health Information website (https://www.cihi.ca/en), *the Alternate level of care (ALC) is a system classification used in Canada that is applied in hospitals to describe patients who occupy a bed but do not require the intensity of services provided in that care setting.*

The Alternate Level of Care has three clinical states:

1. Stable and/or the patient's status has stabilized
2. Low risk of rapid decline.
3. No more searching for new additional diagnostics.

A patient must be designated as an Alternate level of care at that time by a doctor or his representative.

The ALC time frame starts on the date and at the time of designation as documented in the patient chart or record. The ALC time frame ends

1. at the time of departure from the ALC setting or
2. at the time the individual's care needs change such that the ALC designation no longer applies.

Alternate Level of Care at SJHH

As of July 25th, 2022, 1 in 6 patients at St Joseph's Healthcare Hamilton are Alternative Level of Care patients. Those who are clinically stable but awaiting discharge from the hospital because they require additional services and support are occupying 15% of the 887 funded beds at SJHH. Depending on the level of care, the patient will be assigned to one of the following discharge destinations, as shown in Figure 1. The left ring of the diagram displays the type of admission bed, and the right ring shows the discharge destinations. Out of these patients, 45% are waiting for Long-Term Care homes, and 17% are waiting for transitional care beds. Transitional care units are transient stops before the patients are transferred to Long-Term Care (Thibeau, ALC coordinator, 2022). Transitional care is separate from the hospital which is a consent-based bedded care program managed by Home and Community Care Support Services (HCCSS) within different retirement homes. However, at SJHH, the patients are recommended to go to transitional care before transitioning into Long-Term Care. The other 41% of designated ALC patients are waiting for a home with/out support, complex care, palliative care sites, rehabilitation, mental health acute or tertiary beds, retirement homes, residential care facilities, developmental services placements, etc

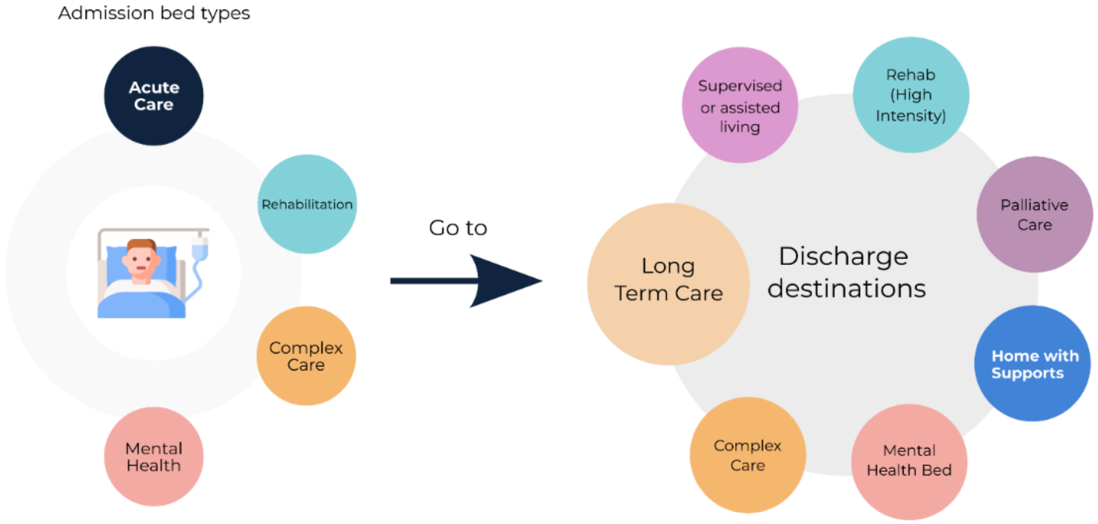


Figure 1: Diagram of admission bed types and Discharge destinations (Deilami et al., 2021, 12)

Project History

Project phase 1 Synopsis

In the first phase of the project from January 2020- August 2020, the cohort students Samuel Seffah-Nyame and Veera Ragavan proposed a decision aid tool to help patients be a part of the decision-making process with the assistance of a family member. The tool allowed the patients to have a fair idea of their options and choose their preferred destination based on their physical, cultural, and psychological needs, and learn about past experiences of other designated ALC patients. (Seffah-Nyame & Ragavan, 2020)

Project phase 2 Synopsis

In the second phase of the project from February 2021 - August 2021, the cohort students Negar Deilami, Venessa Almendariz and Tirenioluwa Ogunmoyero took the recommendations from the first phase of the project to dig deeper into the needs and insights of the key stakeholders and patients. Once these needs and insights were identified, they developed an interactive tool that provides benefits for patients, family members, and the SJHH healthcare team. However, they faced some limitations with the forecasting of the data and thus the Predictive Modelling Tool Project was initiated in coordination with the Interactive Tool project. (Deilami et al., 2021)

Project phase 3 Overview

The third phase of the project officially started in January 2022. Since this project was quite a bit different from its parent project the Interactive Tool project, we had to start everything from scratch. Initially, our main focus was to understand the ALC discharge process in detail along with the strength and limitations of the Predictive Modelling Technology so that we can come up with Patient-centric solutions.

Research Objective

As described in the Cressman Study (Cressman et al., 2013) that uncertainty is integral to the experience of ALC. This uncertainty plays a major role in the Patient's experience with Hospital Institution. The critical challenges could be summarized as below:

1. Individual Level - ALC is a long process and could result in making patients' conditions even worse if not handled carefully and sensitively. Patients might need emotional support in addition to physical care.
2. Organizational Level - Hospitals need to provide timely and accurate information to patients/families to manage uncertainty during the ALC process. The uncertainty could be mitigated by education and social support from caregivers to the patients. More frequent, better, and earlier communication is not available to the patients from the organization due to a lack of accurate information on ALC destinations and their length of stay in the hospital
3. Structural Level - The current structure is not efficiently able to manage the LTC(Long-Term Care) waitlist, again due to the lack of proper coordination in sharing information on patient ALC destination and their length of stay data as this data is managed by a third party organization (i.e. HCCSS) resulting in uncertain wait times for patients.

In a nutshell, the following issues regarding the communication gap and access to relevant information were discovered as a result of our efforts to understand the patient's experience to improve the effectiveness of the ALC process.

For patients and family members:

- The timeframe for difficult conversations around discharge planning for patients and family members doesn't match the Care team's timeframe
- Information overload leads to difficulties in making informed choices
- Uncertainty during the waiting time after ALC designation.
- Financial barriers to effective discharge planning

For the Hospital

- Uncertainty regarding the availability of hospital resources (beds) due to data being collected by different sources(hospital, HCCSS, government, etc.)
- A mismatch between the expectations of the patients and their family members and the reality of the healthcare system.

Stakeholders Research

Stakeholder's map

The purpose of the stakeholder research stage is to understand who is involved in the Designated ALC patients' hospital journey at SJHH. This stakeholder map was a revised version of the one created by previous students. (Deilami et al., 2021). And as per our project expectations, the stakeholders were changed in the circles depending on their involvement in the project. The previous version was essentially for the Interactive tool, while for the Predictive tool, the involvement of the parties involved changed slightly and so did our Stakeholder map.

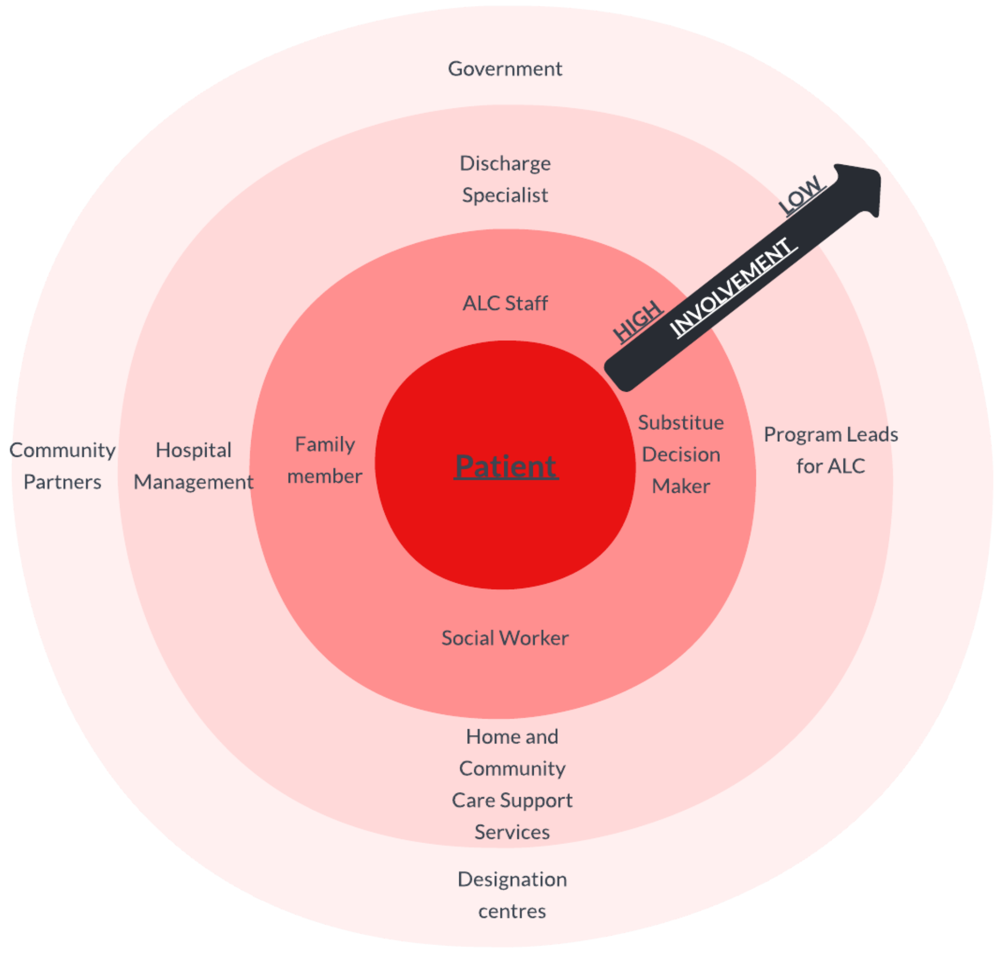


Figure 2: Stakeholders Map

Patient

A patient is designated Alternate Level of Care when they no longer require the services of a specific hospital location and is ready to be discharged. However, there are obstacles or waiting periods for services/resources to be arranged for discharge.

Family Members/Substitute Decision Maker (SDM)

Family members can have the role of caregiver, substitute decision-maker (SDM), or general support for the patients. They are crucial to the patient's journey, and in some situations where patients are deemed to be unable to make treatment decisions, such as when there is a cognitive impairment, they are fully responsible for them.

Social Worker and Discharge Specialist

Social workers and discharge specialists meet with patients/families and collaborate with the healthcare team, Home and Community Care Support Services (HCCSS) and other community partners to ensure that the patients are assigned to the Most Appropriate Discharge Destination (MADD). From admission until discharge, they support patients and family members and assist them in navigating their way to the next stage of treatment. They assist with counseling, discharge planning, and information on available local services, residential care facilities, co-payments, and financial matters. It is significant to note that discharge specialists typically intervene in complex cases where there are barriers to discharge planning, such as disagreements over the safety of the patient's release plan or conflicts between patients, families, or the medical staff.

The Healthcare Team

The healthcare team is usually responsible for assessing the patient and giving recommendations for what level of care is best for them. Patients designated with ALC must get consultations and recommendations from the Most Responsible Physician (MRP).

Depending on the patient’s needs, the members of the healthcare team could be at least one of the following:

- Nurse Practitioner (NP) and Registered Nurse (RN)
- Personal Support Worker (PSW)
- Physiotherapist and Physiotherapy Assistant•Occupational Therapist (OT)
- Recreation Therapist (patient leisure needs and helps to find ways to meet these needs)
- Dietitian and Diet Technician
- Speech and Language Pathologist (SLP)

[Adopted from "Welcome to the Alternate Level of Care Unit (ALC), Providing Patient and Family Centered Care" from St. Joseph's Healthcare Hamilton]

Home and Community Care Support Services (HCCSS)

Home and Community Care Support Services is the new name for Local Health Integration Networks (LHINs). HCCSS plays an important part in Ontario's Healthcare system as they make sure people have access to healthcare at home and in the community. Care coordinators and home care coordinators are the two categories of coordinators at SJHH. In most cases, care coordinators are employed by hospitals and receive referrals to assist with identifying home care assistance and services and/or determining patients' eligibility for long-term care. In order to ensure that patients are transitioned back into the community, the home care coordinators collaborate with patients/families, healthcare teams, and Care coordinators even though they are not employed by the hospital.

Community Partners

Community Partners refers to for-profit or nonprofit businesses like Bayshore Healthcare that collaborate with hospitals to address the difficulties of ALC by providing resources and extra beds.

Program Leads for ALC

Refer to those who hold leadership roles and are in charge of coordinating and managing resources and services.

Designation Centers

Depending on the patient's status, their discharge destination depends on the available destination centers such as Long-Term beds, Mental Health, Supportive housing, etc. as demonstrated in Figure 1.

Government

The government plays an indirect role in the journey of designated ALC patients due to the decisions being made and funding. For example, during the pandemic, one of the decisions made was to move patients to Long-Term Care without their consent which affected many families (Herhalt, 2021).

Target Users

Patients will be the direct target users for our prediction model. They will benefit from the predictions generated by the model as they will be given a list of recommendations for probable ALC destinations, their expected length of stay in the hospital, and approximate costs associated with each ALC destination. And thus in turn patients will be able to make better and more informed decisions due to the information available to them.

Hospital Administration will be the secondary target users and will benefit in a way that the system will be better able to manage their resources like beds, allocation of human resources, and treat more patients with lesser wait times due to the predictions from the model which ultimately will help to improve the patient experience.

Interviews

We kick-started the project by conducting a series of Interviews. Although, our major interviewee was Debbie Thibeau, a Social Worker at SJHH, who we met every two weeks. Additionally, we interviewed Dr. Marjan Alavi who is an Assistant Professor and Program Lead at M.Eng. Manufacturing Engineering at McMaster University. Debbie Thibeau also helped us connect with Sonia Shiels who is a Manager, Coding, Data Quality, Reporting & Documentation Improvement at St. Joseph's Healthcare Hamilton; and Kevin Fatyas who is Data Scientist at SJHH IT Department.

Interview with Debbie Thibeau - Social Worker at SJHH

We met and spoke with Debbie, a social worker, once every two weeks. She contributed significantly to deepening our understanding of the ALC experience. She helped us understand the whole ALC journey with great detail and information. This knowledge was translated into the form of a Flowchart to make it easier for everyone to understand this very complex process of ALC. We have worked on several iterations of this flowchart and with every iteration, the ALC process was made more accurate, simplified, and easy to understand.

Empathy Map

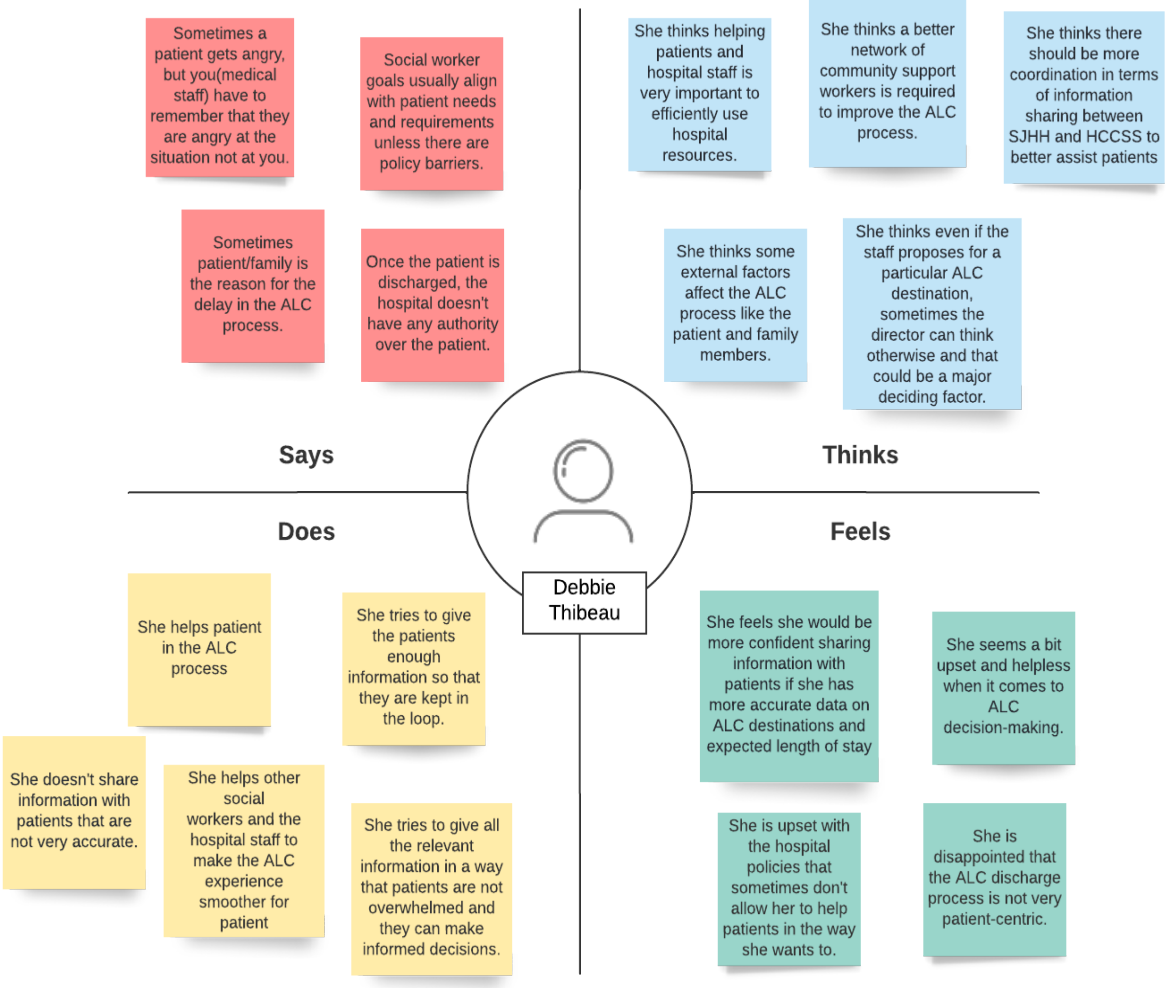


Figure 3: Empathy Map for Debbie Thibeau

Insights

- She is a very emotional person which makes it easy for her to connect with and understand her patients. Years of working in the Medical Sector had made her intimately familiar with all the faults of the ALC Process. Both of these factors made her very invested in improving the patient's ALC experience.
- She sometimes feels helpless about decision-making for ALC patients and that makes her upset.
- Better network of community support could help to make the ALC process more efficient and smooth.

Interview with Dr. Marjan Alavi - Faculty and Chair MEME

Since Dr. Marjan Alavi has subject matter expertise in Digital & Smart Systems along with Machine Learning Technologies, we interviewed her to get her insight on our project.

Empathy Map

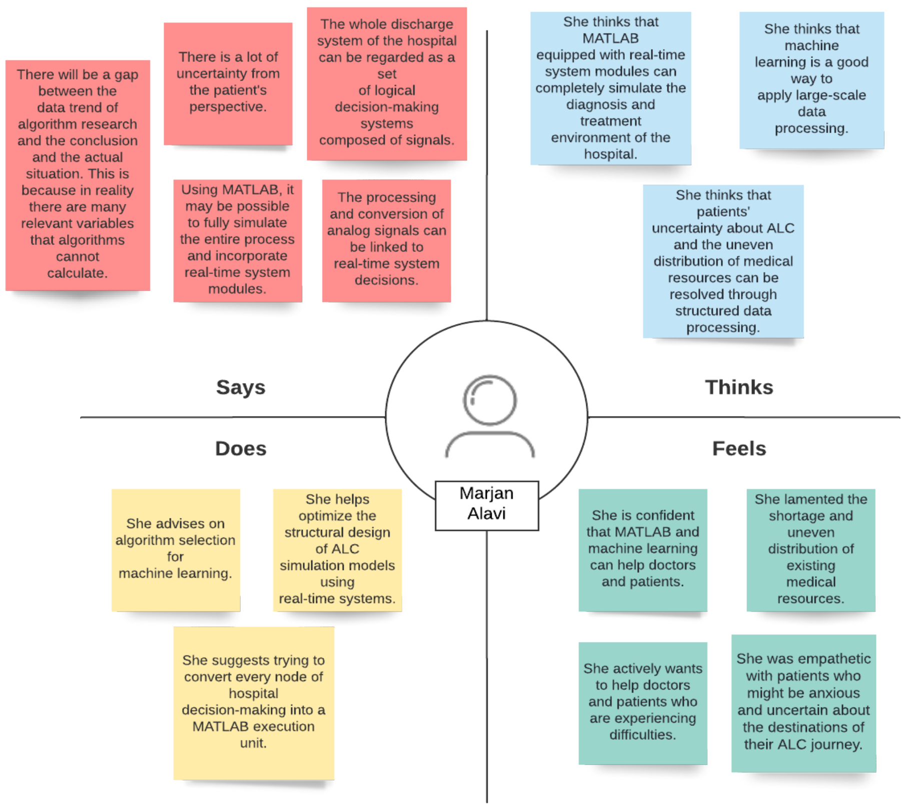


Figure 4: Empathy Map for Dr. Marjan Alavi

Insights

- In the first meeting, after listening to our project she suggested we work on a MatLab simulation model that can translate the patient journey and understand how each ALC decision is made. This suggestion was a godsend to us because at that time we didn't have any patient's data related to ALC so coundn't work on a Machine Learning Model. She advised us to treat the ALC journey as a set of if-else points and create a simulation model in MatLab.
- In the second meeting, She pointed out the importance of Quality Data for the success of the Machine Learning Prediction Model. Because accuracy of the modal is directly proportional to the quality of data.

Interview with Sonia Shiels - Data and Coding Manager,SJHH.

We interviewed Sonia Shiels to understand how Patient's data is handled by the Hospital.

Empathy Map

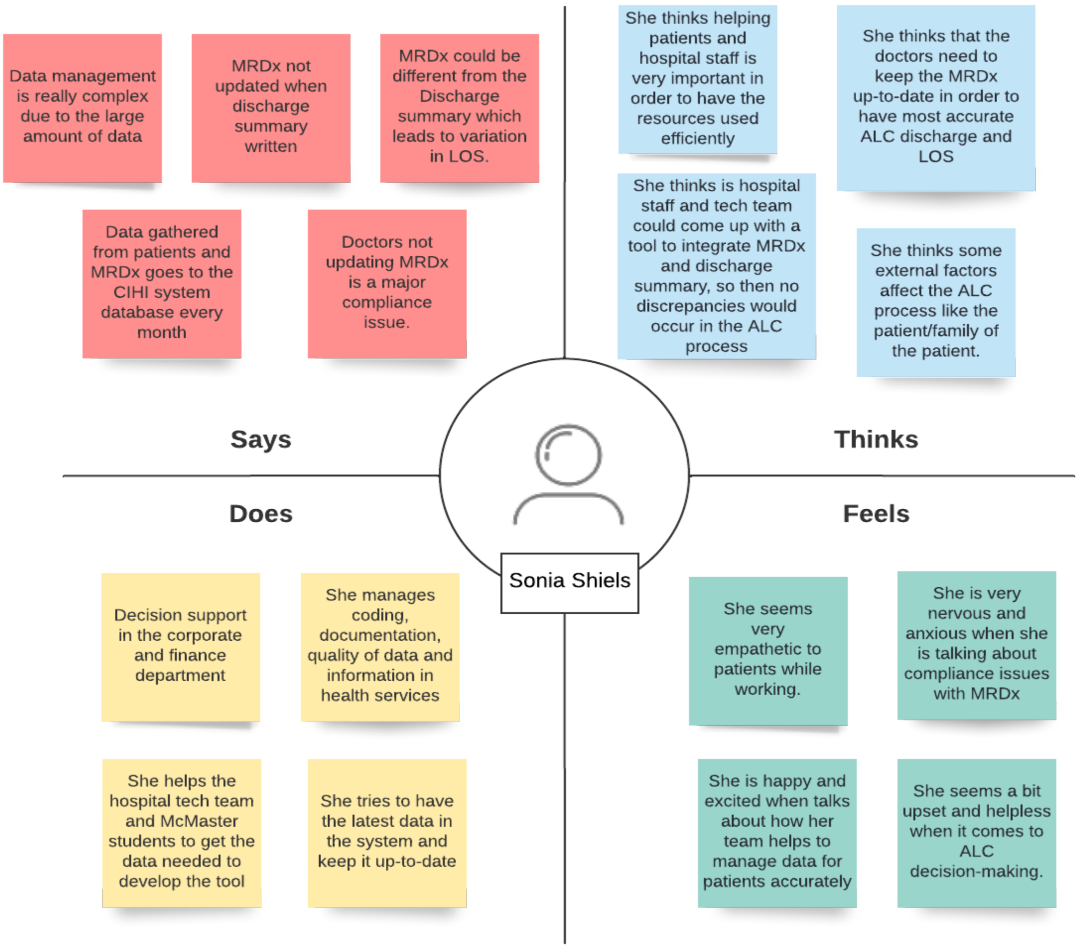


Figure 5: Empathy Map for Sonia Shiels

Insights

- MRDx(Most Responsible Diagnosis) needs to be up-to-date for accurate discharge summary leading to accurate Length of stay.
- Patient/Family of the patient due to their explicit choices about some destinations, are also some external factors affecting the Length of Stay leading to increased wait times for those destinations.

Interview with Kevin Fatyas - Data Scientist, SJHH IT Dept.

Kevin is a Data Scientist at SJHH and we had the pleasure of meeting with him twice and sharing our work and getting suggestions and feedback on it.

First meeting

When we joined this project we were given a document which contains technical details about the Protection Model that SJHH developed previously. Kevin helped us understand that model in great detail. Although he was not one who developed it, with his subject matter expertise he was able to explain the key points effectively. Some of them are:

- He explained what are the different factors that the model took into consideration and how much effect each one has on the outcome. For example, the factors listed below have a major impact on whether a patient is ALC or not
    - ADL (Activities of Daily Living) which signifies a patient's ability to perform daily tasks like taking a bath,
    - Cognition (Gradual Cognitive decline occurs more commonly in ALC patients),
    - Household Income, etc.
- The kind of algorithms were used to train the model and how do they compare to each other. He told us that they tried Random Forest and XGBoost Algorithms and XGBoost performed better with less Test-error
    - Random forest
        - Test-error: 0.078
    - XGBoost
        - Test-error: 0.065
- He also explained the limitations of this model. The major one is that the model can only predict if a patient will be given an ALC designation or not. It doesn't say anything about which type of ALC designation or the waiting time (Expected Length of Stay).

After interviewing him, we got to know the reasoning behind why the model is not currently in use. The model can only predict if a patient will be given an ALC designation or not. It was unable to predict the type of ALC or the length of time the patient would have to wait before being allowed to leave the hospital

Second Meeting

In the second meeting, we questioned him about the software, algorithm, and data SJHH uses to train its prediction model. He told us that the IT (information technology) team at SJHH is quite familiar with working in Python and Javascript (both software programming languages) and that we are free to use any one of those languages to develop our prototype. He also gave mock data which we used to develop a Machine learning model in MatLab software. Key points to note -

- The mock data was not very accurate, so even though we achieved high accuracy in predictions, that can not ensure the same results with actual patient data.
- The actual patient data was not possible to be shared by the hospital due to privacy concerns.
- Kevin mentioned that if the Prediction model was to be developed in python (a programming language), it would be much easier for the hospital IT team to integrate the model with Dovetail (their current and already in use software platform).

Design Research Synthesis

After conducting a series of interviews we identified several key findings related to the patient's discharge experience.

Findings	**Insights**
Uncertainty and Lack of Information	There is a shocking lack of information available to patients as they are waiting to be discharged to an Outside Care facility. *"Because it is the fear of the unknown that stresses people out."* Launi Greenspan (Social Worker) *"Patients are frustrated with vagueness and non-transparency between them and the care team"* Sarah Huckle (Social Worker)
Communication Barrier	The final ALC decision mostly depends upon the hospital's upper management. Social Workers prefer not to give sureties to raise Patient's expectations in case it leads to disappointment. Overall, this leads to a communication barrier. *"Sharing info after approval often leads to mistrust in their social worker"* Sarah Huckle (Social Worker)
Long Waiting Time	Patient has to wait: ● For HCCSS to approve their ALC file forwarded by their Social Worker. ● And for the bed to be available in the outside care facility so that they can be discharged from the Hospital. In both cases, the waiting time could be quite long.
Decentralized discharging system	SJHH works closely with HCCSS to handle patient discharge, but they are still different organizations. This leads to a situation where it is hard to collect accurate information.
Lack of use of the Existing Prediction Model	The existing Prediction Model is not in use because it is quite limiting in its predictions. It can only predict if a patient will be ALC or not

Initial Design

ALC Journey prototypes

As mentioned earlier, we meet with Debbie, a social worker, every two weeks. And in those meetings, we usually ask questions related to the ALC process itself which we then translated into an ALC flowchart which we iterated continuously. With the different iterations of the ALC process, we tried to not only deepen our own understanding but also simplify the process so that anyone can understand it easily.

First Iteration

This was the first iteration of the flowchart. It starts with the patient getting admitted to the hospital for acute treatment. All the blue boxes represent a process, while all the diamonds represent a decision-making step.

After the acute treatment, the patients are accessed if they need any additional medical and care needs. These needs are determined by asking conditional questions in the diamonds. Medical/care needs were categorized into 4 types - Medical/functional ability, Self-care/Activities of Daily Living, Financial status, and Functional Mobility.

Based on this assessment, the patient is being evaluated for being an ALC patient and determined if Home and Community Care Support Services can meet the care needs or not. Either a referral is sent to HCCSS or the Home First plan is discussed with the Patient/Family members. After that, the consent is taken from the Patient for the agreement to the discharge plan and the ALC order is initiated.

Based on 6 other conditions described in the flowchart, the final ALC destination, and then the patient is discharged based on the availability and waiting times for that ALC destination. These six factors are the most critical for deciding which ALC destination the patient will be designated to.

This was the simplest form of flowchart developed based on the meetings with Debbie and her explanation.

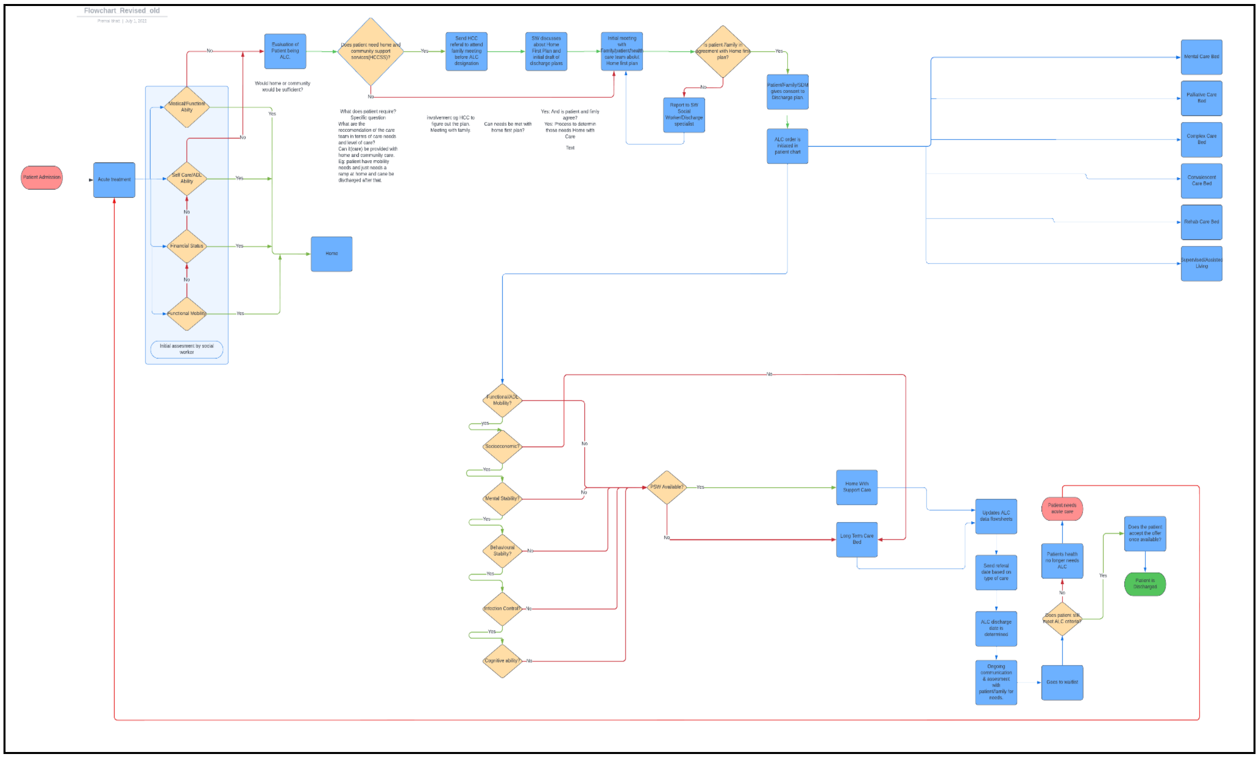


Figure 6:First Iteration Link

Second Iteration

Based on the feedback of Debbie on the first iteration, a few changes were made to the second iteration. The base structure remained the same. The changes made were to make the flowchart more detailed and easily understandable at each step.

Since we decided to go with only the most critical ALC destination - The Long Term Care, for our project, we developed the flowchart focusing only on that destination.

The major change from the first iteration was the addition of the Escalation process. There are 3 situations where the process could be escalated -

- If the patient/family members are not in agreement with the home first plan, then the process is escalated to the Discharge specialist to consider the option of the patient being discharged to a long-term care destination.
- If the care/medical needs could not be fulfilled by Home and Community Care Support Services then the process is escalated.

The critical factors deciding the ALC designation are the same as in the previous iteration.

We also revised the terms in the flowchart to represent the exact Medical terms that the Care team uses so that the flowchart could be the most exact representation of the ALC process.

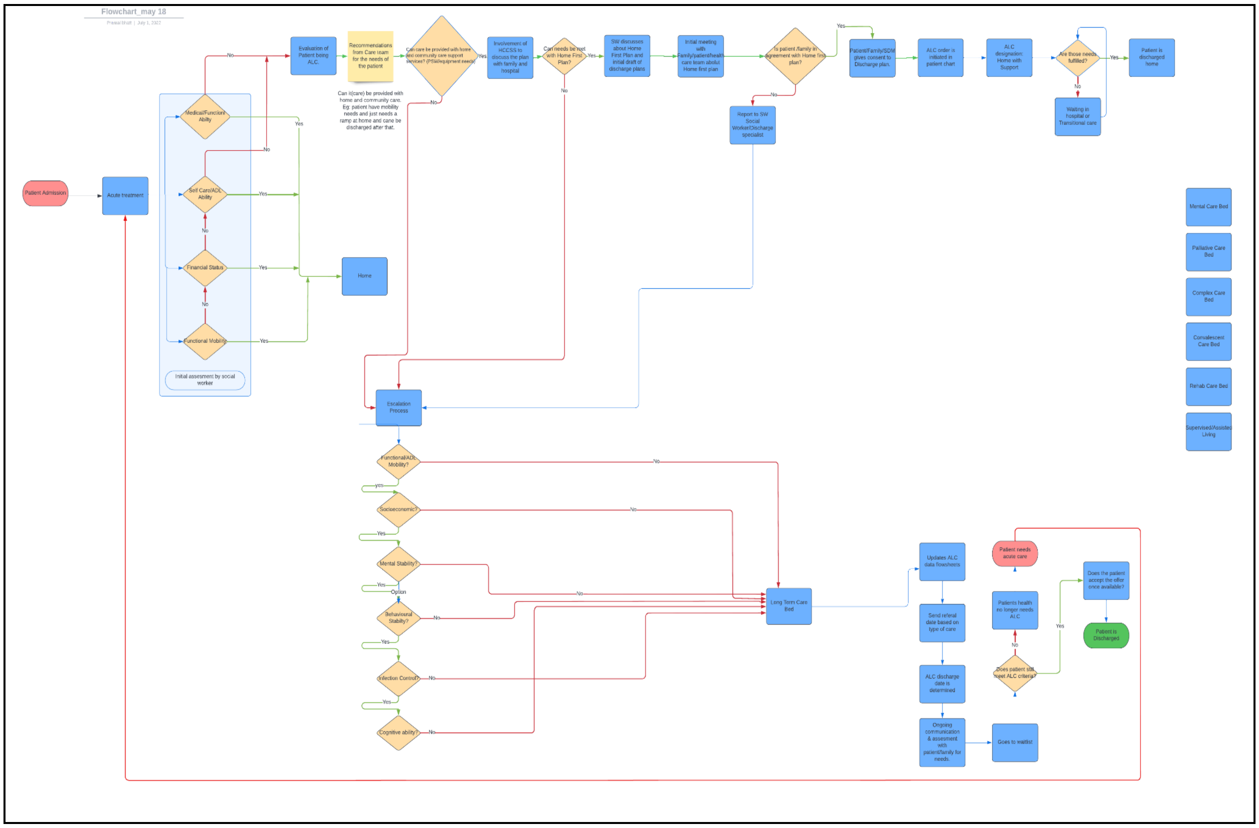


Figure 7: Second Iteration Link

Third Iteration

For this iteration we tried to simplify the flowchart visually and make it more comprehensible. We divided the ALC process into different color-coded stages from patient admission to discharge.

Also in the Waiting phase, we added the transitional wait phase which wasn't separately shown in the second iteration. We marked all the stages where the patient experience was improved by their involvement in the decision-making process by recording their opinions/consent for the discharge plans made. Additionally, we added sticky notes in steps to explain them in a more detailed manner to make it easier for the reader to understand. The easier it is to understand the process, the easier it is to improve the patient experience during the ALC process.

The basic structure is still the same as the first iteration. But each iteration added more details to clearly explain the ALC process and simplify it to a level where a person from a non-medical field could also understand it very easily.

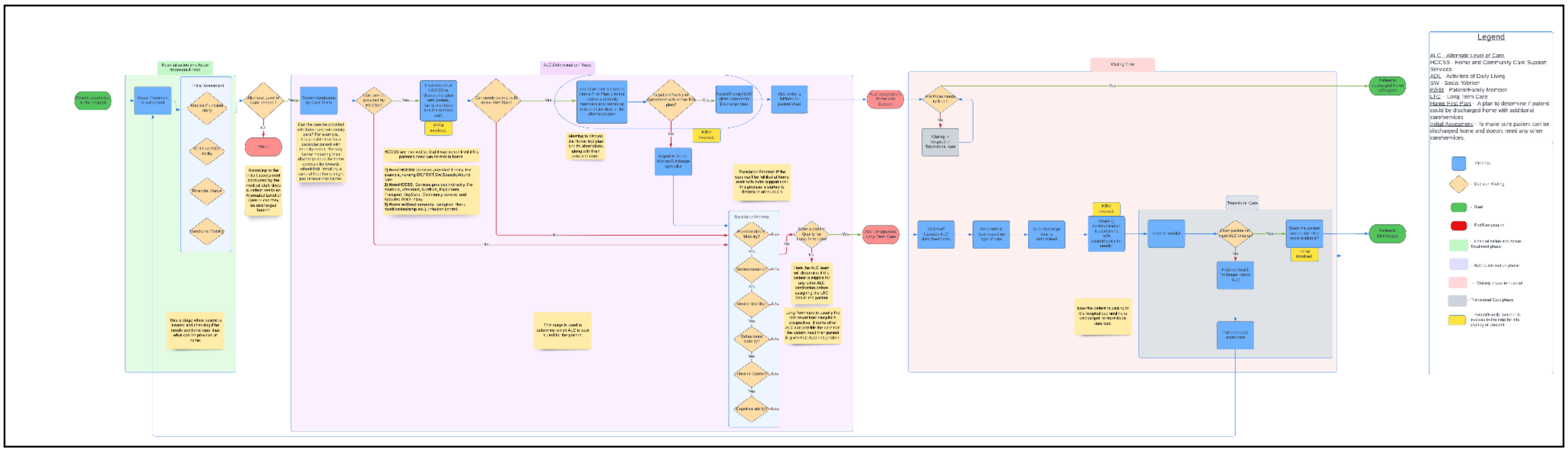


Figure 8: Second Iteration [Link]

ALC Simulation Model

For that, we used Simulink in MATLAB to simulate the actual flow of patients in the hospital and the branch selection that finally enters the destination of ALC patients. The final destination includes several options. The final destination of each individual patient is different. Patients may be able to go home directly or may need to be admitted to ALC care. At each option, the node is a restoration of the patient's journey in the hospital. By entering yes/no in each option, you can influence the final destination. The whole model is a huge logic gate option conditional signal processor. Healthcare workers can make predictions about a patient's final destination by stimulating the patient's experience. Patients can also use this way of changing different conditions to understand their future arrangements and possible trends.

The biggest benefit of this simulation was that it helped us in asking the right question to Debbie so that we know how the patient reaches a particular ALC destination. This simulation model is a reconstruction and restoration of hospital processes, using a digital form to Reflect the hospital's decision logic in the software for a comprehensive understanding of a patient's journey.

First Iteration

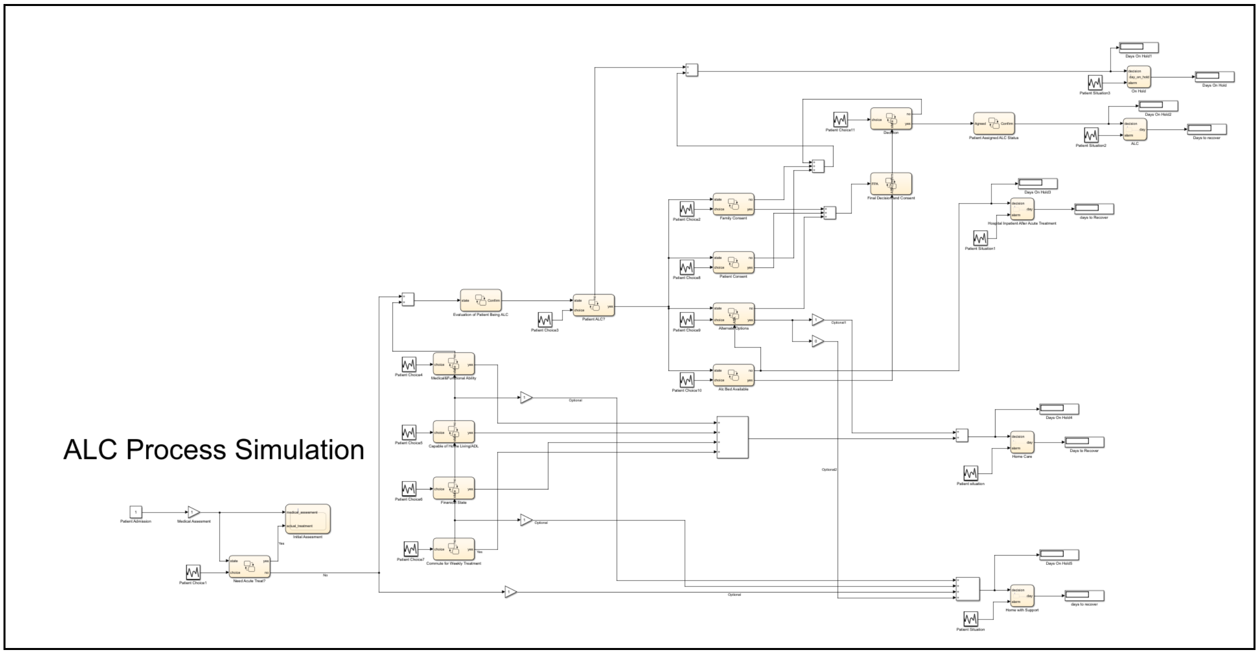


Figure 9: First Iteration of MatLab Simulation

The simulation model is designed in MATLAB according to the original flowchart model. In iteration 1, the entire simulation model uses the MATLAB instruction system equipped with the Stateflow module. The Stateflow module provides a graphical language solution including state transition diagrams, flowcharts, state transition tables, and truth tables. Users can use Stateflow to describe how MATLAB algorithms and Simulink models react to input signals, events, and time-based conditions. Stateflow supports the design and development of supervisory control, task scheduling, and hybrid systems. Input commands use random variables. Through multiple simulations of random variables, the simulation model ran various scenarios throughout the ALC journey. The final result shows the final ALC target for each run.

Second Iteration

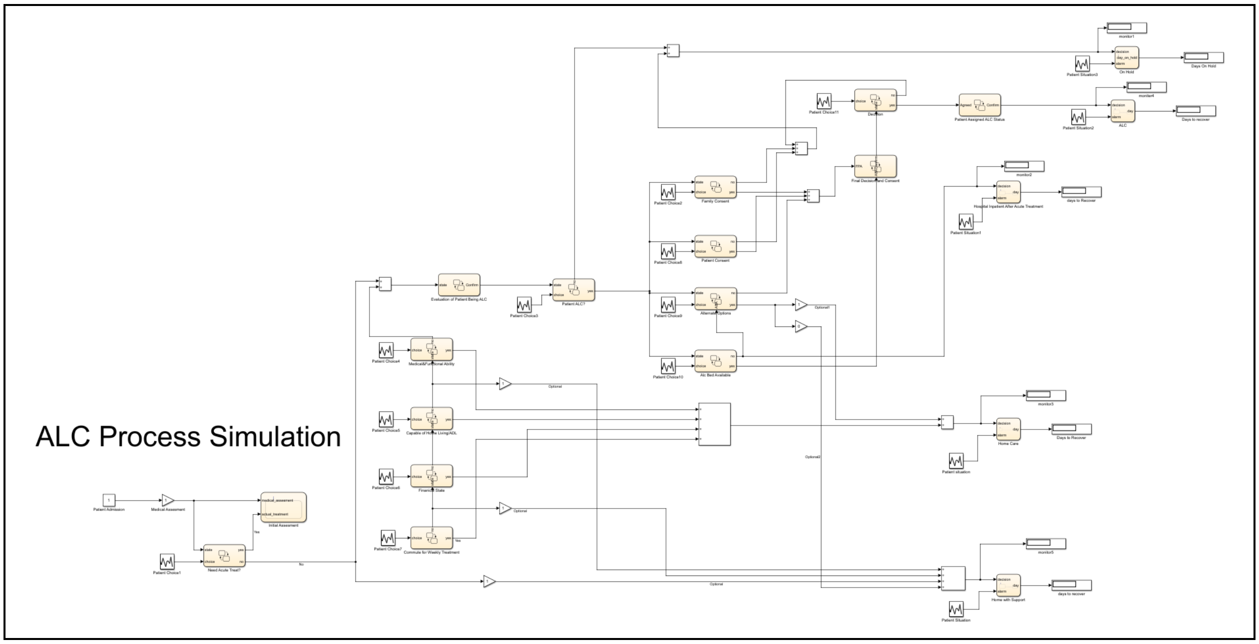


Figure 10: Second Iteration of MatLab Simulation

In iteration 2, the overall design structure remains the same as in iteration 1. In this iteration, the simulation model adds a function of the patient's time during treatment as a reference to the patient's time in the hospital. Through the simulation algorithm, the output results of the length of stay in the hospital and the destination of discharge can be given. The destination of discharge has on hold (no judgment has been made for the time being), ALC (entering the ALC process), Home Care (self-care by the patient and family members at home), Home with Support (instructed by professional nursing staff at home) and Hospital Inpatient After Acute Treatment. This further enhances the effective cognition of the user's ALC journey and increases the scope of use of this simulation model.

Third Iteration

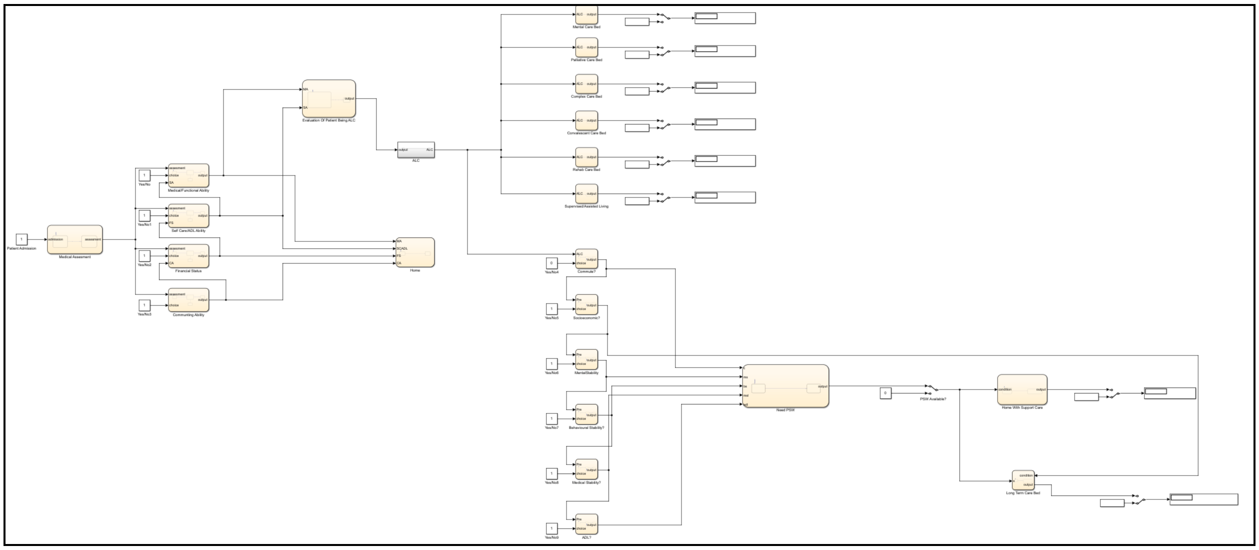


Figure 11: Third Iteration of MatLab Simulation

In iteration 3, the simulation model design was consistent with the revised flowchart, and a lot of optimizations and innovations were made in the overall structure to better fit the execution of the actual ALC journey. In this iteration, all independent variables are changed from random input to the form manually entered by the user. The user can freely adjust each input according to the actual situation of himself or the patient. The system can output the final destination and estimated discharge time through the algorithm.

Machine Learning Model

Using MatLab's simulation model to generate prediction was one way we approached this problem, but that was not the only one.

After we were provided with the patient's mock data, we started looking into ways to use it. We were given 500 rows of (fake)patient data.

Data Type (As in dataset)	How did we use it?
Patient's Number (fake)	Discarded this because it was just an identifier unique to the patient
Admission Date Time	Discarded because we were unable to implement it in our model
Discharge Date Time	Discarded because we were unable to implement it in our model
Discharge Specialist (fake)	Discarded this because it was just an identifier unique to a discharge specialist
Hospital Service	The service for which the patient received treatment during their most recent hospital encounter. We tried to use it but the model gave errors. There could some problem in the code which we were not able to solve.
Discharge Disposition	Where the patient went post-discharge. This was the output of the model.
Patient's level of Care	The level of care of the patient at the time of discharge. Although this column contained useful data like ALC - Palliative Care Bed, ALC - Long Term Care etc., it also had a lot of irrelevant data like 'Newborn'. We don't know how to handle the irrelevant data. So, we discarded this.
ALC Designation	Whether the patient was given an ALC designation or not. This was one of the inputs. 0 means Non ALC 1 means ALC
Deceased	Whether the patient passed away before discharge. This information was one of our input.

Financial Capability	Whether the patient has payment coverage (including OHIP). This was one of the inputs.

But after analyzing the data more deeply we realized that the data doesn't contain enough relevant information that can be used to generate a prediction.

It doesn't contain information like

- the reason behind patient's visit to hospital,
- Age,
- The type of ALC designation, If a patient was given an ALC designation,
- Ultimate discharge destination
- How long it took the patient to be discharged after given an ALC

Nonetheless, our goal was not to build the most accurate prediction model. We just have to do enough to prove the feasibility of the Prediction Model. The only information that we were able to use was the patient's level of care, ALC designation, and Financial Capability. Additionally, when we explored the data for these fields we realized that more than 80% of the data of the patients who were discharged Home. So, if our model just predicted Home as the ALC designation there is a pretty high chance that it will be true.

Still, we went along with developing a Python machine Learning model because the effectiveness of the model can be increased with better data later on.

We used AdaBoost, an adaptive boosting machine learning method, algorithm to develop this Machine Learning Model. Although we gave algorithms like Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) some thought, we eventually chose AdaBoost because, when these algorithms were tested on mock data, AdaBoost performed better. For more details regarding these algorithms and how we compared them along with the instructions on how to set up the programming environment refer to Appendix 1.

Testing the model

Let us consider the hypothetical case of Paul who is suffering from Diabetes and looking for Alternate Care.

For his case, we can pass the following data to our model.

```
InputData = [1,0,0]
```

These values represent his level of care at the time of discharge is ALC-related

ALC Designation	1 means patient designated as an ALC patient.
Deceased	0 means the patient is alive while getting treated.
Financial Capability	0 means they can’t cover their bills

The program cross-compared the three algorithms and obtained the result [0] as shown in the image below.

```
Adaboost:
Accuracy: 0.952
The Confusion Matrix is
[[118   0   0   0   0   0   0]
 [  1   0   0   0   0   0   0]
 [  1   0   0   0   0   0   0]
 [  2   0   0   0   0   0   0]
 [  1   0   0   0   0   0   0]
 [  0   0   0   0   0   1   0]
 [  1   0   0   0   0   0   0]]
Prediction Result: [0]
Reference: [1]
Training time (second): 0.004242100000055871
Testing time (second): 0.000650099999916165
```

The predictive modeling gave [0] as the Prediction Result. This means that Paul's final destination is "Discharge Home without Support Services"

Although the model is very limited as it is only taking three values as input which can't be compared to the real-world scenario, still, this Machine Learning model proves the feasibility of developing an effective Predictive modeling tool and it is anticipated that the predictive accuracy will increase with better input data.

Final Design

Final Iteration of Flowchart

Based on the feedback on the third iteration by Debbie, this was the most updated version of the flowchart. The major change was around the Patient hospitalization after they are ALC designated. The patient's ALC would reinitiate once the patient recovers within 48 hours of hospitalization. They don't go back to the initial acute treatment step in the process which was misunderstood while developing iteration 3.

Also, a clear legend and a list of all the abbreviations used were added to make it more clear and simplified for the readers to understand this process. All the terms were checked to closely represent the actual medical terms that the Care team currently uses in the hospital with the patients as per the feedback and suggestions from Debbie. Critical Factors in determining ALC-LTC eligibility as shown in the diamonds can be listed as follows:

- Functional/ADL (activities of daily living) Mobility
- SocioEconomic Stability
- Mental Stability
- Behavioral Stability
- Infection Control (Specifically for COVID)
- Cognitive Stability

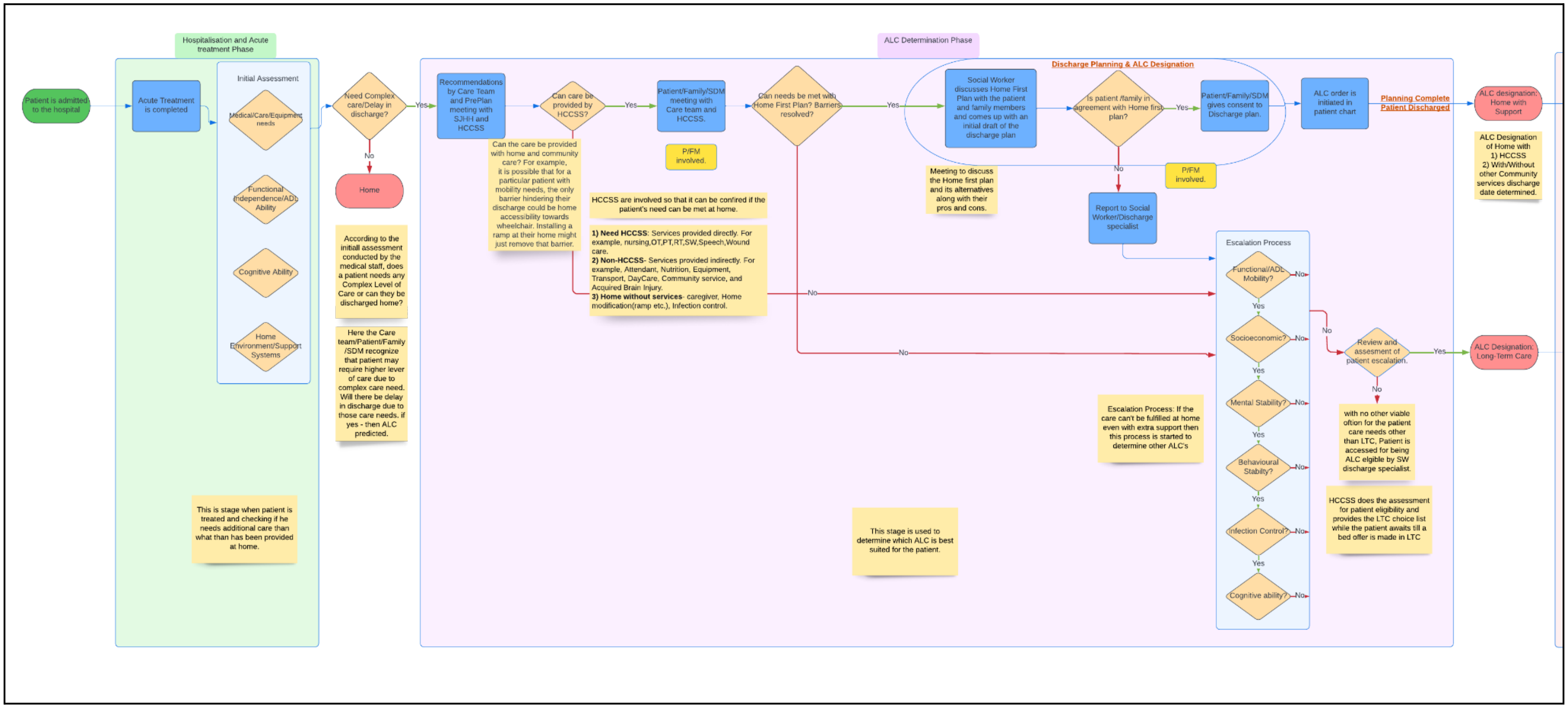


Figure 12: Final Iteration of Lucid Flowchart (1/2) [Link]

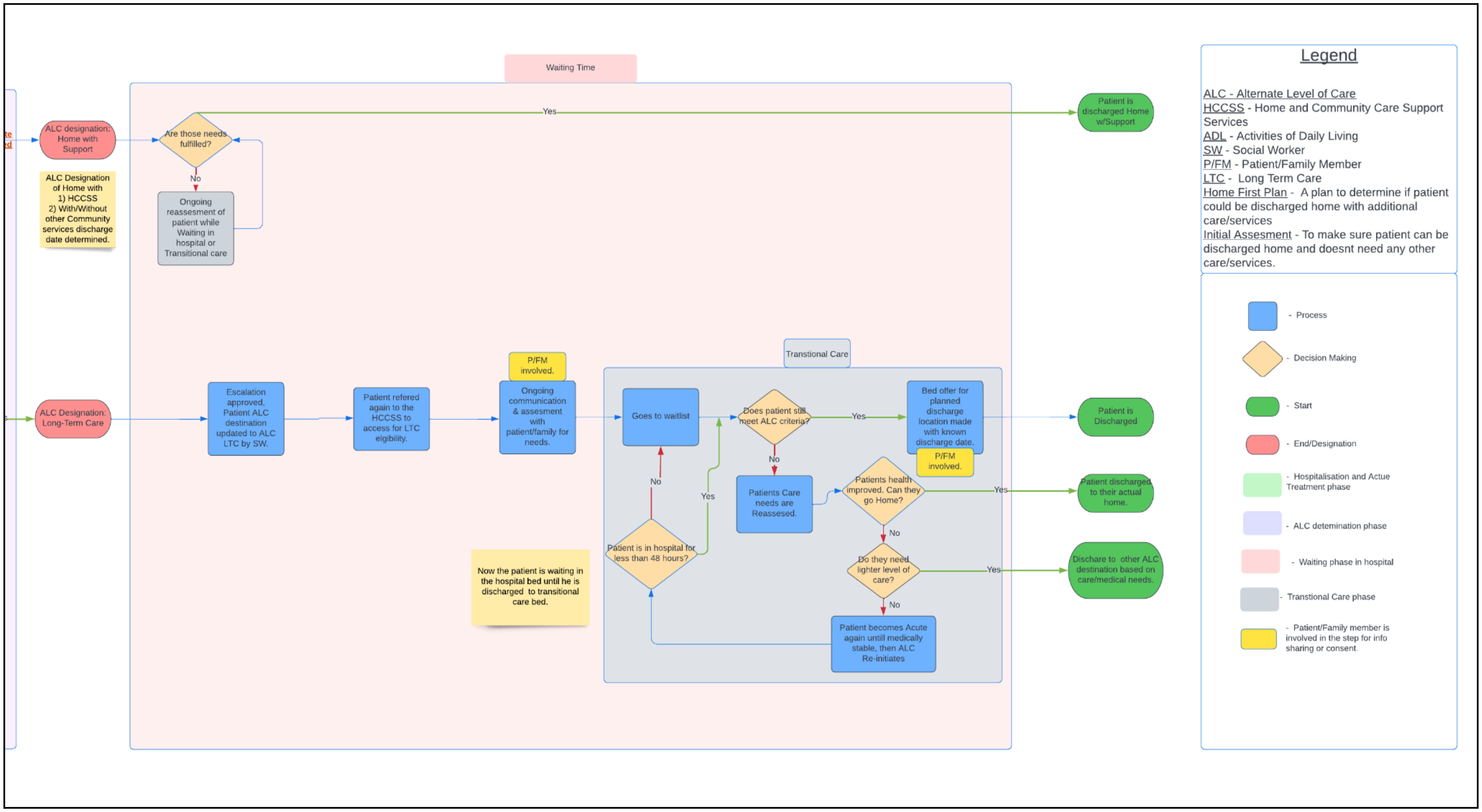


Figure 13: Final Iteration of Lucid Flowchart (2/2) [Link]

Phase Wise Separation for better Clarity

Hospitalization Phase

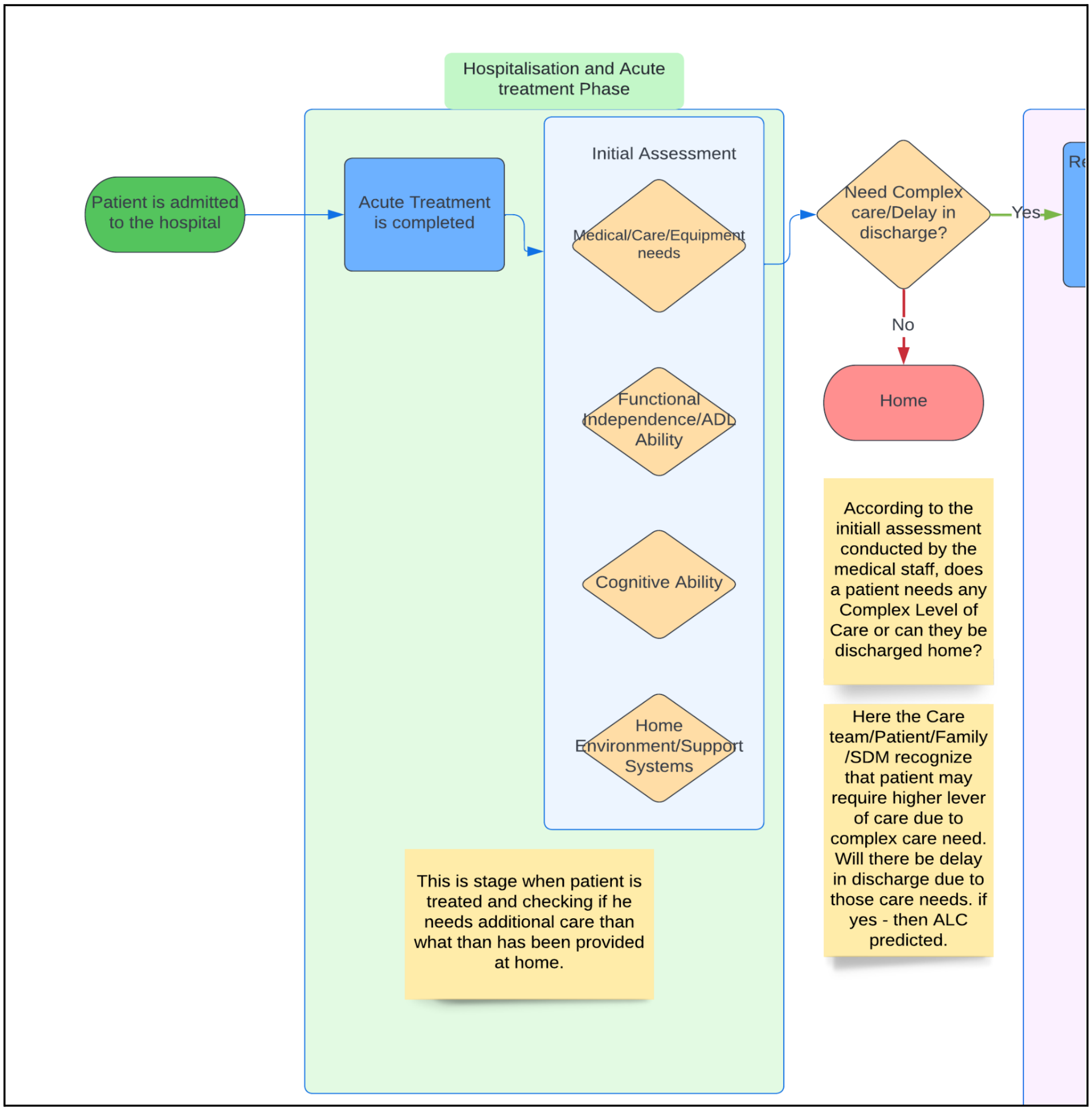


Figure 14: Final Iteration- Hospitalization Phase (1/3) Link

ALC Determination Phase

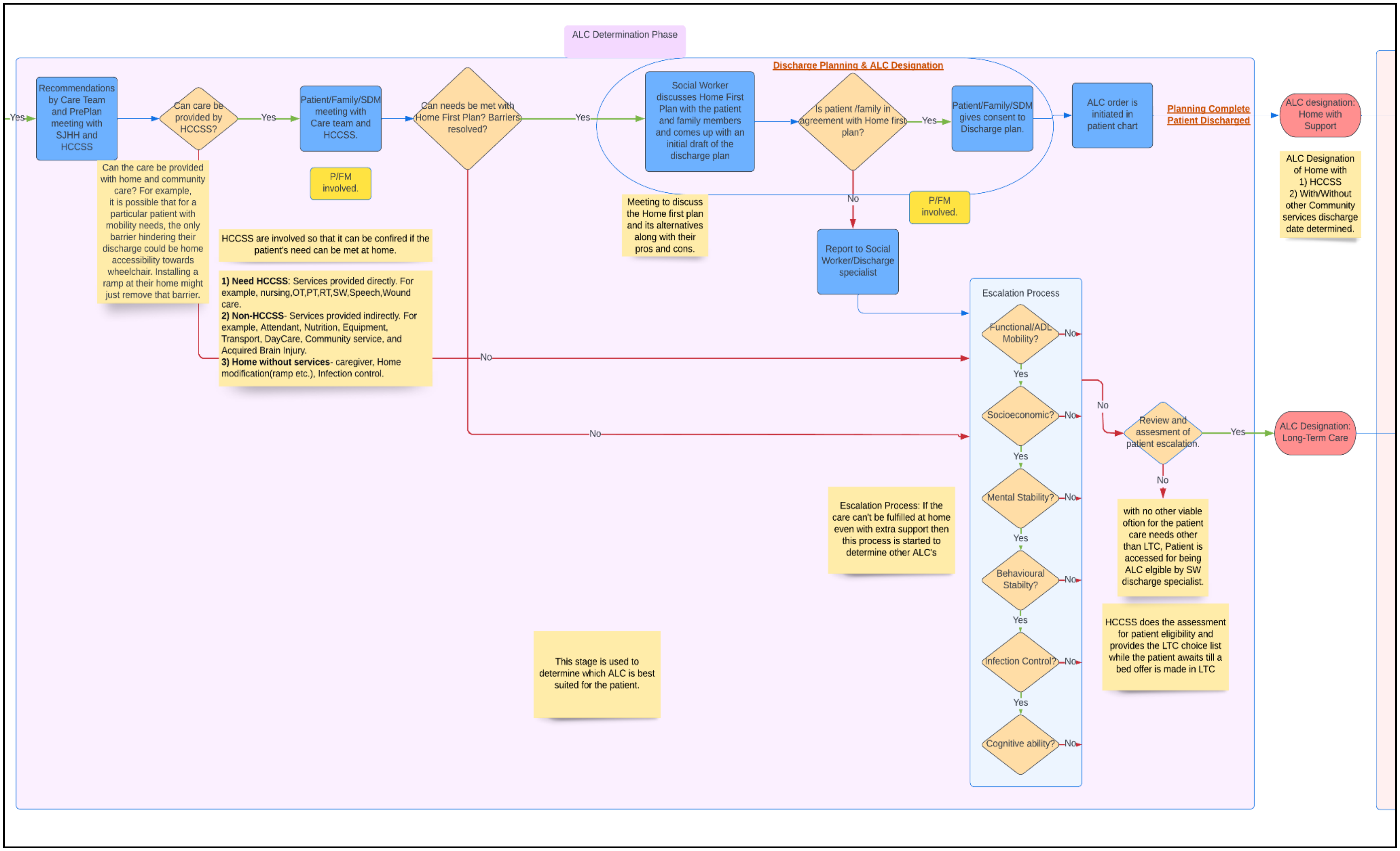


Figure 15: Final Iteration- ALC Determination Phase (2/3) [Link]

Waiting Time Phase

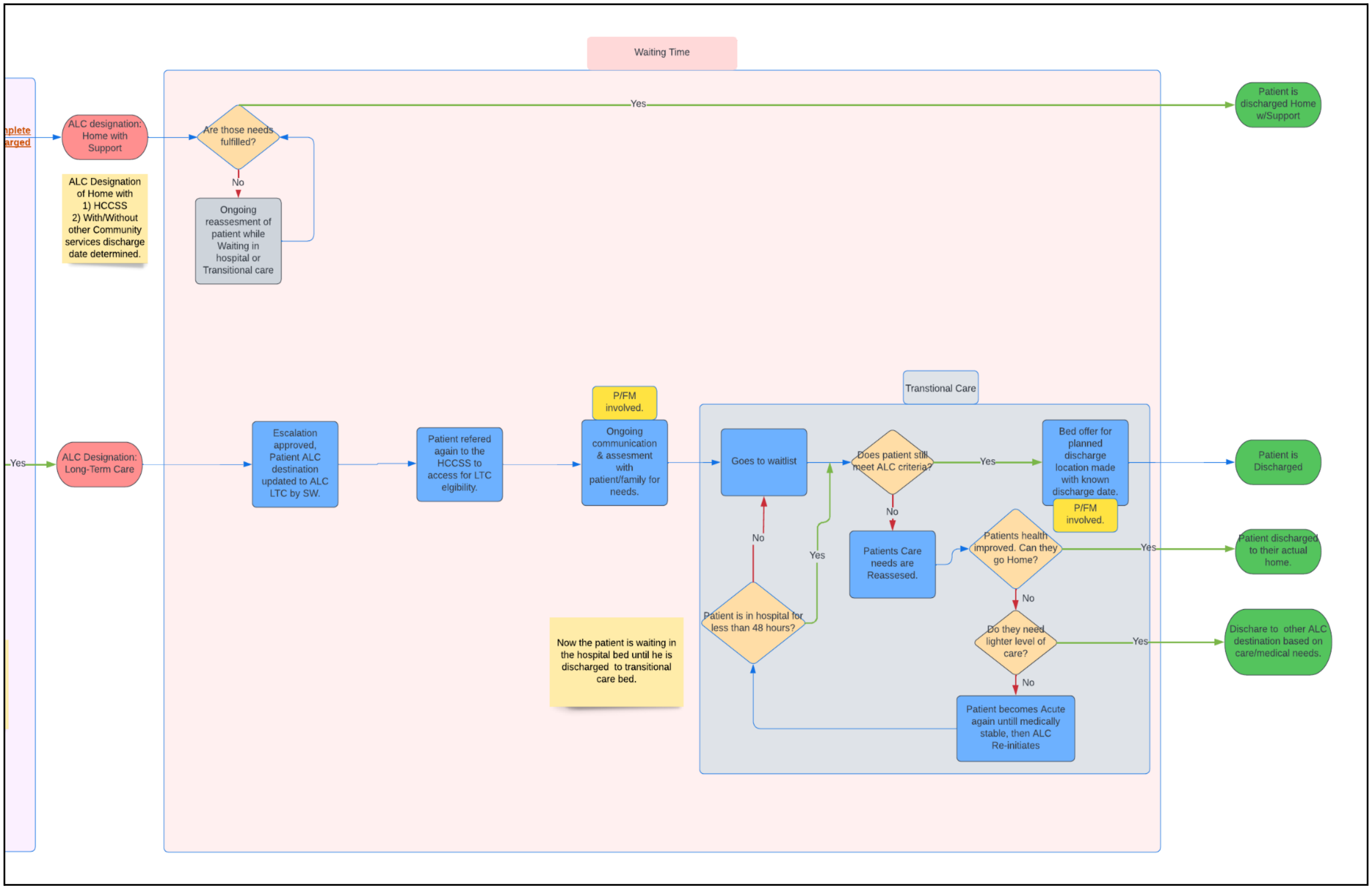


Figure 16: Final Iteration- Waiting Phase (3/3) Link

ALC Journey Map

Based on the final flow chart we created a journey map to illustrate the current and reimagined scenario for a patient. The journey map shows how the knowledge of the probable ALC destinations will help the patient in making more informed decisions during the ALC Determination phase. It also decreases the uncertainty during the Waiting time Phase.

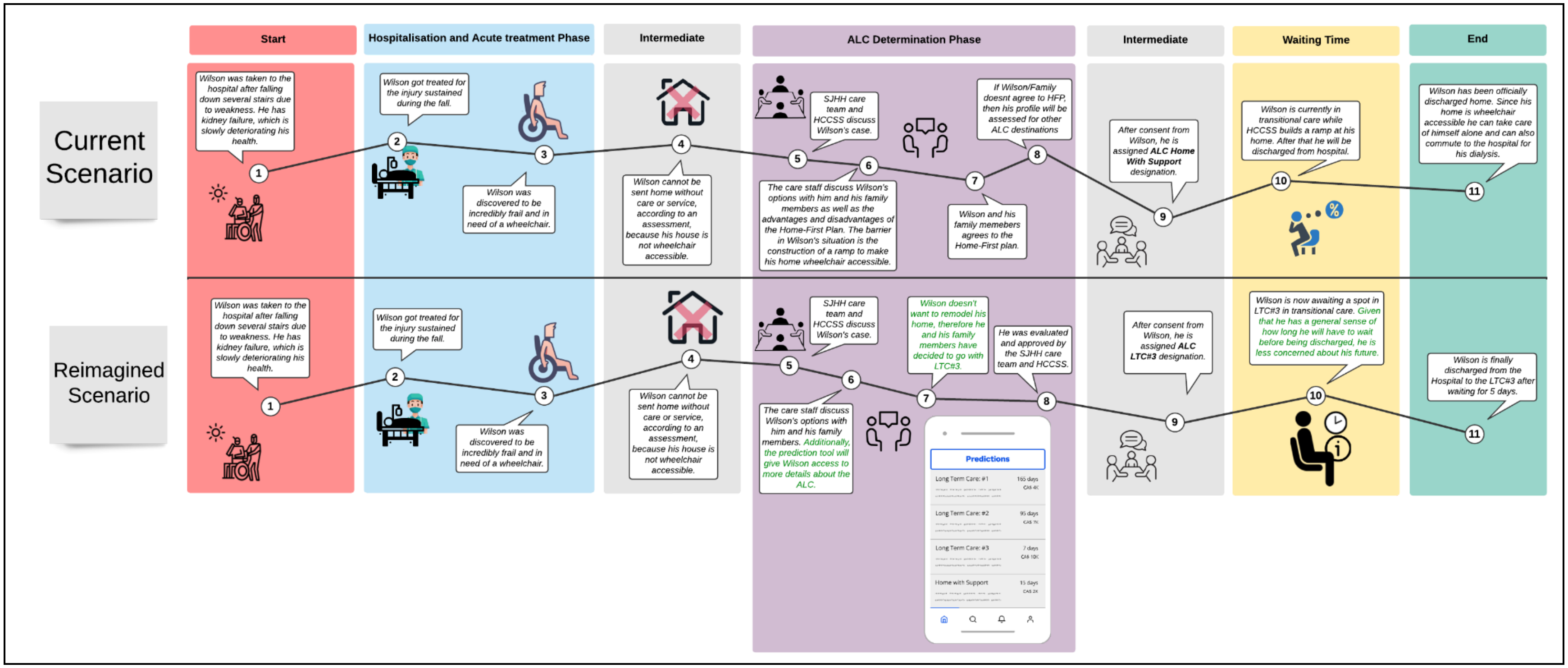


Figure 17: Current and Reimagined Scenario of Patient's Journey [Link]()

Final Iteration of MatLab Simulation Model

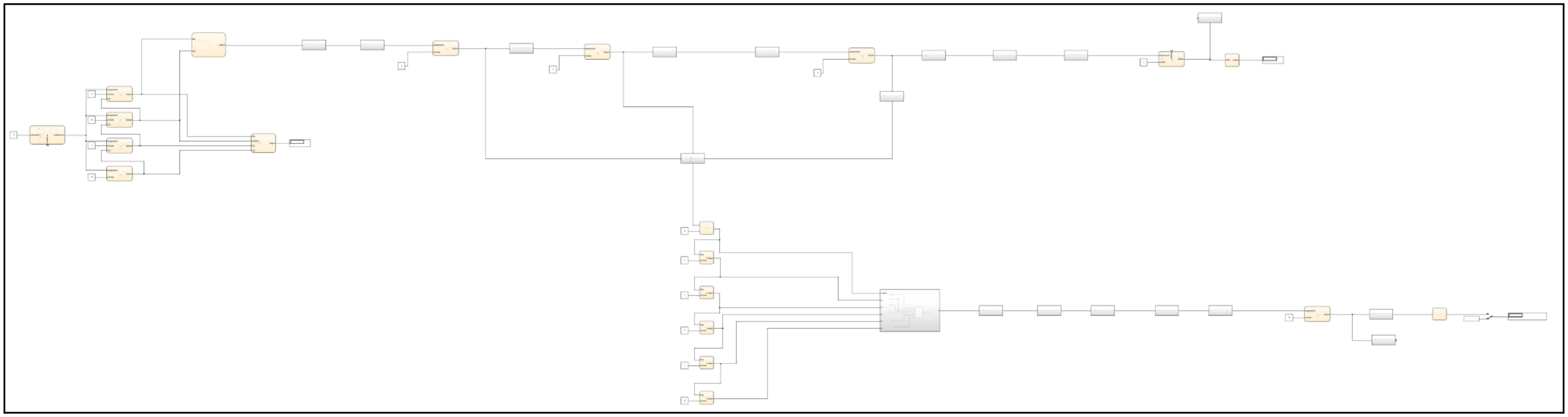


Figure 18: Final Iteration of MatLab Simulation Model Link

In the final iteration, just like the final flowchart, the simulation model is repeatedly tested and refined to achieve the best simulation state. After the user manually enters each independent variable, the system will collect and convert the signal into the adjustment parameters of the final result service, which will be aggregated into the final prediction module. The prediction module can give a reasonable prediction result given the previous training data. This part needs to be used in conjunction with the machine learning part to achieve the best application effect.

Final Design Validation

We are extremely grateful for Debbie Thibeau who was constantly on lookout for finding people we can interview. She forwarded quite a few patients' and family members' contact information. Unfortunately, despite all her efforts we were only able to interview two users overall, both social workers.

Debbie gave us the contact information of Andrea Perrins, a Nephrology Unit patient and Alex Mattioli, a family member (daughter) whom we were unable to get in touch with. She also connected us with Jaime Babin, a Complex Care Unit Patient and Nadina Cowburn, a social worker. The scheduling difficulty prevented us from interviewing these two.

We listened to the experience of the two Social Workers we were able to interview before presenting the final prototype to them because we did not want to create any bias in their answers. By listening to their work stories and how they feel and interact with the system, we were able to validate the insights we found regarding the uncertainty the patients experience during the discharging process.

Sarah Huckle

Background

Sarah is a Social Worker on the Surgical Floor of St. Joseph's Healthcare Hamilton. She has been working as a Social Worker for 5 years.

General Feedback

“Patients will be better informed with information available from probable predictions by the tool and thus patients will have a good idea of the next steps in their journey and would be less frustrated than before.”

“Usually, social worker goals align with patients, but sometimes due to the hospital policies, goals might not align, and then there are disconnects between patients and the care system”

“Transparency is needed between the care team and patients as it plays a critical role in building patient trust on the care team and the system.”

“Sharing info by social workers with patients after management approval often leads to patient distrust in their social worker”

According to her, Empathy and support from the Care Team towards patients fix the barriers, if any, are in place.

Limit sharing the most critical information because as things change, priorities change and critical things change for patients·

Feedback on Tool's usability during ALC Determination Phase

"Patients are frustrated with vagueness and non-transparency between them and the care team"

"Surely, the probable ALC destination choices in form of a list could make patients more informed and thus better decisions could be made"

Feedback on Tool's usability during Waiting Phase

"There would be Less frustration due to Length of Stay predictions. A ballpark info at least could be given to patients"

Launi Greenspan

Background

Launi is another Social Worker at St. Joseph's Healthcare Hamilton with more than 6 years of experience working in this field.

General Feedback

"Sometimes what patients want and what the hospital's recommendation is for ALC can be different. A patient might want to go home but the team might think the LTC is the most appropriate destination for the patient."

"Oftentimes it is what the patient wants vs. what the team(Social Worker), even though we are guiding them. The final decision is the patient's even though it might not be recommended by us(social workers). And we support that decision"

"Sometimes I just keep it simple and don't even tell them what the ALC is".

“How much information I give them depends on person-to-person, how much they can take it. There is always a fine balance. If we give too much information they become paralyzed and don’t know what to do.”

Feedback on Tool’s usability during ALC Determination Phase

“What worries me is that I don’t want patients to think that LTC is an option until all the other ALC options are explored. If your app says LTC, the patient would be like, ‘Oh this app says I am eligible for LTC, I’ll just wait for it no matter how long the wait is, sign me up! ”

According to her, the app should give the prediction in stages. First all the non-LTC predictions and after all the other ALC options are considered and rejected then LTC predictions.

Feedback on Tool’s usability during Waiting Phase

“This app will definitely be useful. Because it is the fear of the unknown that stresses people out. I even think that the predictive tool could also help when a patient reaches LTC designation. There is a long waiting time when we refer the patient to Home Care and when the Home Care can actually see them. So, waiting time could be for how long the patient has to wait to get designated to ALC-LTC”

“The most challenging situations are those when the patient wants to go to their home even when they can’t without the additional support. Another difficult topic to discuss with patients is money but unfortunately, this topic comes up a lot in my field.”

“Patients sometimes don’t want to work with us. They get climatized to a hospital environment with nurses and doctors all around them. They become ‘institutionalized’. And we can’t refer the patient until they give their consent.”

Outcome

Term 1 (From January 2022 to April 2022)

We studied the previous team's work at this time, looked into the limitations of the current predictive modeling technology, and more.

Although the SJHH forwarded their research overview on the predictive model that they developed it was severely limiting. So from the project's inception, we have been considering ways to build and enhance our predictive modeling tool. The existing predictive model by SJHH is well-established for conditional forecasting. They can forecast if a patient will be assigned an ALC designation or not with pretty good accuracy. However, that prediction, while useful, was not nearly sufficient.

We spoke with Debbie, a social worker, and with her assistance were able to gain a deeper understanding of how the ALC discharge procedure functions. And we also created a MatLab model to simulate a patient's journey so that we know what factors are responsible for each ALC designation. One of the main benefits of translating the ALC journey to the MatLab model was that it gave us insight into what questions to ask which ultimately increased our understanding of the patient's ALC Discharge journey.

Term 2 (From May 2022 to August 2022)

In this stage, we looked more closely at the inner details of the ALC discharge system in an effort to provide additional context for the patient's journey. We developed the ALC journey Flowchart after several iterations by getting feedback and suggestions from Debbie.

We chose the most complex ALC destination, long-term care, and made an effort to comprehend every component that contributed to the classification of ALC-LTC. We found and categorized the critical factors contributing to the ALC-LTC decision-making into six main categories as shown in the flowchart.

We then looked at how the patient will benefit from our tool throughout their ALC journey. We clearly highlighted the steps where the patient/family members were involved in the process and how we can help to improve their experience by using the predictions from our tool.

This helped us to narrow our scope which could help to improve patient experience through the ALC journey by providing the predictions like ALC destinations, length of stay, and the associated costs to ALC destinations. Additionally, we developed the Matlab Model - a mathematical version of the flowchart, that could make different choices and produce outcomes in the form of predictions for the patients.

Furthermore, we developed a Machine Learning model that we trained on the Mock data provided by the hospital to get the predictions. Although the usability of the Prediction Model is not realistic, that could be attributed to the limitation of the Mock data. Nonetheless, the feasibility of the Prediction model is not in question as with better input data the model will give better predictions.

Future Scope

Although we achieved a lot, this project is still far from completion. Due to time constraints, we were only able to explore the ALC Long-Term Care designation. The future team, if they so desire, could explore other ALC destinations.

Moreover, since we lacked the necessary expertise in data analysis and machine learning, we were only able to verify the feasibility of the Predictive Modeling Tool but we were unable to explore the actual working of the model in detail. Future engineering design students can team up with students who are researching these technologies to gain deeper insights.

In the subsequent MATLAB simulation model, the human-computer interaction interface can be added in the future, so that the user can directly input the variable value. The system process also has room for upgrading and optimization. In the MATLAB module, a data processing module based on artificial intelligence algorithms can be added to assist users in making more accurate decisions. The simulation model of MATLAB is built based on the flowchart replicating the SJHH decision node. In the future, students can continue to upgrade and optimize the model in cooperation with the facilitators of SJHH, making the model structure more convenient for users to use and meet the needs of users.

At the algorithm level, the Python-based Adaboost algorithm is used in this project to iteratively calculate the data set to evolve from a weak classifier to a strong classifier. In this project, LDA and SVM are used as cross-comparison objects to improve the reliability of AdaBoost. In future algorithms, designers can try to use other algorithms such as PCA (Principal component analysis) or KNN (K-Nearest Neighbors). Programming Languages C++ can also be explored as MATLAB supports the use of C-based programming. If you can use C++ to implement machine learning algorithm functions, you will be able to integrate your own machine learning algorithm modules in MATLAB.

For the data used, the project team obtained the mock data rewritten by SJHH according to the actual data. In the future data processing and data cleaning, designers can try to increase the amount of data, increase data categories, increase data abundance, and maintain the balance of output results to improve the credibility of data processing results. Designers can try to convert text variables into data variables, remove redundant data noise and rearrange data distribution, etc., to obtain more possibilities for data processing results.

This project requires experience in data processing, machine learning programming, mastering a variety of basic programming languages and engineering design experience. The project team recommends that future teams take graduate courses in deep learning, artificial intelligence, IoT, machine learning and data processing. Online Learning Platforms like W3school (https://www.w3schools.com/), and resource-sharing websites like GitHub can be helpful to increase the familiarity with the technical aspect of this project. Or better yet, they can team up with students who are majoring in these technologies so as to get better insight into the inner workings of the Predictive Model and contribute.

References

Callen, B. L., Mahoney, J. E., Grieves, C. B., Wells, T. J., & Enloe, M. (2004). Frequency of hallway ambulation by hospitalized older adults on medical units of an academic hospital. 10.1016/j.gerinurse.2004.06.016

Cressman, G., Ploeg, j., Kirkpatrick, H., Kaasalainen, S., & McAiney, C. (2013, December). Uncertainty and alternate level of care: a narrative study of the older patient and family caregiver experience. 10.1177/084456211304500403

Deilami, N., Almendariz, V., & Ogunmoyero, T. (2021). A Collaborative and Interactive Tool for an effective hospital discharge journey.

Fortuna, J. (2022, Jan). McMaster University-W Booth Tech-CHEMENG 787 Machine Learning: AdaBoost (Class Notes).

Fortuna, J. (2022, Jan). McMaster University-W Booth Tech-CHEMENG 787 Machine Learning: Linear Discriminant Analysis (Class Notes).

Fortuna, J. (2022, Jan). McMaster University-W Booth Tech-CHEMENG 787 Machine Learning: Support Vector Machine (Class Notes).

Seffah-Nyame, S., & Ragavan, V. (2020). THE CASE OF THE ALC PATIENT AND A DECISION AID TOOL.

Thibeau, D. (2022, Social Worker at SJHH).

APPENDIX 1

Candidate Algorithms for ALC Destination Prediction Models

AdaBoost

AdaBoost is an adaptive boosting machine learning method. The adaptation of the AdaBoost method is that the samples misclassified by the previous classifier will be used to train the next classifier. The AdaBoost method is sensitive to noisy data and outlier data. But in some problems, the AdaBoost method is less prone to overfitting than most other learning algorithms. The classifier used in the AdaBoost method may be weak, such as with a large error rate. But its classification performance can improve the resulting model. Weak classifiers with higher error rates than random classifiers are also useful, because, in the final linear combination of multiple classifiers, negative coefficients can be assigned to them, which can also improve the classification performance.

The AdaBoost method is an iterative algorithm that adds a new weak classifier in each round until a predetermined small enough error rate is reached. Each training example is assigned a weight indicating the probability of it being selected into the training set by a certain classifier. If a sample point has been accurately classified, its probability of being selected in the construction of the next training set is reduced; on the contrary, if a sample point is not accurately classified, its weight is increased. In this way, the AdaBoost method can "focus" on the more difficult samples. In the specific implementation, the weight of each sample is initially equal. For the kth iteration operation, we select sample points according to these weights and then train the classifier Ck. Then, according to this classifier, the weight of the wrongly classified samples is increased, and the weight of the correctly classified samples is reduced. Then, the weight-updated sample set is used to train the next classifier Ck. The whole training process goes on iteratively. (Fortuna, 2022)

LDA

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant method, trying to find a linear combination of the features of two classes of objects or events to be able to characterize or distinguish them. The resulting combination can be used as a linear classifier or, more commonly, for dimensionality reduction for subsequent classification. Linear discriminant analysis is a classic linear learning method. The method of linear discrimination is: given a set of training samples, try to project the samples onto a straight line so that the projection points of similar samples are as close as possible, and the projection points of different samples are as far away as possible; when classifying new samples, project it onto the same straight line, and then determine the category of the new sample according to the position of the projected point. LDA is also closely related to Principal Component Analysis (PCA), both of which are looking for linear combinations of variables that best explain the data. LDA explicitly attempts to model the differences between data classes. PCA, on the other hand, does not consider any difference in classes. The discriminant analysis must distinguish the difference between independent and dependent variables. LDA works effectively when each observed measurement of the independent variable is a continuous quantity. (Fortuna, 2022)

SVM

Support Vector Machine (SVM) is a kind of generalized linear classifier that performs binary classification of data according to supervised learning, and its decision boundary is the maximum margin for solving the learning sample. The maximum-margin hyperplane. SVM uses the hinge loss function to calculate the empirical risk and adds a regularization term to the solution system to optimize the structural risk. It is a classifier with sparsity and robustness. SVM can perform nonlinear classification through the kernel method, which is one of the common kernel learning methods. Linear separability is given input data and learning objectives in classification problems, where each sample of the input data contains multiple features and thus constitutes a feature space, and the learning objective is a binary variable representing negative and positive classes. If the feature space where the input data is located has a hyperplane as the decision boundary to separate the learning targets according to the positive and negative classes, and the distance from the point to the plane of any sample is greater than or equal to 1, the classification problem is said to be linearly separable, and the parameter is the normal vector and intercept of the hyperplane, respectively. The distance between the two interval boundaries is defined as the margin, and the positive and negative class samples located on the interval boundaries are the support vectors. The loss function is that when a classification problem does not have linear separability, using the hyperplane as the decision boundary will bring classification loss, that is, some support vectors are no longer located on the interval boundary, but enter the interval boundary, or fall into the decision. The wrong side of the border. The loss function can quantify the classification loss, and its mathematically available form is a 0-1 loss function. The classifier generates risk when it is learned and applied to new data, and the types of risk can be divided into empirical risk and structural risk. The empirical risk is defined by the loss function, which describes the accuracy of the classification results given by the classifier; the structural risk is defined by the norm of the classifier parameter matrix, which describes the complexity and stability of the classifier itself. Complex classifiers are easy to produce overfitting and are therefore unstable. If a classifier determines its model parameters by minimizing a linear combination of empirical risk and structural risk. Some linearly inseparable problems may be nonlinearly separable. Using nonlinear functions can map nonlinearly separable problems from the original feature space to a higher-dimensional space, thereby transforming them into linearly separable problems. This method is called the kernel method. (Fortuna, 2022)

Exploring Mock Dataset

PAT_MRN_ID	HOSP_ADM_DTTM	HOSP_DISCH_DTTM	DISCH_ATTENDING_PROVIDER	HOSPITAL_SERVICE	DISCH_DISPOSITION	PATIENT_LEVEL_OF_CARE	ALC_BOOL	DISCH_DECEASED_BOOL	PAYMENT_COVERAGE_BOOL
J0002007911	5/26/22 11:59 AM	5/26/22 12:08 PM	10007205	General Internal Medicine	Discharge Home without Support Services	Acute	0	0	1
J0002007858	5/26/22 9:13 AM	5/26/22 10:29 AM	10007205	General Internal Medicine	Discharge Home without Support Services	Acute	0	0	1
J0002007904	5/25/22 3:53 PM	5/25/22 4:05 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002007899	5/25/22 9:55 AM	5/25/22 10:44 AM	10007205	General Internal Medicine	Discharge Home without Support Services	Acute	0	0	1
J0003206681	3/15/22 6:46 PM	5/24/22 2:38 PM	NULL	Cardiology	Discharge Home without Support Services	ALC - Unknown	1	0	1
J0002005361	2/3/20 7:27 AM	5/24/22 2:37 PM	10007205	Respirology	Discharge Home without Support Services	NULL	0	0	0
J0003207205	3/15/22 6:38 PM	5/24/22 2:36 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002007854	5/18/22 10:23 AM	5/24/22 2:17 PM	10007205	General Internal Medicine	AHF PATIENT - Discharge/Admit to Acute Care SJHH	Acute	0	0	1
J0002007886	5/24/22 12:45 PM	5/24/22 12:51 PM	10007205	Gastroenterology	Discharge Home without Support Services	NULL	0	0	1
J0002007853	5/18/22 10:15 AM	5/24/22 12:01 PM	10007205	Cardiology	Discharge/Admit to Acute Care - AHF	Acute	0	0	1
J0002007436	2/18/22 12:15 PM	5/24/22 8:57 AM	E104502	Pediatrics	Discharge Home without Support Services	NULL	0	0	1
J0002007860	5/19/22 10:25 AM	5/19/22 12:30 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0003207042	3/15/22 6:48 PM	5/19/22 11:09 AM	NULL	Cardiology	Discharge Home without Support Services	ALC - Unknown	1	0	1
J0003207043	3/15/22 6:50 PM	5/19/22 11:04 AM	NULL	Cardiology	Discharge Home without Support Services	ALC - Unknown	1	0	1
J0002007623	3/15/22 6:56 PM	5/19/22 10:58 AM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002001450	5/19/22 10:55 AM	5/19/22 10:56 AM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0003207203	3/15/22 6:31 PM	5/19/22 10:37 AM	NULL	Cardiology	Discharge Home without Support Services	ALC - Unknown	1	0	1
J0003206829	3/4/22 4:52 PM	5/19/22 10:32 AM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J000204279	3/7/22 3:19 PM	5/18/22 11:59 AM	E1005	Obstetrics Delivered	Discharge Home without Support Services	NULL	0	0	0
J0003205921	4/22/22 2:48 PM	5/18/22 11:23 AM	E1005	Obstetrics	Discharge Home without Support Services	NULL	0	0	1
J0002007846	5/17/22 2:41 PM	5/17/22 2:42 PM	E1035	Nephrology	Discharge Home without Support Services	NULL	0	0	1
J0002007845	5/17/22 1:58 PM	5/17/22 1:59 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002000791	3/8/18 10:29 AM	5/16/22 2:11 PM	10007413	Complex Care	Discharge Home without Support Services	NULL	0	0	1
J0002007824	5/13/22 1:09 PM	5/16/22 9:41 AM	10018702	Schizophrenia	Discharge/Admit to Acute Care - St. Joes	NULL	0	0	1
J0002005026	12/6/19 12:23 PM	5/13/22 8:46 AM	10007506	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002002462	5/28/19 9:32 AM	5/13/22 8:46 AM	E1058	General Internal Medicine	Discharge Home without Support Services	ALC - Long Term Care - Copayment	1	0	1
J0003205808	3/15/22 6:42 PM	5/13/22 8:45 AM	10007441	Cardiology	Discharge Home without Support Services	ALC - Unknown	1	0	1
J0002005697	7/31/20 11:10 AM	5/13/22 8:45 AM	E1056	General Internal Medicine	Discharge Home without Support Services	Acute	0	0	1
J0002007818	5/12/22 4:43 PM	5/12/22 4:53 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002007443	3/15/22 10:09 AM	5/12/22 3:04 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002005703	8/10/20 11:26 AM	5/10/22 4:33 PM	E1001	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0003208638	11/3/17 4:06 PM	5/10/22 4:32 PM	10007205	Critical Care	Discharge Home without Support Services	NULL	0	0	1
J0002005126	10/10/19 10:32 AM	5/10/22 4:26 PM	10003052	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002001927	4/26/19 8:54 AM	5/10/22 4:23 PM	E1027	Orthopedics	Discharge Home without Support Services	NULL	0	0	1
J0002006270	2/12/21 11:40 AM	5/10/22 4:23 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002005364	2/3/20 11:06 AM	5/10/22 4:22 PM	10007205	Respirology	Discharge Home without Support Services	NULL	0	0	1
J0002006126	2/4/21 11:05 AM	5/10/22 4:22 PM	10007205	General Internal Medicine	Discharge Home without Support Services	NULL	0	0	1
J0002005971	10/21/20 11:05 AM	5/10/22 2:50 PM	10007205	General Internal Medicine	Discharge to Complex Care/LTC - Outside Hospital	NULL	0	0	1
J0002006210	5/10/22 1:50 PM	5/10/22 1:57 PM	10007205	Critical Care	Discharge/Admit to Complex Care - St. Joes	NULL	0	0	1
J0002006882	9/7/21 8:20 AM	4/27/22 10:38 AM	10007205	Critical Care	Expired (Cadaveric Donor)	NULL	0	1	1
J0002007484	3/8/22 10:52 AM	4/27/22 9:42 AM	10018702	Acute Mental Health	Discharge/Admit to Tertiary Mental Health - St. Joes	NULL	0	0	1

The mock data has ten columns, each with 500 rows of data. The raw data contains a large amount of unprocessed textual data, and there are unrelated classes in the data such as identification data and treatment data. This is why raw data needs to be processed and then analyzed and processed.

DISCH_DISPOSITION	ALC Designation	Deceased	Financial Capability		DISCH_DISPOSITION Translation Table
1	0	0	1		Discharge Home without Support Services=0
1	1	0	1		AHF PATIENT - Discharge/Admit to Acute Care SJHH=1
1	1	0	1		Discharge/Admit to Acute Care - AHF=2
1	1	0	1		Discharge/Admit to Acute Care - St. Joes=3
1	1	0	1		Discharge to Complex Care/LTC - Outside Hospital=4
1	0	0	1		Discharge/Admit to Complex Care - St. Joes=5
1	1	0	1		Expired (Cadaveric Donor)=6
1	1	0	1		Discharge/Admit to Tertiary Mental Health - St. Joes=7
1	1	0	1		Inpatient who does not return from a pass=8
1	1	0	1		Discharge to Shelter=9
1	1	0	1		Discharge to Hospice=10
1	1	0	1		Expired (Non-Donor)=11
1	1	0	1		Discharge Home with Support Services=12
1	1	0	1		Discharge to Outpatient Clinic=13
1	0	0	1		Discharge/Admit to Acute Mental Health - St. Joes=14
1	0	0	1		
1	1	0	1		
1	1	0	1		
1	1	0	1		
1	1	0	1		
1	0	0	1		
1	0	1	1		
1	0	0	1		

After data cleaning and transformation, the project uses three raw data as input data. "DISCH_DISPOSITION" as output data expresses the final destination of the patient in the ALC process. In “ALC Designation”, 1 means patient designated as an ALC patient. In “Deceased”, 0 means the patient is alive while getting treated. In “Financial Capability”, 0 means they can’t cover their bills. The "DISCH_DISPOSITION" in the original data uses text data. The text information that is difficult to use directly is converted into numerical information that is convenient for training (represented by 0~14, please refer to the translation table for details).

In the existing data provided by the collaborators, there are a large number of weak classifiers. The final output of 92.4% of the data points is Discharge Home without Support Services. Nearly 88% of data categories are less than 5% of the overall data volume. Since AdaBoost has a high fault tolerance rate and is not prone to overfitting, a more optimized algorithm model can be obtained. The reason for this is that the data noise was not removed during data processing. There is a large amount of non-ALC related data entered into the training data. Non-ALC-related data interferes with the training results, thus reducing the confidence value of the training results. At the same time, the combined results of the datasets also have shortcomings. The input data uses a large number of binary combinations of 0 and 1, and the data results have a low number of combination groups. This makes training data more difficult and results in conflicting results.

Machine Learning Unit Instructions

Machine Learning Unit Input Data Description

ALC_BOOL	True (1) if the patient's level of care at the time of discharge is ALC-related, otherwise false (0).
DISCH_ Critical_Illness _BOOL	True (1) if the patient is very ill at the time of discharge, otherwise false (0).
PAYMENT_COVERAGE_BOOL	True (1) if the patient had full payment coverage for their most recent encounter, otherwise false (0).

Machine Learning Unit Output Data Description

DISCH_DISPOSITION	The discharge disposition (i.e. where the patient went) post-discharge.

Machine Learning Unit Output Translation Table

Discharge Home without Support Services=0
AHF PATIENT - Discharge/Admit to Acute Care SJHH=1
Discharge/Admit to Acute Care - AHF=2
Discharge/Admit to Acute Care - St. Joes=3
Discharge to Complex Care/LTC - Outside Hospital=4
Discharge/Admit to Complex Care - St. Joes=5
Expired (Cadaveric Donor)=6
Discharge/Admit to Tertiary Mental Health - St. Joes=7
Inpatient who does not return from a pass=8

Discharge to Shelter=9
Discharge to Hospice=10
Expired (Non-Donor)=11
Discharge Home with Support Services=12
Discharge to Outpatient Clinic=13
Discharge/Admit to Acute Mental Health - St. Joes=14
Discharge Home without Support Services=0

Python Libraries Used in ALC Test Prediction Model

import numpy
import sklearn.discriminant_analysis
import AdaBoostClassifier
import train_test_split
import cross_val_score
import matplotlib.pyplot
from timeit
import pandas as pd

Instructions For ALC Prediction Model (Python-Based Cross-Validation)

Step 1: Install Python and Necessary Libraries

Step 2: Open ALC_Pred_Model.py with Python jupyter ALC_Pred_Model 最新检查点: 2022年7月1日 (已自动保存)

File Edit View Insert Cell Kernel Widgets Help

In [3]:

```
import numpy
import sklearn.discriminant_analysis
from sklearn.metrics import accuracy_score
from sklearn.ensemble import AdaBoostClassifier
from sklearn import metrics
import pandas as pd
from timeit import default_timer
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import roc_curve, auc
from sklearn import svm
from sklearn.model_selection import cross_val_score
'''load the data file'''
df = pd.read_excel(r"ALC_TRAIN.xlsx",engine='openpyxl')
'''Define the columns that from second to the last as input attributes '''
X= df.iloc[:,1:4].values
'''Define the first columns(Sex) as output '''
y=df.iloc[:,0].values
'''Split the input data for training and testing '''
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)
print("No.Data point:", len(X))
print("No.Training data:", len(X_train))
print("No.Testing data:", len(X_test))
print("-"*30)

'''Start counting time'''
start=default_timer()
'''Using LDA'''
lda = sklearn.discriminant_analysis.LinearDiscriminantAnalysis()
lda.fit(X_train, y_train)
end=default_timer()
y_predlda = lda.predict(X_test)
'''Calculate accuracy'''
accuracylda = accuracy_score(y_test, y_predlda)
end2=default_timer()
print("LDA:")
print('Accuracy is:', '%.8f'%accuracylda)
'''Genrate the confusion matrix'''
cmlda = confusion_matrix(y_test, y_predlda)
print("Confusion Matrix is:")
```

Step 3: Upload the data file (.xlsx) to Python model (Jupyter Notebook)

Step 4: Enter the target variable at new&new1

```
new = [0, 1, 0]
new1 = [1, 1, 1]
```

Step 5: View the output

```
Adaboost:
Accuracy: 0.952
The Confusion Matrix is
[[118   0   0   0   0   0   0]
 [  1   0   0   0   0   0   0]
 [  1   0   0   0   0   0   0]
 [  2   0   0   0   0   0   0]
 [  1   0   0   0   0   0   0]
 [  0   0   0   0   0   1   0]
 [  1   0   0   0   0   0   0]]
Prediction Result: [0]
Reference: [1]
Training time (second): 0.004242100000055871
Testing time (second): 0.000650099999916165
```

Step 6: If an error is displayed, re-run with ALC_Pred_Model.py after simplifying the data

Analysis of ALC Test Data Results Based on AdaBoost Algorithm

The original ALC test data has 10 columns for a total of 500 datasets. For example, PAT_MRN_ID, HOSPITAL_SERVICE, and DISCH_DISPOSITION are category data. ALC_BOOL, DISCH_CRITICAL_ILLNESS _BOOL, and PAYMENT_COVERAGE_BOOL are true/false data. Through data cleaning and transformation, the raw ALC test data were exported as numerical data containing three independent variables and one dependent variable. Then, import the processed dataset into a python program equipped with a machine learning module to obtain a regression-convergent algorithm model. This model uses AdaBoost, SVM, and LDA as three contrasting machine learning methods to test whether the model can obtain ideal ALC destination prediction results in the ALC data environment. Due to the "multiclass format is not supported" situation in the SVM test, this data test had to split the data set into two parts. ALC_TRAIN is the benchmark data and has all 14 kinds of dependent variables. ALC_GRAPH is used as adjustment data, and the corresponding variable has only 0/1 results, that is, "patients enter the ALC destination" and "patients do not enter the ALC destination".

Test results on the ALC_TRAIN dataset are

LDA:
Accuracy is: 0.928
Confusion Matrix is:
[[115 0 3 0 0 0 0]
 [1 0 0 0 0 0 0]
 [0 0 1 0 0 0 0]
 [1 0 1 0 0 0 0]
 [1 0 0 0 0 0 0]
 [1 0 0 0 0 0 0]
 [1 0 0 0 0 0 0]]
Training time (second): 0.0014
Testing time (second): 0.00035
[0]
[2]

Adaboost:
Accuracy: 0.952
The Confusion Matrix is
[[118 0 0 0 0 0 0]
 [1 0 0 0 0 0 0]
 [1 0 0 0 0 0 0]
 [2 0 0 0 0 0 0]
 [1 0 0 0 0 0 0]
 [0 0 0 0 0 1 0]
 [1 0 0 0 0 0 0]]
[11]
[1]
Training time (second): 0.0043
Testing time (second): 0.00065

SVM cannot be used in this dataset because it cannot support the projection of multi-level datasets on high-dimensional planes.

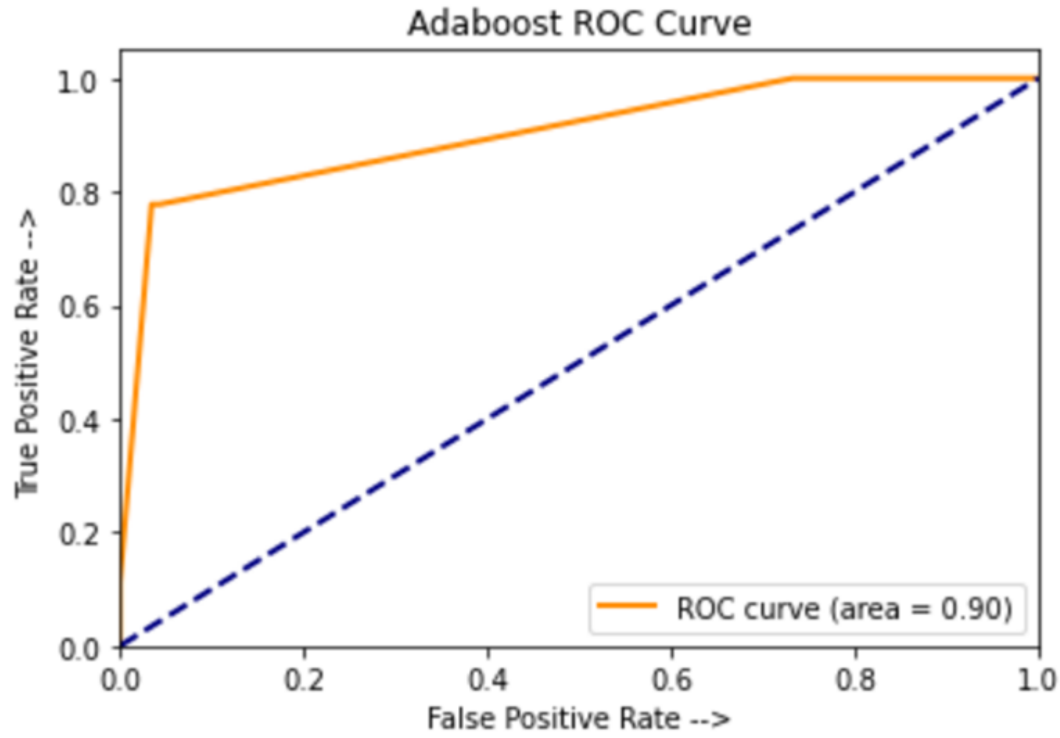
By comparison, it is found that AdaBoost has higher accuracy in processing weak classifier datasets and LDA, but the corresponding time is slower. This is because AdaBoost requires more computing power for algorithm model iteration to train a strong classifier from a weak classifier. The end result, AdaBoost, is also more practical.

Test results on the ALC_GRAPH dataset are

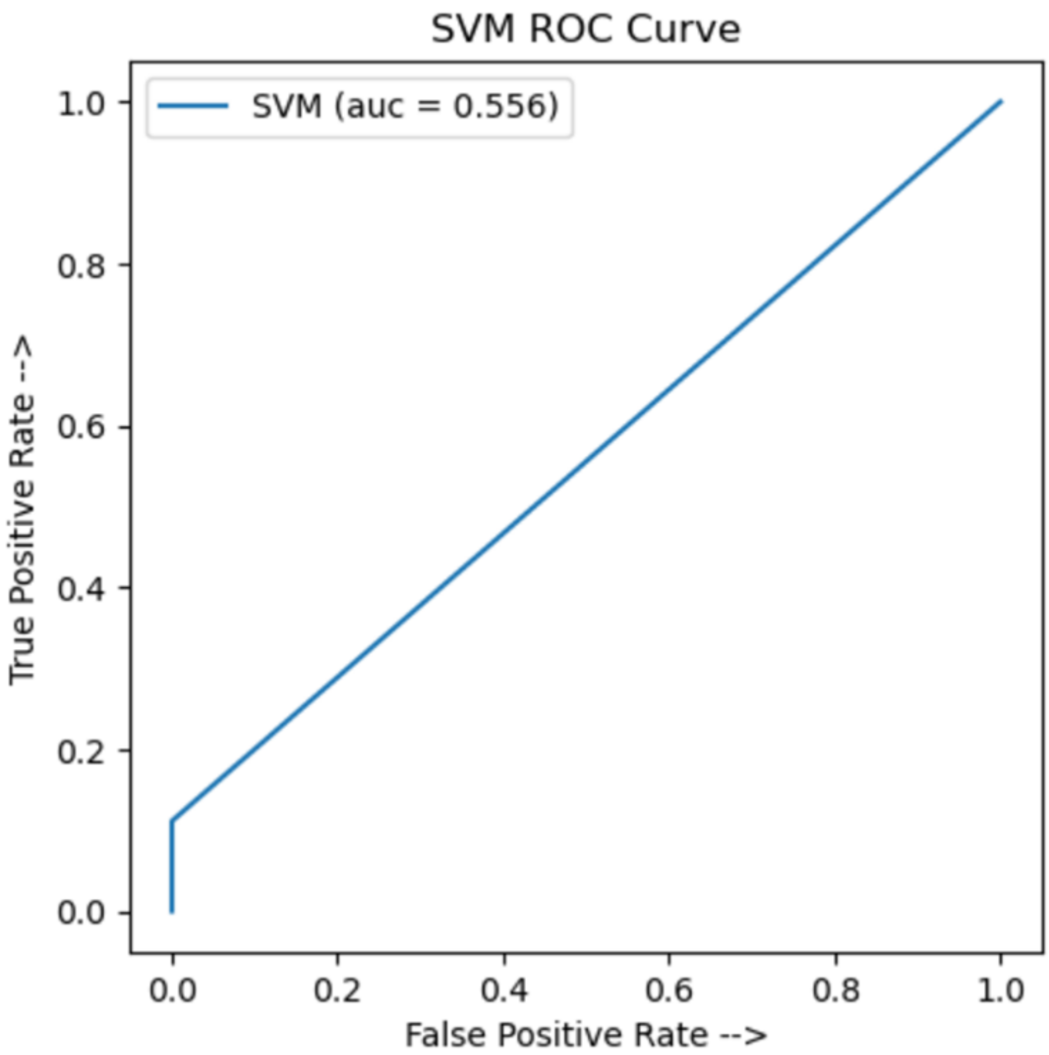
LDA: Accuracy is: 0.944 Confusion Matrix is: [[111 5] [2 7]] Training time (second): 0.0115 Testing time (second): 0.000334 [1] [1]	Adaboost: Accuracy: 0.936 The Confusion Matrix is [[116 0] [8 1]] [1] [1] Training time (second): 0.00365 Testing time (second): 0.000707	SVM: Accuracy: 0.936 The Confusion Matrix is [[116 0] [8 1]] Training time 0.00301 Testing time 0.00129 [1] [1]

Through longitudinal comparison, LDA has better performance in accuracy for the ALC_GRAPH dataset after the reduction of the dependent variable. This is because LDA's algorithm has significant advantages over AdaBoost and SVM in linearly convergent data. However, the LDA algorithm sacrifices the Confusion Matrix, which is less reliable than AdaBoost and SVM. The test time is still AdaBoost, which takes the longest and relies on higher computing power.

By comparing the ROC curves of AdaBoost and SVM, it can be found that AdaBoost is more reliable than SVM in ALC test data. The confidence interval for AdaBoost reaches 0.9, while the SVM is only 0.556. From this, it can be inferred that AdaBoost has a better overall performance in the training environment of ALC test data.



An ROC curve（receiver operating characteristic curve）is a graph showing the performance of a classification model at all classification thresholds. The smooth straight line in the middle of the ROC curve represents a 50% probability of true/false events occurring, that is, regardless of whether the classifier output is false or true, Both are only 50% correct. This can be understood in terms of a coin toss, a predictive model can predict the outcome (heads and tails) of a coin toss. The result is a smooth straight line in the middle that is not affected by any external force. The ROC curve (yellow curve) of this model is a function curve on a smooth straight line (ROC value is 0.9>0.5). This proves that the prediction results of this model are credible and relevant to the results (better than the 50% correct rate of predicting a simple coin toss).



The above figure represents the ROC curve of SVM. It can be seen that although SVM is better than AdaBoost in accuracy, its performance in ROC has lower reliability than AdaBoost. The ROC curve accuracy of SVM is only 0.556, which means that it is only slightly better than the prediction effect of coin toss. If the result of the SVM is lower than 0.5, then it means that the prediction result is not as reliable as the prediction of a coin toss. Compared with the previous ROC curve of AdaBoost, the prediction result of AdaBoost is more reliable.