



MSC IN COMPUTING - TEAM PROJECT

INTERIM REPORT

TU059/TU060

Year 2



Team Name:

AVINASH PERUBOINA - D19123919

GEETIKA AGARWAL - D18128873

AYUSHI NAYAK - D19123831

VAIBHAV NARAN GALA - D18130272

HASSAN YAQOOB - D18130130

CONTENTS

1. Introduction	4
2. USER Scenario	4
2.1 Target Users	4
2.2 Personas	5
2.3 Necessity of user persons	5
2.4 Problems aim to be solved.....	5
3. Technical Problems	6
3.1 Core Technical Problem	7
3.2 Architecture View	7
3.2.1 Layered Architecture.....	8
3.3 AI Model.....	8
3.3.1 Probabilistic Model	9
3.3.2 Recommendation Model	9
3.4 Existing Products	9
4. Technical Solution	10
4.1 Functional Requirements	10
4.2 Cross-Functional Requirements	11
4.3 Technology View	12
4.4 Identity View	12
4.4.1 Security Principles	13
4.4.2 Sequence: Unauthenticated user login Request.....	13
4.4.3 Sequence: User Authenticated user login Request	14
4.5 Technological Stack.....	14
4.6 User Interface	15
4.7 Intended Models.....	17
4.7.1 Probabilistic Model	17
4.7.2 Recommendation Model	18
4.8 Data	19
4.8.1 Dataset Collection and Generation.....	19
4.8.2 Data Storage.....	20
5. Evaluation	20

5.1 Success	21
5.2 Evaluation Criteria.....	21
5.2.1 Technology Evaluation	21
5.2.2 Initial Evaluation	22
5.2.3 Smoke Testing	23
5.2.4 System Testing	24
5.2.5 User Acceptance Testing.....	24
5.2.6 Implications of GDPR	24
5.2.7 Comparison	25
6. Conclusion.....	25
6.1 Project Management	25
6.2 Challenges	27
6.3 Limitations.....	27
6.4 Timeline.....	28
7. References	29
Appendix A.....	30
A.1 – Katie Patel	30
A.2- Lyra Foster	31
A.3 – Zoro Uzumaki.....	32
A.4- Aine Danley.....	33
A.5- Harvey Spector	34
A.6- Mark Smith	35
Appendix B.....	36
B.1- Haley system context diagram	36
.....	36
B.2- Haley Webapp Container Diagram	36
B.3- Component Diagram for Haley WebApp – User Home Page	37
B.4- Component Diagram for Haley WebApp – Historical Details.....	37
B.5- Component Diagram for Haley WebApp – Diagnostic Tool	38
B.6- Component Diagram for Haley WebApp – After Diagnostic Result	38
Appendix C.....	39
Appendix D.....	39
D.1- Level 2 Container 1.....	39

D.2- Level 2 Container 2.....	40
D.3- Level 2 Container 3.....	40

HALEY

YOUR HEALTHCARE ASSISTANT

For a better quality of life

1. Introduction

As we reach midway of our journey to make Haley come to life, we would like to present our vision in this report. At the outset, Haley is a personal health assistant designed for comprehensive health monitoring by tracking the most common symptoms. It will ask users a few questions and based on the user's responses; it will give a diagnosis based on sound, ethical, and well researched deep learning models. It aims to inform users about the doctors/General Practitioners to approach in their locality. We will base these results on an initial diagnosis and what HSE approved home-based remedies they can take in the meanwhile.

While there are many projects and well-developed products already market-ready with substantial user bases, our Haley aims to be a one-stop solution for all. With its roots in the firm belief that medical care should be universal, accessible, and affordable, we derive our business logic from advanced & modern-day medicine while also considering cultural and social beliefs. Taking all of the above while designing Haley, we made features including user data security and the right prediction our priority. In our upcoming sections, we will talk about our technical roadblocks, how we faced and overcame them, limit of the design and development for our project's scope, and how we are utilizing the rest of the time left. We sum up by talking about our visions to market Haley and scaling it as a community platform in the future.

2. USER SCENARIO

2.1 Target Users

The health application mainly focuses on users aged between 18 and 65. We have chosen this user group due to various technical and legal reasons.

1. 18 is the legal adult age; before that, a user would require their guardians to handle documents legally.
2. If below the age of 16, a user may not be familiar with the technicality of medicine. We will have to make provisions for our younger user group to have joint accounts with their caretakers on Haley.
3. A user above the age of 65 may not technically sound to use the application. According to a survey conducted by the United Nations 65+ years may be usually denoted as old age.¹

¹ https://www.who.int/healthinfo/survey/ageing_mds_report_en_daressalaam.pdf

4. Users might prefer familiar ways to deal with their problems.

The application focuses on all users falling under the target age groups but the application aims to capture the users who are students, parents, travelers, and tech enthusiasts. We came to this conclusion after doing a [survey analysis](#) and [persona segmentation](#).

2.2 Personas

We created a few examples for the decided user group, which will help visualize the users' daily life problems.

1. Meet 'Katie Patel'; she is 21 years old. Katie is an international student who recently moved from India to Ireland to become a cyber-crime lawyer.
2. Meet 'Lyra Foster'; she is 34 years old. Lyra is a high school teacher who is married with two kids. She lives a simple and fulfilling life.
3. Meet 'Zoro Uzumaki'; he is 53 years old. Zoro is a tech product director who is married with two kids. He believes in keeping his body healthy and fit.
4. Meet 'Mark Smith'; he is 30 years old; Mark is a freelancer and a travel blogger who aims to travel the world on a budget. His moto in life is to stay healthy and travel.
5. Meet 'Aine Danley'; she is 42 years old. Aine is a human right activist who is a widow with kids. She aims to live a fulfilling life and help underprivileged children.
6. Meet 'Harvey Spector'; he is 58 years old; Mark is a lawyer who is married with kids and grandkids. Harvey aims to research law, give guest lectures, and spend time with is grandkids.

Note- Full screenshot of the personas can be found in Appendix A

2.3 Necessity of user persons

User group identification and persona creation is an important step in any market analysis it provides us with an idea of the views of users regarding the problems they are facing, this in turns helps us with identifying the requirements and finalizing the user stories for the project. This also helps in gauging the feasibility of the project and prioritizing the features.

2.4 Problems aim to be solved

We did a persona segmentation for the project and identified the problems which the user faces in their day to day life in respect to health. In this project we are taking these observations as the problem statement and trying to solve them for the users.

Proto Persona Questions for Segmentation	Student Persona - Katie	Tourist Persona - Mark	Senior Persona - Harvey	Teacher Persona - Lyra	Tech Manager Persona- Zoro
How conscious are you with your health?	When I was living back in India my mother use to handel things related to my health, even now when I feel ill I first call up my mom. She is my goto person for everything.	I am very conscious of my health. I need to be healthy to travel the world.	I am old now its natural for me to get sick often but I do try to follow the doctors order to remain healthy	I am a little careless with my health but my children's health is my top priority	I am very attentive to my health. I treat my body as a temple and would always do so.
Do you have a record of all your medical history?	I do not have a full record of my health but I can name any major problems that I have in the past 5 years. I have forgotten how many times I have taken over the counter medicine for simple headache or fever.	I do have a record of all the major health problems I have faced but due to my constant travels I sometime loose my records or forget to add somethings to my record.	I have a file which I maintain with all my medical documents and I am very careful with it	I am always loosing one paper or the other, I have a good memory but in terms of actual prescription I am not sure.	I use my PC to scan and store all my medical documents
What is your biggest concern with regards to your health?	After moving to Ireland I am scared of what will happen if I fell ill. I do not have any friends and relatives to take care of me. I also do not know whome to contact if I feel ill.	My biggest concern is that on my travels one day I will fall terribly ill and would not be able to do anything	I have a lot of health conditions which I am currently managing but I do feel that some day one of these will get bad and due to my other conditions I am be unable to recover.	I am very healthy and do not have big health concers, it is just that sometimes I confuse my children's symptoms specially when they are sick at the same time.	I have always been a healthy person and have proper full body checkups every 6 months
How easy is to get professional help?	Back in India it was easy to get professional health I could feel ill today and be able to see the doctor by tomorrow, but since moving to Ireland I have discovered a waiting period of 2-3 days to see the doctor.	I am always in different places sometimes it is easy to find doctor and sometimes it is not	My GP is with me for 25 years now and is pretty much family. He I always there when I need him	I can easily get appointment with my family doctor for the next day whenever I need.	I am usually not plagued with sudden illness and even if I am I can get in contact with my GP within 2-3 days
Would you prefer to use home remedies to manage your symptoms?	I would love to use home remedies to manage my symptoms and already do so now.	I would like to have that option in case I am unable to contact a doctor	I am not a big fan of home remedies but if something is there that my doctor approves I will try it.	I already experiment with some home remedies for my kids I cant always give them chemicals for every small thing.	I do not believe in home remedies
How do you feel about using a health care assistant?	A health care assistant could like a brilliant idea I will have one less thing to worry about.	I would give it a try and test out how it works	I have my children for that I do not need a piece of technology for this.	I don't trust technology and would rather spend my time doing something else	I will be intrested in using one I am always curious towards new technology
Is there anything particularly frustrating when it came to managing your health? Why	I don't have much time to manage my health due to my busy schedule of classes and usually just belive to walk it off when small problems are considered. I fear that my careless nature will land me into big trouble some day and develop into a big problem	when I am outside EU to some different continent I do not understand the medications because they are non-EU standard and hence when I normally visit my own GP things sometimes just get complicated.	There is 10 years worth of medical history with me and a big file sometimes I fear that the papers will get old and disintegrate and I will loose that information.	Everything is pretty much like it would be for a mother of two kids. I just want to give my all to my kids and their wellbeing.	I manage my health quite properly

Note- Full and detailed document is attached by the name of Persona Segmentation.xlsx



Persona
Segmentation.xlsx

3. TECHNICAL PROBLEMS

We based Haley on the problems that we observed from the initial survey of the users and popular idea of building an artificial intelligence-based medical diagnosis which is backed by sound logic, science, and ethics. We came up with a solution where we focused on solving problems and not just the technology.

We discovered our problems were fourfold:

1. We noted that well-versed doctors do not need the tool for themselves as they train for their trade for almost a decade. However, if a well-trained algorithm were to pose the same questions as they would, much time could be saved in advance of the consultation.
2. Patients could hit a plethora of search engines to diagnose what ails them, but they usually get caught by platforms devised using SEO-equipped manipulations.
3. The already existing products lacked in creating a tool that considers minor health issues as major symptoms that could make or break a pandemic.
4. Opportunistically, there is no application in the current Irish market that helps the users manage their health and symptoms and provide a platform to avoid a big pandemic like the current one. Also, the project helps with getting in touch with GPs around the user's local area. Finally, we aspire to build a healthy community where we follow the concept 'each one care one' with Haley's assistance.

3.1 Core Technical Problem

Use-case oriented implementation must be a technical design, so we need to consider the issue that Haley is trying to solve first from a business perspective. Afterward, understanding the domain will make it easier to understand the technical design better. We highly recommend understanding our mission, as well.

3.2 Architecture View

The Architecture View section transforms our business domain into a conceptual view of our architecture. We mean that it captures less information on the technology we are using (but may be affected by it) and is motivated more by the problem domain. This section will also capture some of the high-level trends that we are embracing.

We have referred the Domain-Driven Design.

The context map displays all the different languages within the scope that we use or the illustration.

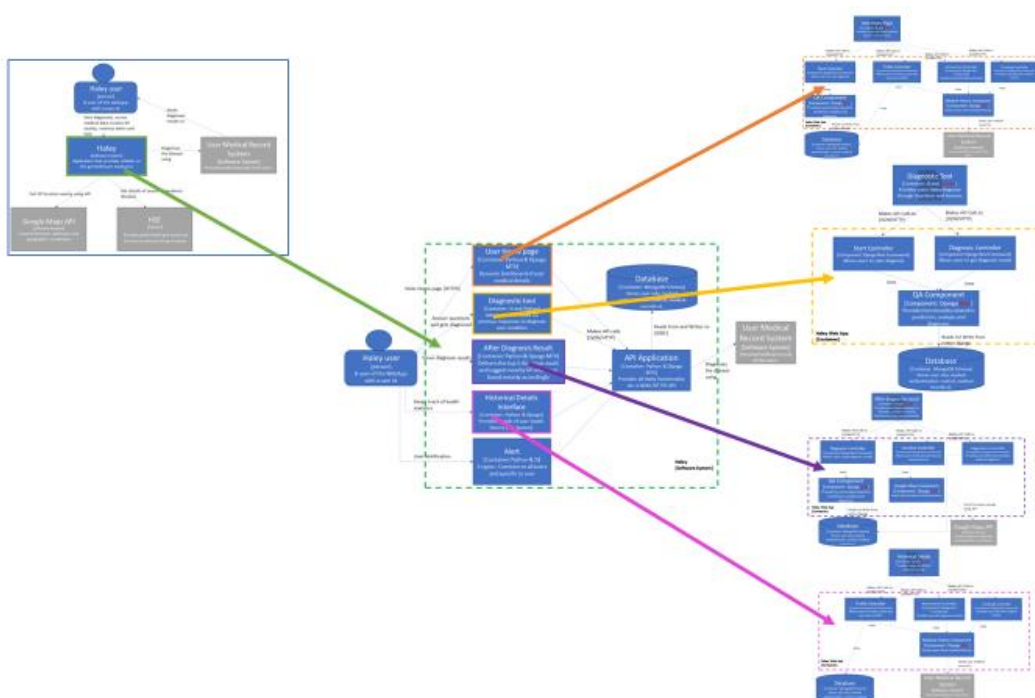


Figure 1: Context Map (Separate subsystems are available in Appendix B)

The bounded setting is an essential concept in our architecture because it is the foundation of our architecture's modularity. Upon starting development, we have realized that we need more granularity to our design and implement the same during subsequent sprints of the development.

Also, we have populated the [Tactical Design](#) section on the index of this document to dig deeper into each bounded setting (Appendix C).

3.2.1 Layered Architecture

There would be a different rate of change in the various parts of our architecture, and we need to make sure that a change in one component does not unintentionally break the other. Our architecture organizes the layered architecture pattern such that each of the layers can have its responsibility.

We borrow the layer names from DDD, [Hexagonal Architecture](#), and expand the layers further to help our device requirements. There are shared jargon among different patterns as much as we want to follow prior arts. We may have called them differently to prevent confusion between these various patterns. These are the layers that we have got:

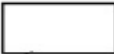
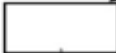
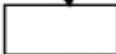




Short name	Full name	Dependency direction	Responsibilities
Ui	UI Components		Provides tangible interface for users to interact with Contains minimum domain logic and delegates the rest to the client
Clients	Clients		Abstracts integration protocol details into a library
Driving	Driving Adapter		Note: There is not separate application services layer, they are contained here Provides application services Coordinates domain objects Handles protocol details
Domain	Domain		Contains our business rules and use cases Encapsulates business logic into DDD tactical patterns
Driven	Driven Adapter		Persists entities into data storage Queries entities from data storage Interacts with infrastructure services e.g email etc.
Data	Data Pipeline		Ingests external data into our data storage
infra	Infrastructure		Provisions infrastructure via infrastructure as code

Table 1: Layered Architecture

This pattern is to be extended but can come in varying degrees to all boundary setting. For Core bound setting, we rigorously implemented and for others, less.

3.3 AI Model

To effectively implement the features of the project we need to make two neural network models to process the input received from the user and provide responses based on it.

3.3.1 Probabilistic Model

This model deals with taking symptoms and their weights as an input from the user process these inputs on the bases of the learning dataset and provide a set of diseases and their probabilities which may ail the user.

3.3.2 Recommendation Model

The output of the probabilistic model is taken as the input of this model, this model then processes the output and provides the user with some home remedies that they can use to manage their symptoms till the time professional help is unavailable and even after that if the users prefer to do so.

3.4 Existing Products

We have done an analysis of the features offered by other similar products in the market and compared it to what our product aims to offer. The below table shows a tabular representation of this analysis.

	ADA HEALTH	WEB MD	HEALTHLINE	JustAnswer	HALEY
Historical Symptom tracking	✓	✗	✗	✗	✓
Symptom Diagnosis	✓	✓	✓	✗	✓
Easy to use	✓	✓	✓	✓	✓
Web application	✗	✓	✓	✓	✓
Mobile application/ can be accessed on smartphones	✓	✓	✓	✓	✓
Remedy recommendation	✗	✓	✓	✗	✓











Document Upload					
List of professionals nearby					

Table 2: Haley vs Existing products

Note- A more detailed analysis of the products can be found in the document attached named Competitor Analysis.docx

4. TECHNICAL SOLUTION

4.1 Functional Requirements

S.No	Use case	Description	Priority	Rationale behind priority
1	Create and Manage user profile	Once the user registers on the app an account is automatically created as user profile.	High	As the data is personal and critical, we should maintain separate profile for each user
2	Diagnostic tool	A user answers questions popped up by the app and based on these they get diagnosis results	High	This is the core functionality of the app that we intend to implement.
3	GP details	Based on the diagnostic results, a user is suggested of nearby GP's through Google maps API.	Medium	This is mainly based on diagnosis details, so relatively its priority is lower
4	Assessment details	Once the user diagnosed, the results will be maintained and can be seen through this assessment feature	High	This is another core functionality of the app to maintain the user assessment records.
5	User Feedback	Users can provide feedback related to their diagnosis	Medium	User feedback is important to improve our model accuracy which can be alternatively improved with synthetic data like how we implemented it during its initial development.

6	File system	Users can maintain their medical files/data digitally and can be accessed on the go.	Medium	We are giving the assessment details feature, so our database captures a lot of medical history of user. Hence, the priority is lower compared to others.
7	Remedies	Users are suggested home based remedies approved by HSE and WHO.	Low	Since the data set we have used do not provide remedies and we can scrape data once other features are implemented.
8	Alerts	In case of outbreak, users are notified about an epidemic in their locality.	Medium	Only once the system is ready to use and our models are trained on real time data, can we send out alerts based on diagnosis in a demography and HSE

Table 3: Functional Requirements

4.2 Cross-Functional Requirements

Requirement on extreme right are most important.

1	2	3	4	5
Portability	Reuse (as app / device)	Robustness / Errors / Data integrity	Observability	Usability
Reliability / Recoverability	L10n I18n	Availability / Support hours	Cross browsers / device	Authorization
Auditability	Compatibility / Integrability	Legal / Privacy	Maintainability	Accessibility

Documentation	Capacity	Scalability	Backup	Security / GDPR
			Performance	

Table 4: Cross Functional Requirements

4.3 Technology View

The Technology View transforms our Architecture view into a more physical view of our architecture. This viewpoint makes it more applicable to the technologies we use by integrating the data we have on our [requirements technology](#).

We took the primary view of our Technology view from the [C4 concept](#). However, we do not rigorously follow the model, as we do not cover levels 1, 3, and 4. Levels 3 and 4 will change regularly; however, if we try to capture them, we feel it is best to look at them from the code or a high-level pattern. The reason we do not identify Level 1s is that we can easily observe the same on the diagram of the System Landscape, and our Context map has been confused by the term [System Context](#).

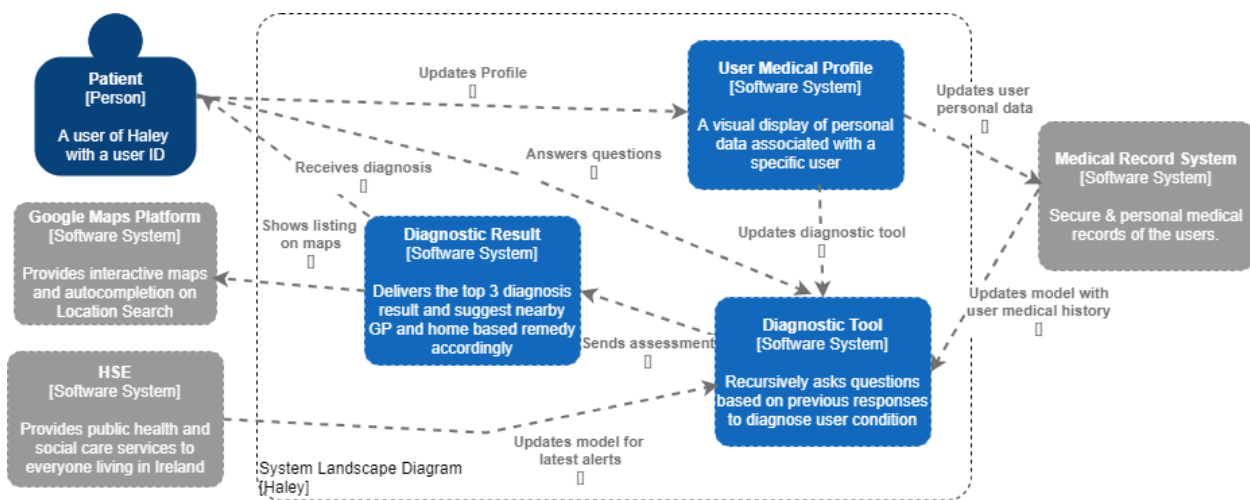


Figure 2: System Landscape (Level 2 container Diagrams are available in Appendix D)

4.4 Identity View

This view contains the strategy of how we are handling Validation, secure traffic handling, history and file system, strategy to handle attacks, and database security. Auth0 is the data front-runner for identity, so we are using it universally for the whole system. However, as users'

sensitive data is critical to our system, we have listed the following for additional security measures.

4.4.1 Security Principles

1. Traffic between clients (web, desktop, and mobile) and Haley is encrypted using HTTPS. By default, all Haley services talk to each other via a localhost connection or an encrypted SSL connection. To keep communications secure, Haley runs over HTTPS only using an SSL/TLS certificate. We will be using <https://letsencrypt.org/>.

2. Haley requires CSRF tokens in all interactions with any web API to prevent CSRF attacks.

3. Haley supports password authentication.

4. Haley supports user-uploaded files. Ideally, they should be hosted from a separate domain from the central Haley server to protect against various same-domain attacks (e.g., haley-user-content.example.com). We support hosting them on the basic LOCAL_UPLOADS_DIR file storage backend, where we store them in a directory on the Haley server's filesystem. (Appendix C)

5. Data Encryption: We will encrypt few sensitive details of the user's stored in database using field level encryption in mongodb.

4.4.2 Sequence: Unauthenticated user login Request

This sequence diagram demonstrates how, when no authentication is needed, we secure our API endpoints, but the resources are still sensitive enough to be secured, requiring captcha use.

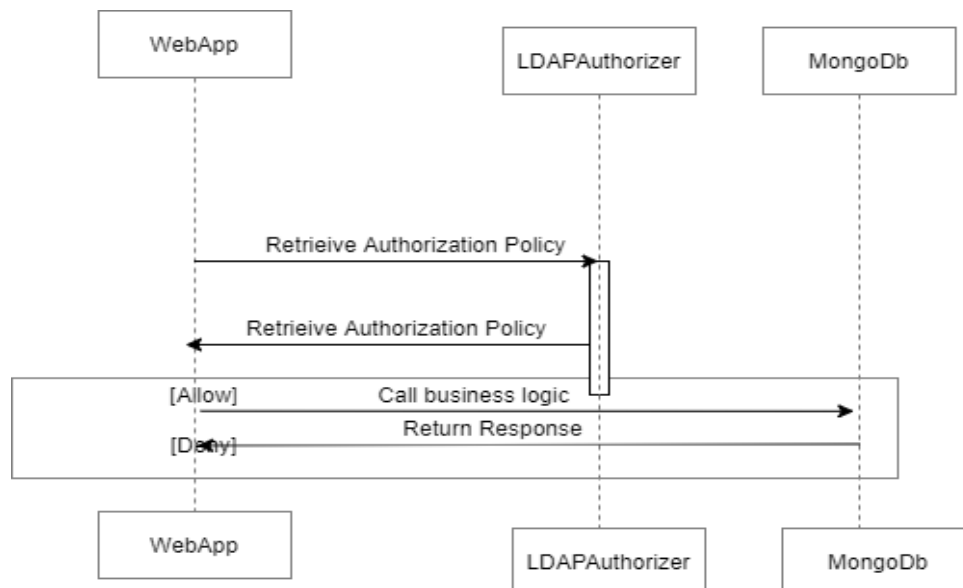


Figure 4: Unauthenticated API request

4.4.3 Sequence: User Authenticated user login Request

This sequence diagram illustrates how, when a user has authenticated themselves, we protect our API endpoint.

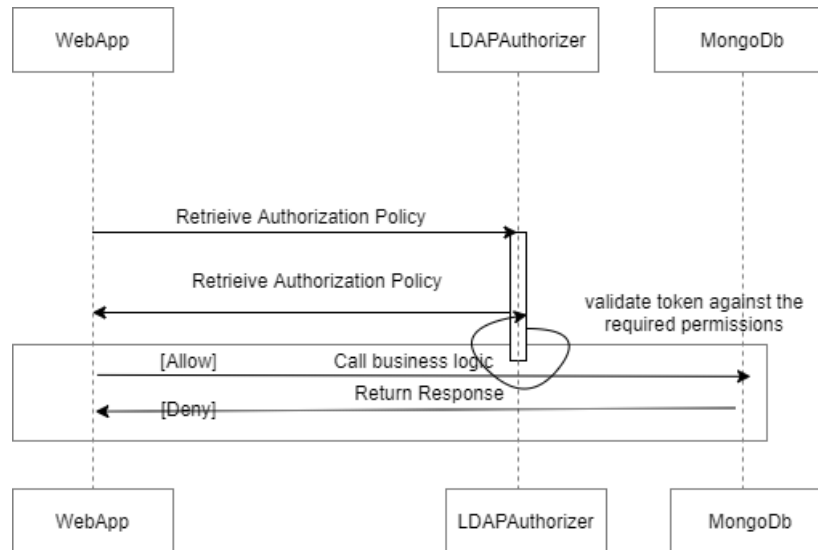


Figure 5: User authenticated API request

4.5 Technological Stack

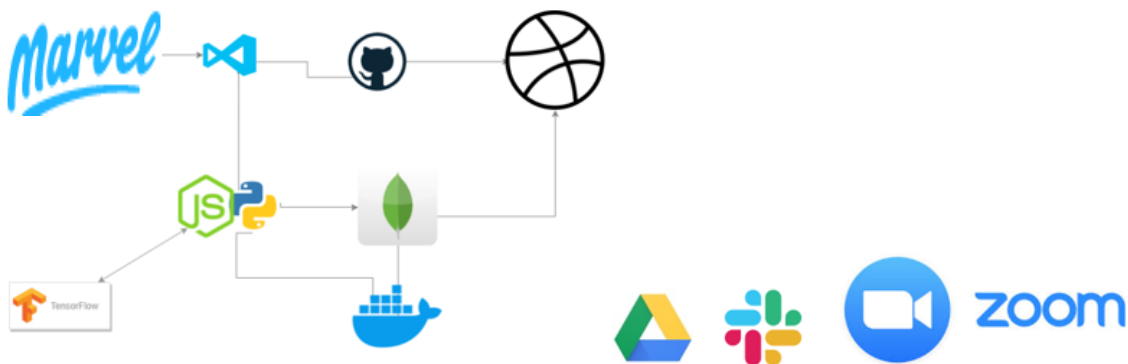


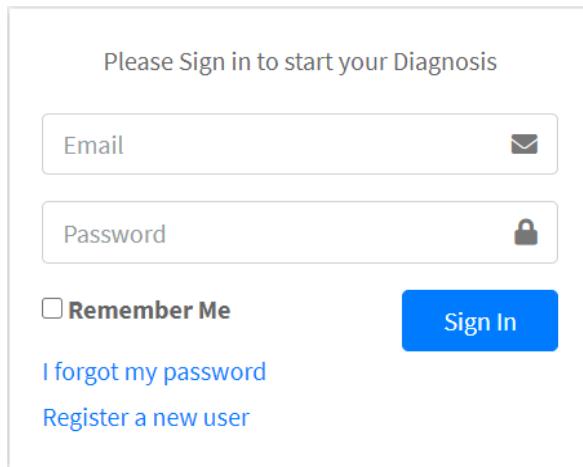
Figure 6: Technological stack

Note: Detailed description about the technical components have been provided in the [evaluation](#) section of this document


4.6 User Interface


We are using Marvel app to design the prototypes and mock-ups for our application. As we are already in the middle of the project development features like login, register, profile update, landing page, user dashboard, question generation are already implemented and fully functional. We are using bootstrap, HTML, CSS, jQuery, JavaScript and ajax to do so.

The following screenshot shows how are application initially looks, we are still working on refining these pages and making it more interactive and interesting-



Please Sign in to start your Diagnosis

Email 

Password 

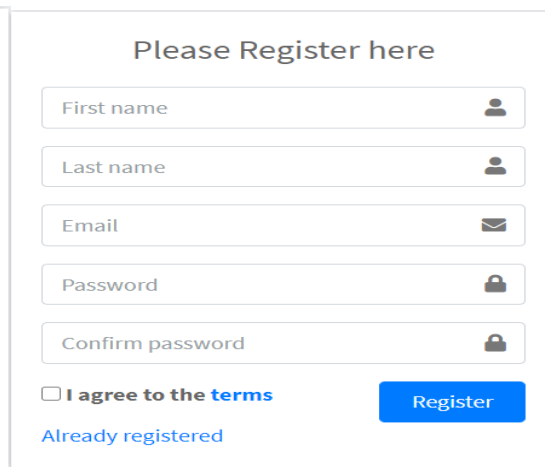
☐ Remember Me

[I forgot my password](#)


[Register a new user](#)


Sign In


Figure 7: Login Page





Please Register here

First name 

Last name 

Email 

Password 

Confirm password 

☐ I agree to the terms

[Already registered](#)

Register

Figure 8: Register Page

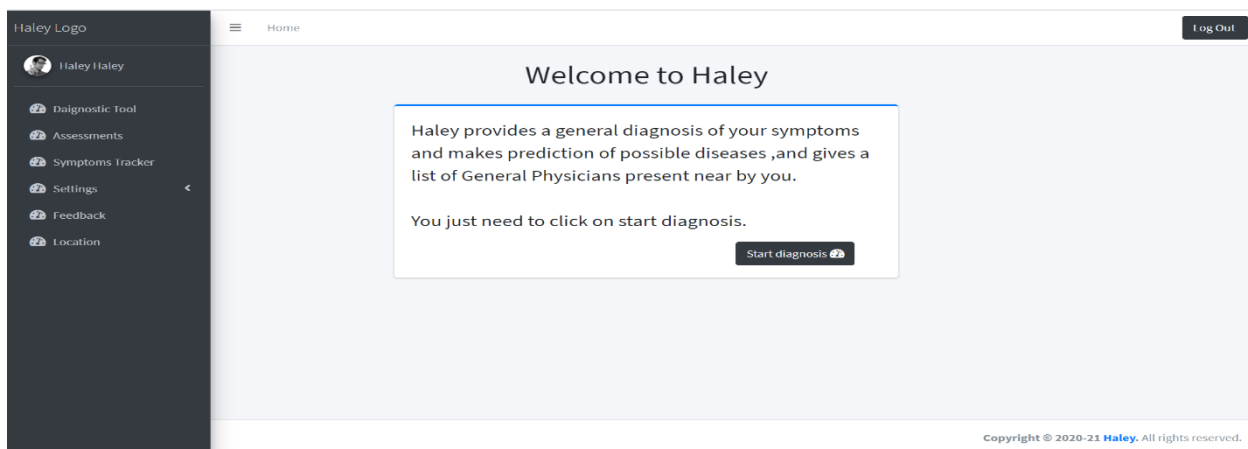


Figure 9: User Home Page

127.0.0.1:8080/health

Copyright © 2020-21 Haley. All rights reserved.

Figure 10: Update user profile Page

Copyright © 2020-21 Haley. All rights reserved.

Figure 11: Diagnostic tool Page

Copyright © 2020-21 Haley. All rights reserved.

Figure 12: Password Management Page

4.7 Intended Models

We have implemented the probabilistic model and are in the process of integrating it with the application to publish the desired results. A little more research is needed for the recommendation model which is our target for the upcoming weeks.

4.7.1 Probabilistic Model

We have decided on neural 'BayesianCNN' model for the probabilistic model. The proposed model is a combination of convolution neural networks and Bayesian posterior distribution over the CNN filters. Here CNN filters are used for feature extraction and instead of single point estimates, pooling is done on the basis of probabilistic approach. This method gives advantage over single point estimates both in terms of controlling overfitting and giving precise predictions.

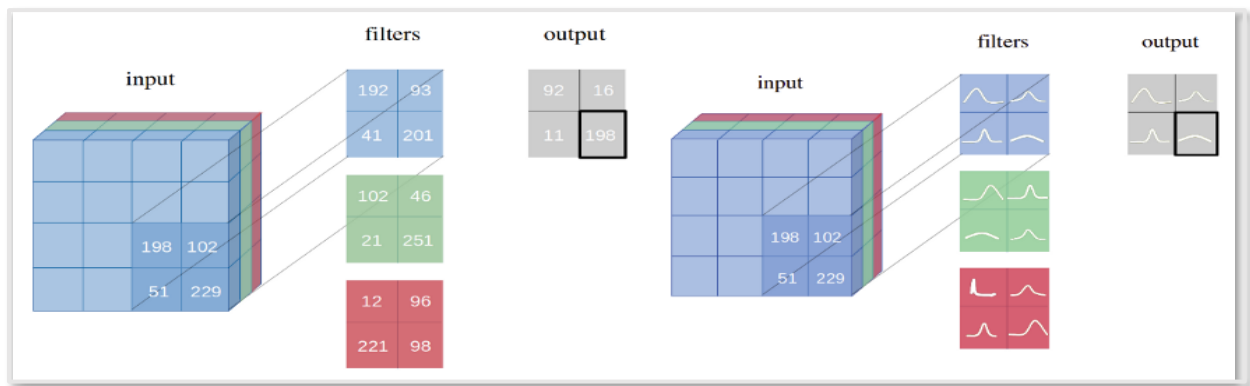


Figure 13: Network with point-estimates as weights vs Network with probability distribution as weights.²

Layers of the model are-

- 1- Input layer with shape (100,1)
- 2- 1d bayesianConv layer with shape (100,100)
- 3- 1d bayesianConv layer with shape (98,200)
- 4- 1d bayesianConv layer with shape (96,300)
- 5- Dropout layer
- 6- Global max pooling layer
- 7- Dense layer with shape (1000,500)
- 8- Dense layer with shape (500,25)
- 9- Dropout layer
- 10- Dense layer with shape (25,1) and activation as sigmoid.

The Settings for each layer are as follows-

² <https://arxiv.org/pdf/1806.05978.pdf>

Setting Type	Layer
Activation -Relu	Layer 1 to 9
Activation Sigmoid	Layer 10

Other settings are-

- a. Loss function: Sparse categorical cross entropy
- b. Train test split: 75-25
- c. Kernel size = 3
- d. Strides = (1,1)

4.7.2 Recommendation Model

We aim to make a model which takes the diagnosed disease as the input and recommends home remedies on the basis of it. The model will be trained on a dataset which contains the 3 columns being disease, home remedy and rating. The model will recommend home remedies based on this dataset. We are still in the process of narrowing down the specific model we will be using for implementing this, we aim to be done with this in the next two weeks. This is also known as collaborative filtering. The following Diagram depicts the working of recommendation model-

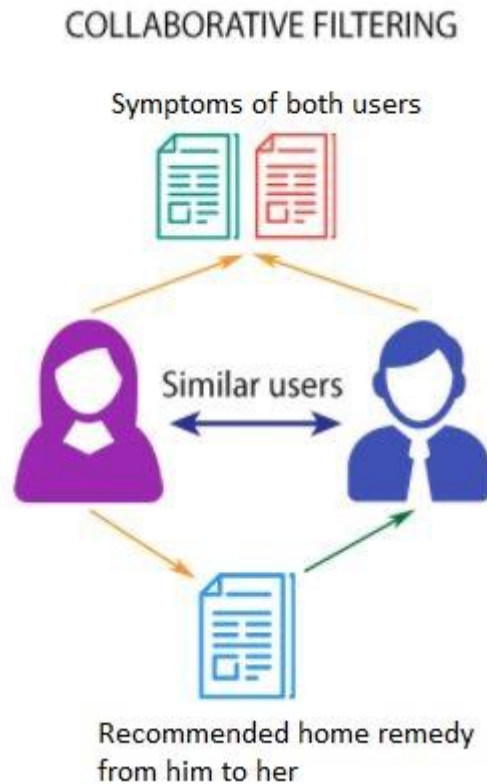


Figure 14: Recommendation Model

4.8 Data

Due to the scope of the project we were not able to find a dataset large enough to meet the requirements. To overcome this challenge, we made use of the hybrid approach for data gathering. The approach included-

1. Finding existing datasets
2. Scrapping data
3. Synthetically generating data

4.8.1 Dataset Collection and Generation

- 1- Data sources were manually scraped from websites and subreddit in a csv format

Data Source	Data Found
Kaggle, UCI, GitHub, Socrates	Data found was in csv format, which contains the diseases and their corresponding symptoms. Some of these datasets were specific to diseases related to one system of the human body or one organ.
WHO, Red Cross	Data found was in text format, which contains the diseases, its symptom, causes and its effect on the human body
Professional	After talking to professional we had a better idea of the nature of medicine and human body.

Table 5: Data Acquisition sources

- 2- Different data file was understood and then attributes corresponding to diseases were extracted, to extract the disease and deciding the scope of the project popular surveys and statistics were referred to from various government and private websites.³
- 3- Extracted attributes were passed on to Synthea Patient Generator. Synthea is a Synthetic Patient Population Simulator. The goal is to output synthetic, realistic (but not real), patient data and associated health records in a variety of formats.
- 4- Records established from Synthea were passed on to java code to populate the dataset with very low variance.
- 5- Java jar file helped us to create 1000 rows for each disease accounting to 25001 rows and 101 columns
- 6- This Generated dataset was normalized and mapped according to the disease.
- 7- Normalized dataset was grouped in 25 different sheets where each sheet corresponds to one disease, and a summary statistic was generated which showed symptoms involved in a particular disease and which are the major symptoms recorded by the patients, this step was done to ensure that there is no bias in the data.
- 8- Finally, rows were shuffled and merged back into one file using python.

4.8.2 Data Storage

To store the data, we have used the MongoDB Atlas where data can be stored in the form of documents in cloud. The reason for choosing this is the data is not in same structure and it varies from user to user. Thus, we decided to use a NoSQL database where we can store data without any schema restrictions Also, it provides more flexibility in maintaining such data in documents form, which is similar to JSON format.

5. EVALUATION

The hypothesis for the project can be defined as follows-

H1- If AI based web application is developed for health monitoring and keeping track of the past records which also provides the users' features like health tracking, symptom diagnosis, home remedies and virtual documentation then the users' can have a detailed information about his ailments and would take calculated decision to maintain their health.

³ <https://www.cdc.gov/nchs/index.htm>

<https://www.nhsinform.scot/illnesses-and-conditions/a-to-z>

<https://www.healthcheckup.com/general/most-common-diseases-in-the-world/>

5.1 Success

⁴Success of a project is different from different prospective. 'Haley' is not just about implementing what has been taught but also about continuous learning. Firstly, in terms of technology success can be defined as, understanding of python, Django, jQuery and all the other technologies used, as well as making use of the inbuilt libraries provided by the technology to increase the efficiency of the system.

Secondly, in terms of features success can be defined as, implementing all the features initially defined by the team, high precision of the diagnostics provided and helpful recommendation of the home remedies.

Thirdly, in terms of the team success can be defined as working well together, adhering to timelines, and harmonious interaction between the team members throughout the duration of the project.

Finally, in terms of users success can be defined as, good user interaction with the application, ease of access, interactive and attractive user interface, and total control on personal data stored in the application at any point of time, the feedback form that will be available to the users will help determine this.

5.2 Evaluation Criteria

The overall testing of the product comprises of the following points-

5.2.1 Technology Evaluation

It was very important for the project to choose the correct technological stack before we begin the actual development. To do so we kept few key points in mind-

- a. The solutions selected must make it possible to create the application as quickly as possible without reinventing the wheel.
- b. The program must be usable in mobile and desktop versions on Android and iOS devices and the leading web browsers.
- c. Machine learning can depend heavily on the application.

The rational behind using the technologies that we choose can be viewed in the following table-

⁴ https://www.researchgate.net/publication/262369059_Information_systems_success_measurement

Technology	Reason of Choosing
Marvel App	To make a more realistic sketch that we would use for user opinion through a survey, we selected the Marvel App. It also has integration with Microsoft Teams, so sharing individual designs became easier.
Slack	Slack provides features like Google Drive and Calendar integration which was helpful for us to maintain the documentation, code and informal discussion for the project.
Google Drive	We made a shared Gmail account to make all the documents and Google Colab Notebooks accessible to all the members at all the time. It helped us better to collaborate and distribute the task load.
Visual Studio code	We push all our code to GitHub, and due to its collaboration with VS Code, it was quickly the best way to run and test our codes. We used many extensions like '.github', '.pep8' for a better grasp of our code and folder structure.
GitHub	We selected GitHub was for its accessible features of pushing code into a single repository and making use of ZenHub so that our project followed a set timeline via Sprints and Milestones. Also, we have made our code open source, and given GitHub is the single most significant source for it.
Django(Python)	We are using Neural Networks, and given Python is the golden standard of Machine Learning based web development at the moment. Using it in production means we can deploy our server architecture. Also, learning Django is both relatively easy and suitable for future employability
MongoDB	At this stage of development, we have limited our data volume, and we wanted a pure NoSQL candidate that was document oriented. MongoDB was the best free solution that could implement.
Docker	With the numerous Docker containers, the architecture as a code allows us to write what we want and start all at once with one command. All the architecture, according to our scripts, will be deployed automatically

Table 6: Detailed Technological stack evaluation

5.2.2 Initial Evaluation

The purpose of this is to understand the need of the users, their willingness to use this product, and to identify what the users want from the system. To do this we designed surveys, users from

different demographics and belonging to different age groups took part in the survey and the results obtained were satisfactory. The first survey organized was to understand the user groups, and general mentality of the users to the product, the results were satisfactory and after analysis the user groups were much clearly defined, and the features were finalized. The second survey was organized to gather the input from the users regarding the prototypes developed by the team, some interesting feedback was received which was then taken into consideration in future development.⁵

Note- Detailed survey analysis can be found in the document attached and named Survey Analysis.docx

5.2.3 Smoke Testing

The purpose of this testing is to keep getting suggestions and check the functionality of the system at the time of development so that a high quality of product can be delivered at the end. The following table depict what is involved in this testing- Firstly, after development of any feature the team within itself conduct testing to see whether the feature is working fine or not. Secondly, the demonstration of the feature will be given to the mentors / peers and insights will be gathered from them to improve the system.

Type	Description
Functionality	Verify that the functionality is working fine on every deployment
Libraries	Verify all libraries are properly called and used
Journey	User-like browser navigation
API	Hitting APIs, no browser involved
Component	Top-level Vue.js test
Unit	Business rules for data conversion.

Table 7: Smoke testing criteria

⁵ Google forms was used for organizing surveys links as follows-

Product Survey - <https://forms.gle/4a19qDBGTqEyxzJR9>

Prototype Survey- <https://forms.gle/36wPm9AAfCK2xCEV9>

5.2.4 System Testing

The purpose of this testing is to determine the working of the system as a whole after development. This further includes –

- a. Feature Testing- Testing the system for proper working of the features and its results.
- b. Result Analysis- Analyzing the results of the models developed and checking for false positives, precision, loss, and comparison of success rate vs error rate. The following table depicts how we aim to evaluate the model results-

Medical Problems	NB	LR	K*	DT	NN	ZeroR
Breast Cancer wise	97.3	92.98	95.72	94.57	95.57	65.52
Breast Cancer	72.7	67.77	73.73	74.28	66.95	70.3
Dermatology	97.43	96.89	94.51	94.1	96.45	30.6
Echocardiogram	95.77	94.59	89.38	96.41	93.64	67.86
Liver Disorders	54.89	68.72	66.82	65.84	68.73	57.98
Pima Diabetes	75.75	77.47	70.19	74.49	74.75	65.11
Haebberman	75.36	74.41	73.73	72.16	70.32	73.53
Heart-c	83.34	83.7	75.18	77.13	80.99	54.45
Heart-statlog	84.85	84.04	73.89	75.59	81.78	55.56
Heart-b	83.95	84.23	77.83	80.22	80.07	63.95
Hepatitis	83.81	83.89	80.17	79.22	80.78	79.38
Lung Cancer	53.25	47.25	41.67	40.83	44.08	40
Lymphography	84.97	78.45	83.18	78.21	81.81	54.76
Postoperative Patient	68.11	61.11	61.67	69.78	58.54	71.11
Primary tumor	49.71	41.62	38.02	41.39	40.38	24.78
Wins	8\15	5\15	0\15	2\15	1\15	1\15

Table 8: Predictive Accuracy of Bayes and other Technique (Al-Aidaros, Bakar, & Othman, 2012)

- c. Security- Checking if the system is truly secure as the information dealt with is medical and therefore sensitive in nature.

5.2.5 User Acceptance Testing

The purpose of this testing is to get the target users involved with the system and get their feedback on the system. This testing will take place across multiple users and different demography's. The users will be given a set of tasks that they will perform on the developed applications, they will be able to judge the system based on the ease of access, profile creation, symptom diagnosis, user interface etc. The users will then be asked to fill out a questionnaire designed about the system. The result of this will be taken into consideration and used to improve the system if possible or taken down as learnings.

5.2.6 Implications of GDPR

Since the application is dealing with sensitive information the application must have the following-

- a. User must have total control over their personal data i.e. a user can update, delete any information personal to them whenever they want.
- b. The username and passwords should be secured when passed to the application.
- c. All PII data stored by the system in the database should be encrypted and secured.
- d. No PII data should be passed to any users via text, emails etc. without proper password protection and encryption.
- e. All medical files uploaded as images/ documents should be secured while storing.
- f. At the time of login users will be provided a detailed description of what data is collected by the system, how it is stored and its purpose of collection and usage.

5.2.7 Comparison

The final phase of evaluation will be evaluating the product developed with the already existing product available. The goal is to develop a product that the users will prefer to use in place of all the other products that are already available in the market currently. The table below depicts the criteria for comparison-

Evaluation Criteria	Description
Features Available	Are we providing all the features from available in the different products and more?
User-Interface	Is our user interface more interesting than that of other application?
Preference of the users	Do users prefer our application over the others available?

Table 9: Comparison criteria's

6. CONCLUSION

6.1 Project Management

Agile methodology is used for software development. Agile is a continuous development – continuous integration methodology that helps break down the project into smaller more manageable pieces. The total timeline of the project is 13 weeks, each week is considered as an individual sprint with some story points to cover.⁶ As per agile the following is done for the project-

⁶ <https://www.wrike.com/project-management-guide/faq/what-is-agile-methodology-in-project-management/>

1. **Product Backlog-** A product backlog is created which contains all the tasks that needs to be done for completion of the project. The tasks are divided into epics and issues. An epic is the feature of the project and the issue is the corresponding user stories for the same, a high level estimation of the time taken o complete each task is also defined so as to help the team adhere to timelines.
2. **Sprint Backlog-** The sprint backlog is essentially the tasks and features that are taken into consideration for the current sprint.
3. **Kanban Board-** To keep track of all the issues and their status a Kanban board is maintained using ZenHub as a tool. Typically, the board has 4-5 headings being, Tasks, In-progress, Done, Review and challenges. For this project, the Kanban board is modified to contain the product backlog as well.
4. **Rotating Scrum Master-** Usually there is only one scrum master throughout the sprints in a project but for this project the scrum master rotates between the team members on a weekly basis, this helps with all team members gaining an experience of the responsibilities of this position and also makes the team more caution while working through the project.
5. **GitHub-** For repository maintenance and collaboration between the team. GitHub tools helps in maintaining version throughout the project and helps maintain the code on a virtual platform. GitHub is also linked with ZenHub that helps in creating boards and maintaining issues for the project as it is closely linked with the repository. ZenHub also has function of reports which helps visualizing the project progress through burndown charts and velocity tracking. Ideally, the team aims to deliver 35-40 story points each sprint.
6. **Virtual Environment-** Taking into consideration the current pandemic situation it is not possible for the team to physically meet and work together. To handle the challenge zoom is used as a communication tool, every day at 1100 hrs, daily scrum call is scheduled where the team updates each other with their previous work and discuss what they are going to do next as well as any challenges they are facing, the team also uses this point to reassess their priorities if needed. Individual zoom calls are also frequently done between the team members for efficient working. Slack is used as a platform to share our ideas and general informal project discussion. Google calendars are used to help work out a schedule between the team. Google drive is used for maintaining any document that needs constant updates or tracking.
7. **Retrospect-** A retrospect is maintained and done by the team at the end of the sprint to go over the previous week and understand what went wrong or what was good in the previous sprint to help take the learnings of the previous sprint and work with them in mind on the current sprint.

- 8. Team Conflicts-** Handling team conflict is one of the top priorities of the team as the team acknowledges that only a harmonious environment can provide good results. The team follows the guidelines Dr. La Jolla⁷ for the same. The team follows the following steps to do so-
- a. Take a break
 - b. Reflect and sort out your thoughts
 - c. Find a third non-bias party to discuss the issue
 - d. Understand how the problem originated and discover any underlying issue
 - e. Listen to all sides and finally decide the best possible solution
 - f. Have a virtual get-together which is not work related and have some fun.

6.2 Challenges

Six weeks into the project the team has come across various challenges and found solutions to some and searching for solution for others some of the challenges yet to be resolved are as follows-

1. Limited users – Due to the current health concerns it will not be possible for the team to find a lot of users for user acceptance testing, the current estimated users are only 30. The team is still working on finding a solution for the same.
2. High Run time- The processing speed of CPU's is limited and hence while testing the results of the models it may take more than 8-10 hours to run the model even after cutting the data into smaller portions. The team is working on ways to enhance this, but no solution is found until now.
3. Scalability – The current system deals with only 25 diseases. The dataset used for this system is also synthetically generated following the principles provided by already existing systems and scrapping of data. If in the future the number of diseases needs to be increased, then the dataset will need to be re-scrapped and generated which will take a lot of man hours. A more flexible and easy solution still needs to be found for this problem.

6.3 Limitations

1. Scope- The scope of the project covers only 25 diseases as of now and hence any other disease cannot be diagnosed by this system.

⁷ <https://blink.ucsd.edu/HR/supervising/conflict/handle.html>

2. Space- The space available to create and maintain the system is limited to 512 MB which is sufficient for now but there is a limit on the number of users who can successfully use all the features of the system.

6.4 Timeline

For our sprint mile stones within the project duration, we have created issues and epics and assigned them to weekly sprints in a Zenhub board. During the first six weeks, the foremost thing which we have learned is time management to complete our weekly sprints and project management through Zenhub.

We have covered the functionalities like User registration, user authentication, updating the user profile, and diagnostic tool, our app's key feature. Now moving forward, we have planned to complete our other required functionalities such as integrating users' answers to predict the results, GP's details, assessments feature, and user's feedback until week nine, after which we would have user evaluation based on which we would make necessary changes in our app.

When it comes to estimating time for the project, we have identified all the remaining work. As a part of this, we have allowed time for meetings, communications, and other discussions on some functionalities or challenges. We have ordered the activities required for the individual functionalities, along with few deadlines as well. According to the priorities, we are working on the individual tasks to carry out the functionality and coordinate the tasks to complete functionality.

The next step for development are defined as follows-

Sprint	Task	Week
Sprint 6	<ol style="list-style-type: none"> 1. Finishing Diagnostic tool 2. Starting Recommendation Model 	7
Sprint 7	<ol style="list-style-type: none"> 1. Refining UX/UI 2. Google Map API integration 3. Recommendation model implementation 	8
Sprint 8	<ol style="list-style-type: none"> 1. Finish UX/UI 2. Uploading feature implementation 	9
Sprint 9	<ol style="list-style-type: none"> 1. Final Report Creation- Draft 1 2. System Testing 3. Result Analysis 	10
Sprint 10	<ol style="list-style-type: none"> 1. Final Report Creation- Draft 2 2. Peer Review 	11
Sprint 11	<ol style="list-style-type: none"> 1. Final Report Creation- Draft 3 2. Individual Report Creation- Draft 1 	12

	3. User Acceptance Testing	
Sprint 12	1. Final Report Finalization 2. Competitive Analysis 3. Demo Creation	13

Table 10: Upcoming Sprint Timelines


7. REFERENCES

1. Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification With Naive Bayes Approach. *Information Technology Journal* .
2. <https://arxiv.org/pdf/1806.05978.pdf>
3. <https://medium.com/neuralspace/bayesian-neural-network-series-post-1-need-for-bayesian-networks-e209e66b70b2>
4. <https://github.com/adalca/medical-datasets>
5. <https://www.kaggle.com/itachi9604/disease-symptom-description-dataset>
6. <https://hpo.jax.org/app/download/annotation>
7. <https://github.com/leanderme/sytora>
8. <https://www.kaggle.com/plarmuseau/sdsort/home>
9. <https://www.kaggle.com/plarmuseau/primer>
10. <https://www.kaggle.com/plarmuseau/symptom-disease-recommender>
11. <https://github.com/Aniruddha-Tapas/Predicting-Diseases-From-Symptoms>
12. <https://themanifest.com/mobile-apps/how-develop-healthcare-app>
13. <https://www.malacards.org/>
14. <https://www.kaggle.com/moradnejad/nhanes-questionnaires-datasets-20172018-csv?>
15. <https://www.kaggle.com/johnsmith88/heart-disease-dataset>
16. <https://www.kaggle.com/mansoordaku/ckdisease> <https://www.kaggle.com/priya1207/diseases-dataset>
17. <http://ecg.mit.edu/time-series/>
18. <https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>
19. <https://physionet.org/data/>
20. <http://47.93.42.104/544#>
21. <https://ada.com/>
22. <https://github.com/GroupProjectSem3/synthea>

APPENDIX A

A.1 – Katie Patel






Katie Patel



Age
21 Years

Highest Level of Education
Associate degree (e.g. AA, AS)

Social Networks

Gender
Female

Bio
Kate is studying law at TUD, she's a vegan and maintains her health with her personal training coach 3 days a week. With her free time, she enjoys going to underground jazz clubs and nights out with her friends in the city.

Work
College Student

Family
Single

Attitude to Technology

- Familiar with basic mobile applications
- Plays lot of games on computer

Goals or Objectives

- Become a Professional cyber crime lawyer
- Travel all countries in the world

Life Motto
Katie just moved to Ireland and wants best medical guidance along with a new GP

Strengths

- Practical

A.2- Lyra Foster

Lyra Foster



Age
34 Years

Highest Level of Education
Bachelor's degree

Social Networks



Gender

Female

Bio

Lyra is a high school teacher. She prefers to stay at home on the weekends and she spends a good portion of her free time helping her children thrive and grow. Her hobbies include making ceramics, cooking, and volunteering. She quotes, "My time is valuable and I do not want to waste it by submitting questions to health experts that go left unanswered."

Work

Hlgh School Teacher

Family

Married with kids

Goals or Objectives

- To get consultation and manage medical documents of the kids at one place
- Complete PhD

Attitude to technology

- She is pretty familiar with basic mobile and web applications.
- She use social networking sites alot

Life Motto

To be the best version of current self.

Strenghts

- Practical
- Organized
- Creative

A.3 – Zoro Uzumaki

Zoro Uzumaki



Age
53 Years

Highest Level of Education
Master's degree (e.g. MA, MS,

Social Networks



Gender

Female

Bio

Zoro is a Product Director at a corporate investment and incubation firm. When not working around the clock, he enjoys going to crossfit 4-5 days week, cycling, and training for triathlons on the weekends.

Work

Tech Product Director

Family

Married

Attitude to Technology

- Well versed with IT terminologies
- Uses Mobile extensively for calling
- Is familiar with technologies like GitHub, Zenhub, Scrum.

Goals or Objectives

- Feels bad wasting and burdening peoples time if they are unable to provide service needed.
- Inspired by tennis pro Roger Federer.

Life Motto

Zoro wishes preventive body care with no injuries

Strengths

- Analytical
- Passive
- Punctual

A.4- Aine Danley

Aine Danley



Age
42 years

Highest Level of Education
Master's degree (e.g. MA, MS,

Social Networks



Industry
Enter text here

Organization Size
Self-employed

Gender

Female

Bio

Aine is a NGO worker who devotes all her time helping underprivileged kids get basic necessities of life such as food, clothing, and education. She prefers to stay at home on weekends and read poetry. Her hobbies include dancing, cooking and knitting.

Work

Human Rights activist

Family

Widow with kids.

Goals and Objectives

- To help as many underprivileged children as she can
- Stay in contact with her children who are adults themselves.

Attitude to technology

- She is familiar with basic computer application like MS Office and can surf the internet
- She is not much into social media and finds it difficult to use

Life Motto


To live a fulfilling life

Strength

- High EQ
- Organized
- Punctual

A.5- Harvey Spector






Harvey Spector



Age
58 years

Highest Level of Education
Professional degree (e.g. MD,

Social Networks

Industry
Enter text here

Organization Size
Self-employed

Gender
Male

Bio
Harvey is a lawyer. He prefers to fight one case at a time and give guest lectures to youngsters, he spends most of his time researching the laws and articles related to them. His hobbies include socializing with fellow lawyers, drinking, and debating.

Work
Lawyer

Family
Married with kids and grandkids.

Goals and Objectives
To retire and spend time with his grandkids.

Attitude to technology
He is not familiar to technology and finds it difficult to operate.

Life Motto
To be kind and live life happily

Strength

- High IQ
- High EQ
- Kind

A.6- Mark Smith

Mark Smith



Age
30 years

Highest Level of Education
Associate degree (e.g. AA, AS)

Social Networks



Gender

Male

Bio

Mark is a travel enthusiast who uses his video blog to share his experiences from different countries he travels and archives to get lots of subscribers, likes, and comments. He freelances as a video editor for different productions. In his free time, Mark enjoys listening to music, gardening and reading about world history.

Work

Freelancer, Blogger

Family

Single

Goals and Objectives

- To travel the world on a budget.
- To be healthy and have his medical documents always on hand

Attitude to technology

- He is very familiar with using high end applications for video editing
- He is also very active on social media platforms

Life Motto

To be healthy and travel the world

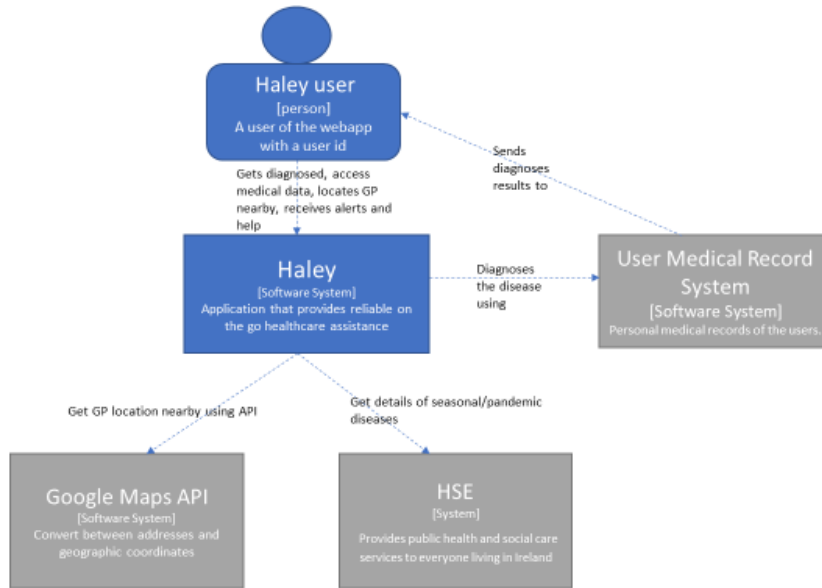
Strengths

- Practical
- Organized
- Creative
- Time Management

APPENDIX B

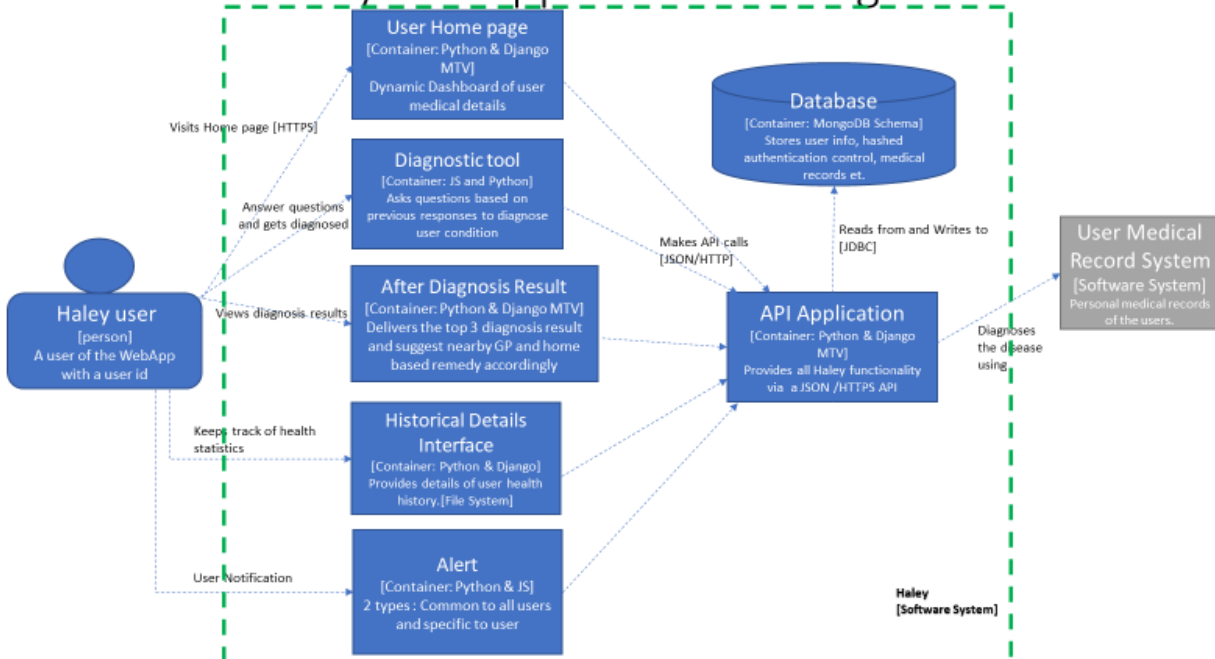
B.1- Haley system context diagram

Haley System Context Diagram



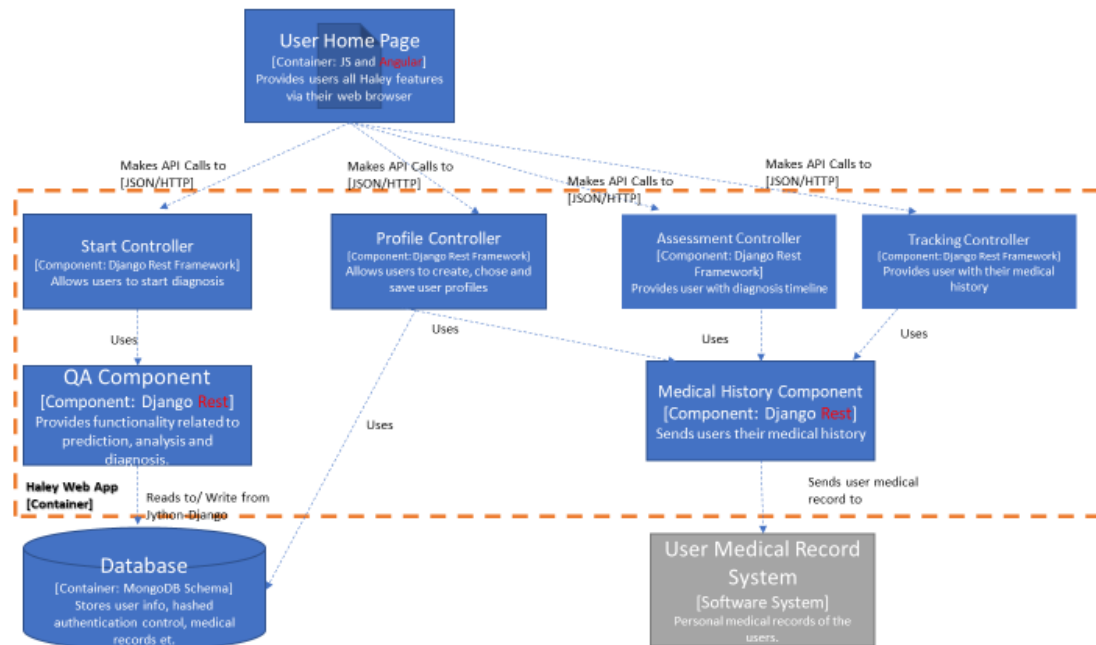
B.2- Haley Webapp Container Diagram

Haley Webapp Container diagram



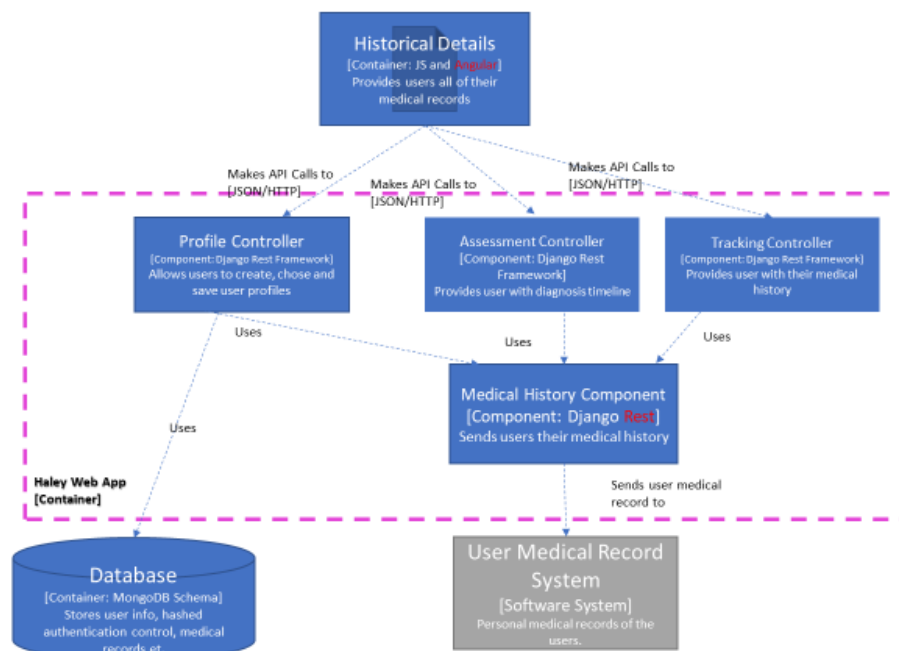
B.3- Component Diagram for Haley WebApp – User Home Page

Component diagram for Haley WebApp – User Home Page



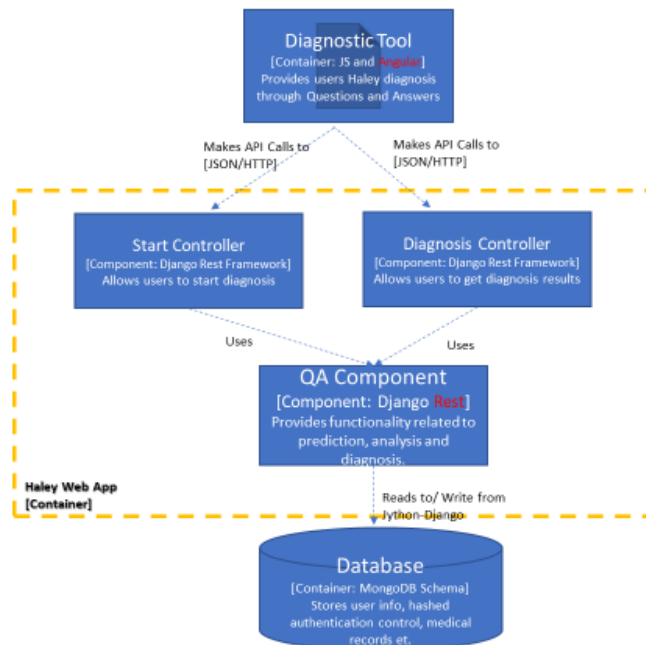
B.4- Component Diagram for Haley WebApp – Historical Details

Component diagram for Haley WebApp – Historical Details



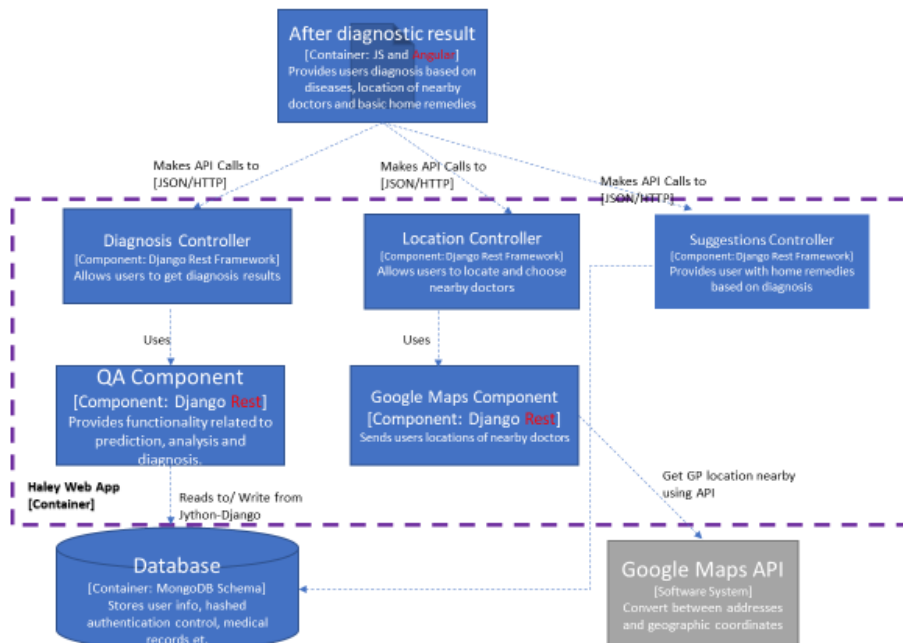
B.5- Component Diagram for Haley WebApp – Diagnostic Tool

Component diagram for Haley WebApp – Diagnostic Tool



B.6- Component Diagram for Haley WebApp – After Diagnostic Result

Component diagram for Haley WebApp – After diagnostic result



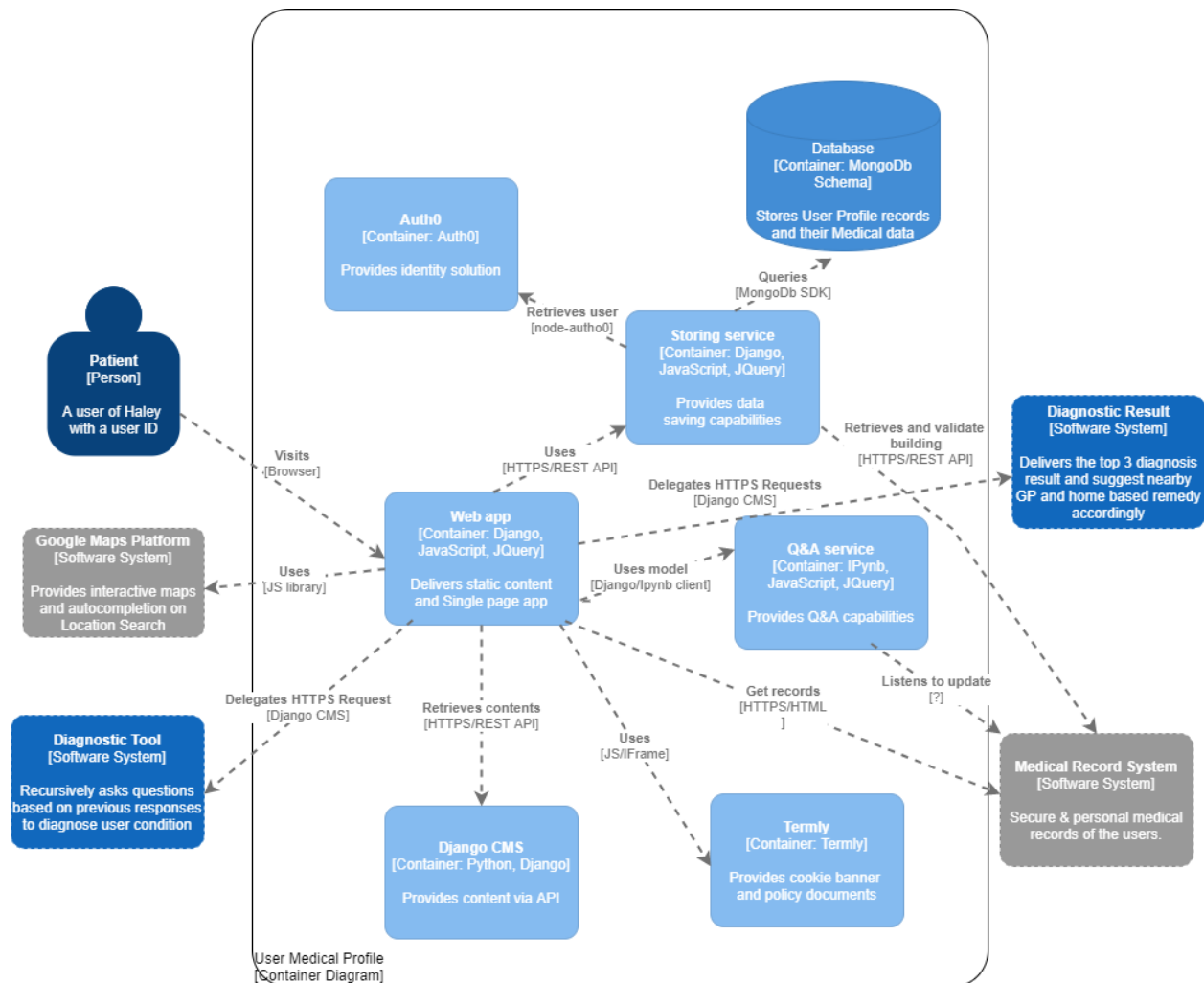
APPENDIX C

Notes: 1. We typically perform DDD forms of Core, Supporting, or Generic at the subdomain level in practice. We understand that the priority of becoming more granular in the Bounded Setting is worth further focusing on, which is why we introduce the color-coding to the Bounded Setting level.

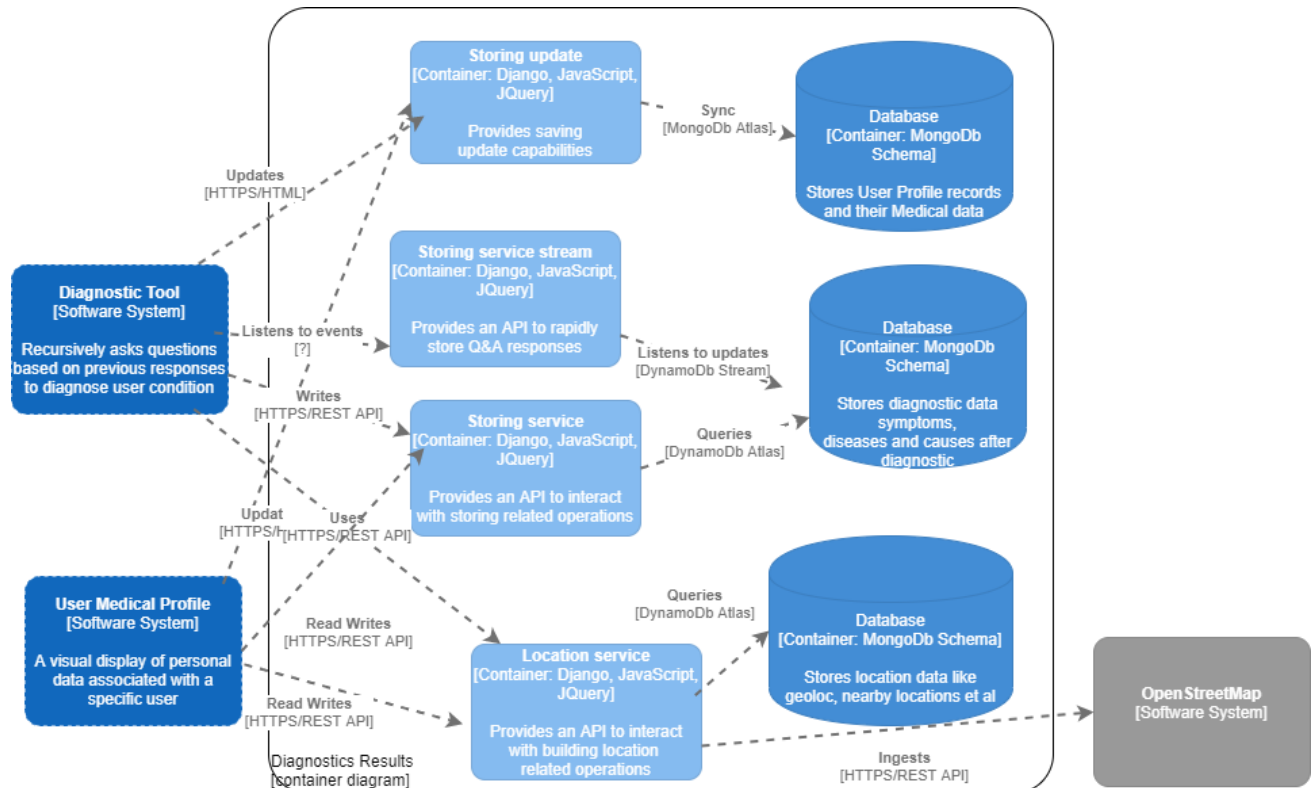
2. In practice, the DDD scope map is generally drawn to represent the facts. We will find the combination of the current state and target state since we use the context map as a design tool here.

APPENDIX D

D.1- Level 2 Container 1



D.2- Level 2 Container 2



D.3- Level 2 Container 3

