Syracuse University
College of Engineering and Computer Science

Lung Cancer Detection System
System Requirements Specification

Qiaolongfei Guo, Haoyu Gong, Qinghan Li

Professor Yu

October 26, 2019

Contents

1. INTRODUCTION

1.1 Purpose.

The purpose of this document is to provide detailed description of lung cancer detection system, which is a web application for lung cancer detection based on CT images and making recommendations of doctors or hospitals nearby the user. This document will provide the purposes, requirements, interfaces and system structures of the system. The document's audience is intended for system stakeholders such as customers, developers, testers, project managers and the software architects.

1.2 Product Scope.

The inspiration and motivation of the project are from the real experience of the team: people are suffering due to limited medical resources especially in the developing countries. Therefore, the main challenge today is how to improve the efficiency of the medical process via the technologies. The team will choose lung cancer as an example and try to give a solution: to build a web-based application that can accelerate diagnosis, process of choosing doctors/hospitals and making appointments. Due to the seriousness of diseases such as lung cancer, medical organizations may take more man power and time to confirm diagnosis. Therefore, the team plans to implement this by using the machine learning technology to save man power and time. Additionally, different doctors/hospitals always have different expertise in different diseases, therefore the system is designed to recommend the best treatment to the users once the diagnosis is confirmed. Lastly, the system should provide an appointment system for the patient and doctors to arrange for a proper time once the patients decide to choose a doctor.

From the customer's perspective, the medical organization wants to improve their efficiency in order to improve their profitability; and the patients want to decrease the whole processing time in order to get their best treatment as soon as possible. Therefore, this application should have a broad market and improve the quality of life especially in the developing countries.

Figure 1– Crowdedness in the hospital in China

2. OVERALL DESCRIPTION

2.1 Product Perspective and Functions:

The lung cancer detection system will be using techniques in machine learning to make diagnosis to the patients. The system requires user to upload their CT images of the lungs to detect the existence of potential tumors. Both the patients and the doctors are required to log in by using Google account and the doctors are required to be certified as a doctor. If the system detects tumor in the image, the system will make recommendations of medical organizations and show their location on the map. Eventually, the patients should be able to see the schedule of doctors who are members of the application, and the user should be able to make appointment directly.

A high level diagrammatic representation of the user activity is depicted below.

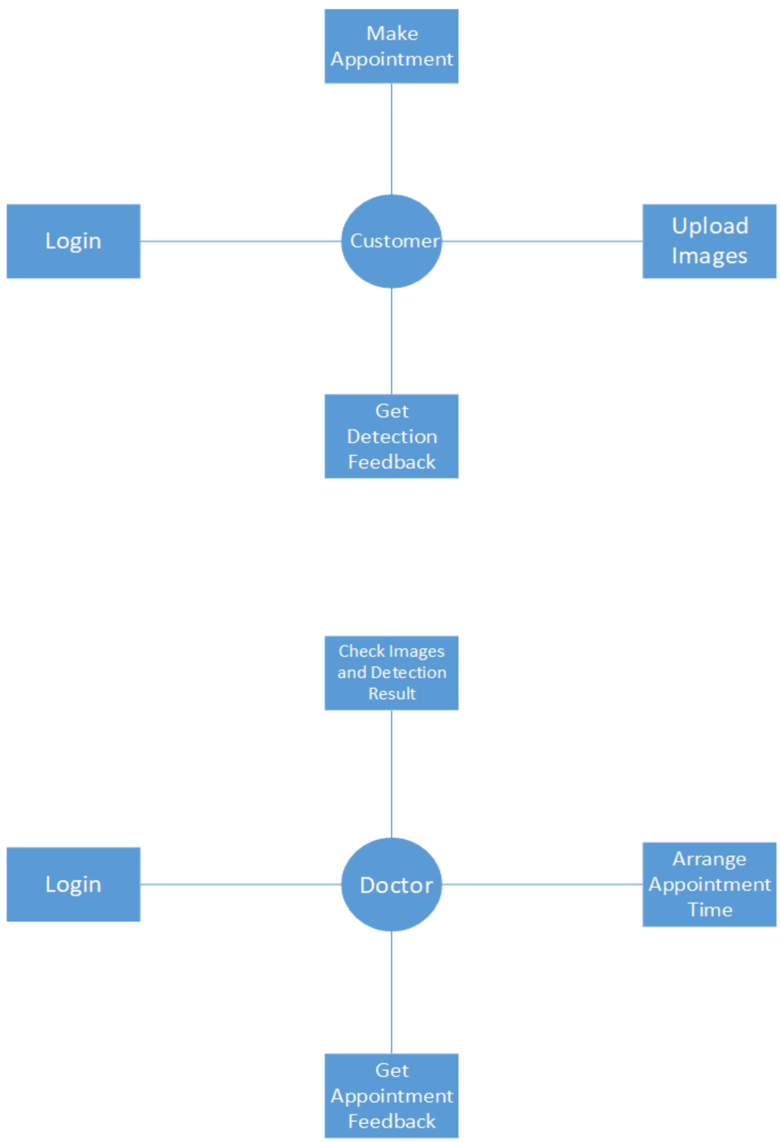

Figure 2 - High Level Activity Diagram

3. INTERFACE REQUIREMENTS

3.1 User Interfaces:

The user interface of the app will utilize HTML, CSS, PYQT and the Google maps. There are two different types of users of the system, patients and doctors. There will be difference for different users such as the user interface for doctors should not contain Google maps since the doctors don’t need to check the locations of medical organizations in our system. For the users, the Google map should be able to show the recommended medial organizations. Additionally, the appointment system is different too where the doctors should provide schedule to the users, and the users should be able to pick one period of available time.

The system will behave consistently across multiple platforms and browsers. Each user flow will have a login screen and a dashboard. To keep the views consistent across the system and easy to use, all users shall use the same to register/login. The user shall just need to select the role and login to the system using a Google account. For the doctor users, they are required to provide their identification in order to be a member of our system for safety and legal reasons.

3.2 Software Interfaces:

The system will be using third party APIs in order to achieve the functionalities. The team will be using TensorFlow mainly to process the images. A more detailed description of the lung cancer detection algorithm will be introduced in the system architecture section.

Google account authentication APIs will be used to provide authentication to the users of the system. Google map APIs will be used for the medical organization recommendation section.

Once the user is authenticated, its previous diagnosis history should be stored in the account. The system will be using the MongoDB as the database to store all kinds of data.

A third-party Python-based framework Django will be used to make the doctor appointment system. Additionally, a third party API Nylas Scheduler might be used for better appearance purposes.

4. REQUIREMENTS

4.1 Functional User Requirements

F-1 The system shall have personal user home page for Lung image uploading.
F-2 User shall identify themselves to the system by using account and password.
F-3 The system shall allow user to review their former history.
F-4 The system shall recommend nearby hospitals/doctors for further diagnosis.
F-5 The system shall allow the user to make appointments with doctors.
F-6 The system shall allow the doctors to register in order to make appointments with patients.

4.2 Functional System Requirements

F-1.1 User should use their Google account to log in this system.
F-1.2 User could create a new account if they don't want to use their Google account.
F-1.3 Both kinds of user could have their own special URL address to show profile.
F-1.4 A user shall not be a doctor and a normal user at the same time.
F-1.5 The doctor verification shall allow user to upload related document.
F-2.1 This system shall use Google Identity Platform sign-in features.
F-2.2 This system shall allow a brand new sign-in features for those who choose not to use Google account.
F-2.3 If user fails to sign in, this system shall have the ability to handle the error and leave customer a notice.
F-2.4 If user signs in this system successfully, they shall be directed into the main functional page.
F-3.1 The system shall use database technique to save user's former history.
F-3.2 The system shall allow user to download their own former history at any time.
F-3.3 The former history shall be top-secret, and don't allow everyone else to check but the account owner, only if get the permission.
F-4.1 The system shall show the normal user to check the closest hospital or clinic by Google Map.
F-4.2 The system shall request doctors to provide their detailed information, including their address, if they are willing to arrange an appointment on this system.
F-4.3 This functionality does not allow any commercials.
F-5.1 When a normal user arranges an appointment with a doctor, he/she must provide their necessary information, like name, phone number, etc.
F-5.2 If a doctor wants to use this functionality on this system, he/she needs to update their available time as soon as possible.
F-5.3 This system allows doctors and normal users to change their appointment before the arranged time 24 hr.
F-5.4 If an appointment is arranged, the system shall show the update immediately.
F-6.1 Doctor shall use their Google account to log in this system.
F-6.2 Doctor need to pass the identity verification by uploading related hard copies.
F-6.3 Doctor shall complete their own profile to be shown on the map.

4.3 Non-Functional User Requirements

NF-1 This system should be available to anyone who is interested, without any previous training.
NF-2 This system shall save all users' data safely and properly.
NF-3 This system shall not allow any actions about sharing users' data, unless arranged upon by user.
NF-4 This system shall provide different roles of login, normal user or doctor.
NF-5 This system shall require doctor users to prove their identity in any form.
NF-6 The performance, development, and maintenance of this system shall be the most recent and suitable.
NF-7 If this system has some important news to announce, like doing maintenance, it shall publish news prior to the action about 24 hours.
NF-8 This system should be designed in the form of adding extended functions in the future easily.
NF-9 The input must be a valid image.

4.4 Non-Functional System Requirements

NF-1.1 Users could access this system by a PC, or mobile phone with Internet connection.
NF-1.2 Users need to use a web browser to visit this system.
NF-2.1 This system shall provide a safe login system on its own.
NF-2.2 In order to provide a more convenient way, this system shall allow users to use their Google account to sign in.
NF-2.3 This system shall not change anything about users' data in the backstage in any case.
NF-3.1 This system shall not share personal data, including contact information and location to any third parties.
NF-3.2 This system shall not allow any access from information crawlers.
NF-4.1 A normal user could sign in this system by a new username and password, or their Google account.
NF-4.2 A doctor must provide their supportive documents to access related functionalities in this system.
NF-5.1 As a doctor user, he/she could upload any forms of supportive document to this system.
NF-5.2 System admin shall check those uploaded supported documents to give the priority of doctor user.
NF-6.1 This system shall be developed by Python 3.8, latest Google APIs, MongoDB, and Pycharm 2019.
[No System requirements for Non-Functional User Requirement NF-7]
NF-8.1 This system shall use an open style UI for the future new functionalities design.
NF-8.2 The whole frame of this system shall be tolerant to the future design.
NF-9.1 If the input is not valid, the system shall return nothing and send a notice to the user.

5. MODELS

5.1 Use Case Diagram

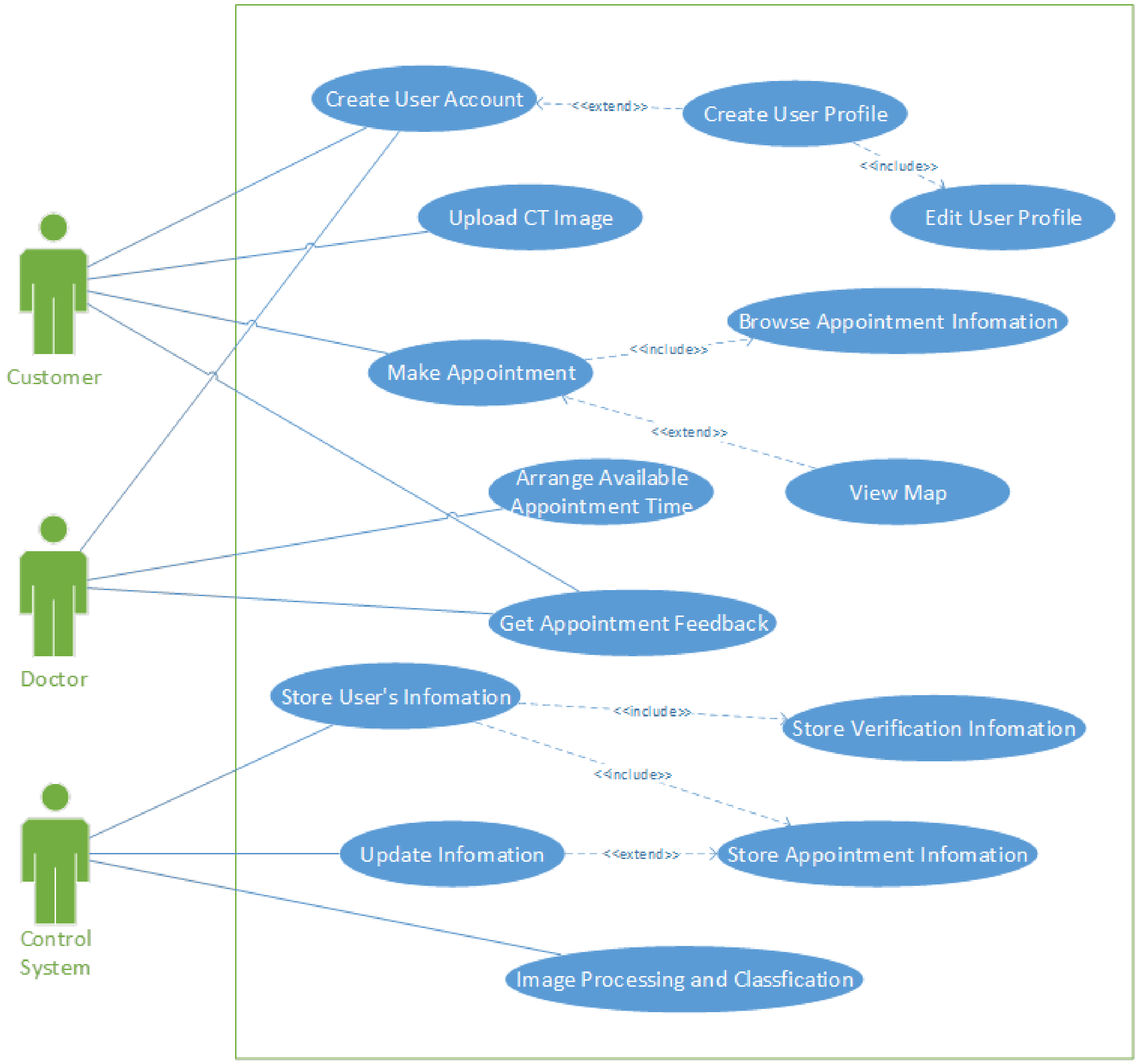

Figure 3 - Use Case Diagram

This use case diagram represents three actors(Customers, Doctors and Control System) and the use cases below:

Create User Account - Both the customers and doctors could create their own account with Google Account or a completely new account, and design their profiles.

Upload CT Image - Customers could upload one or several lung CT images to the system to get the result of automatically detection.

Make Appointment - Customers could browse the information of available doctors, time and location through the map, and then make appointments.

Arrange Appointment Time - Doctors could arrange appointment time as their pleasure in the system.

Get Appointment Feedback - Customers could get the feedback of whether they have made a successful reservation, while doctors could get to know whether a period of time have been checked.

System Updates and Maintenance - The system has to store users' information including account number, password, appointment time and location, and update on time.

Image Processing and Classification - The system could process the uploaded images and give a feedback whether it is benign or malignant.

5.2 Activity Diagram

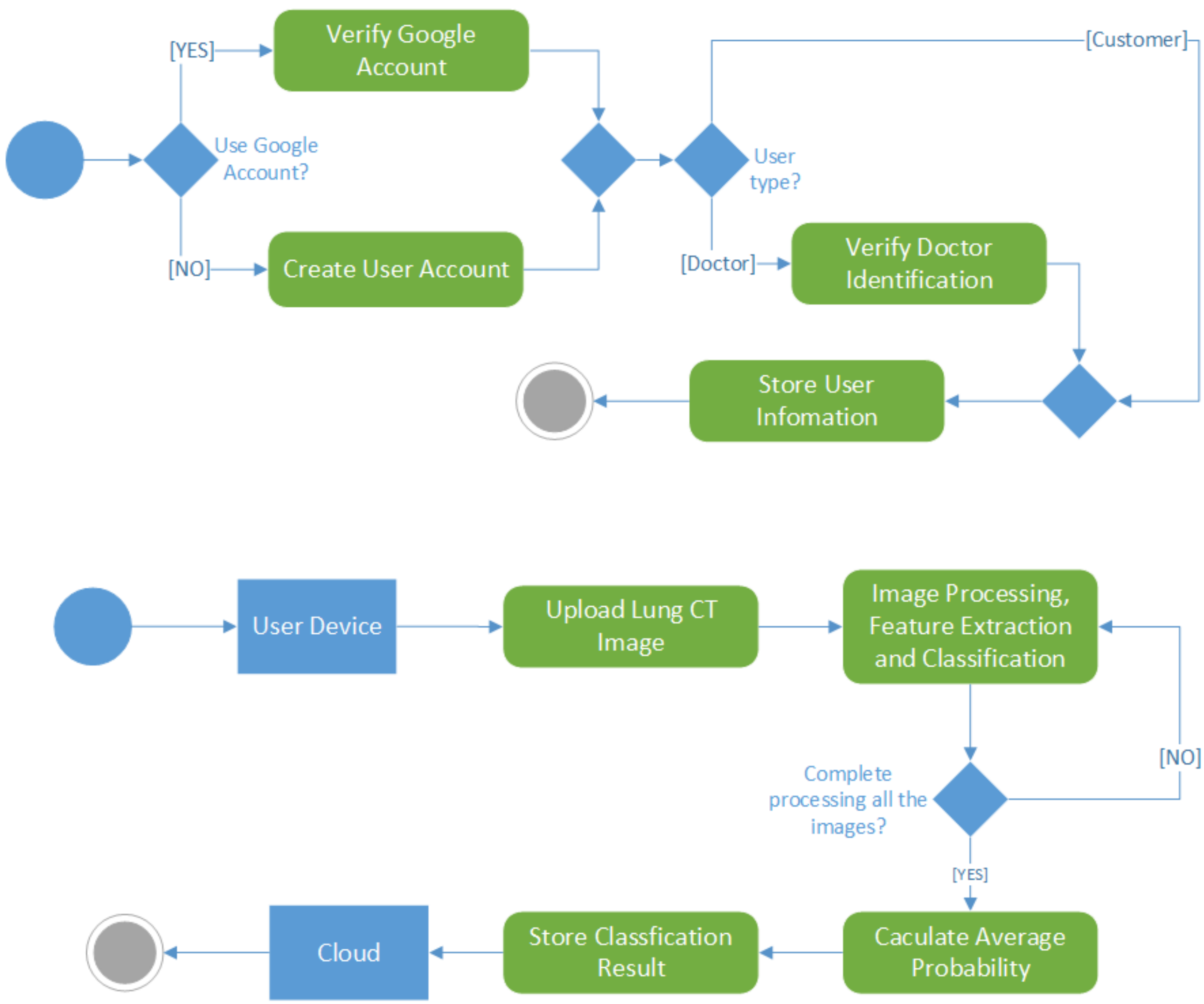

Figure 4 - User Login & Automatically Detection Activity Diagram

User Login (Top)

- The new user attempts to create a new account by selecting whether they want to use and verify their Google Account or create a completely new one.

- After that, he/she needs to confirm their user type. If user is a doctor, he/she needs to upload some relative certificates to verify the doctor identification.
- User is prompted to agree to store the required information into the application secure database.

Automatically Detection (Buttom)

- In order to use this automatically system, what the user need to do is upload some lung CT images and wait for the judgement.
- The system would process these images one by one and calculate the average value.
- Then the system would store them in the database to make sure customers and their doctors could check it again.

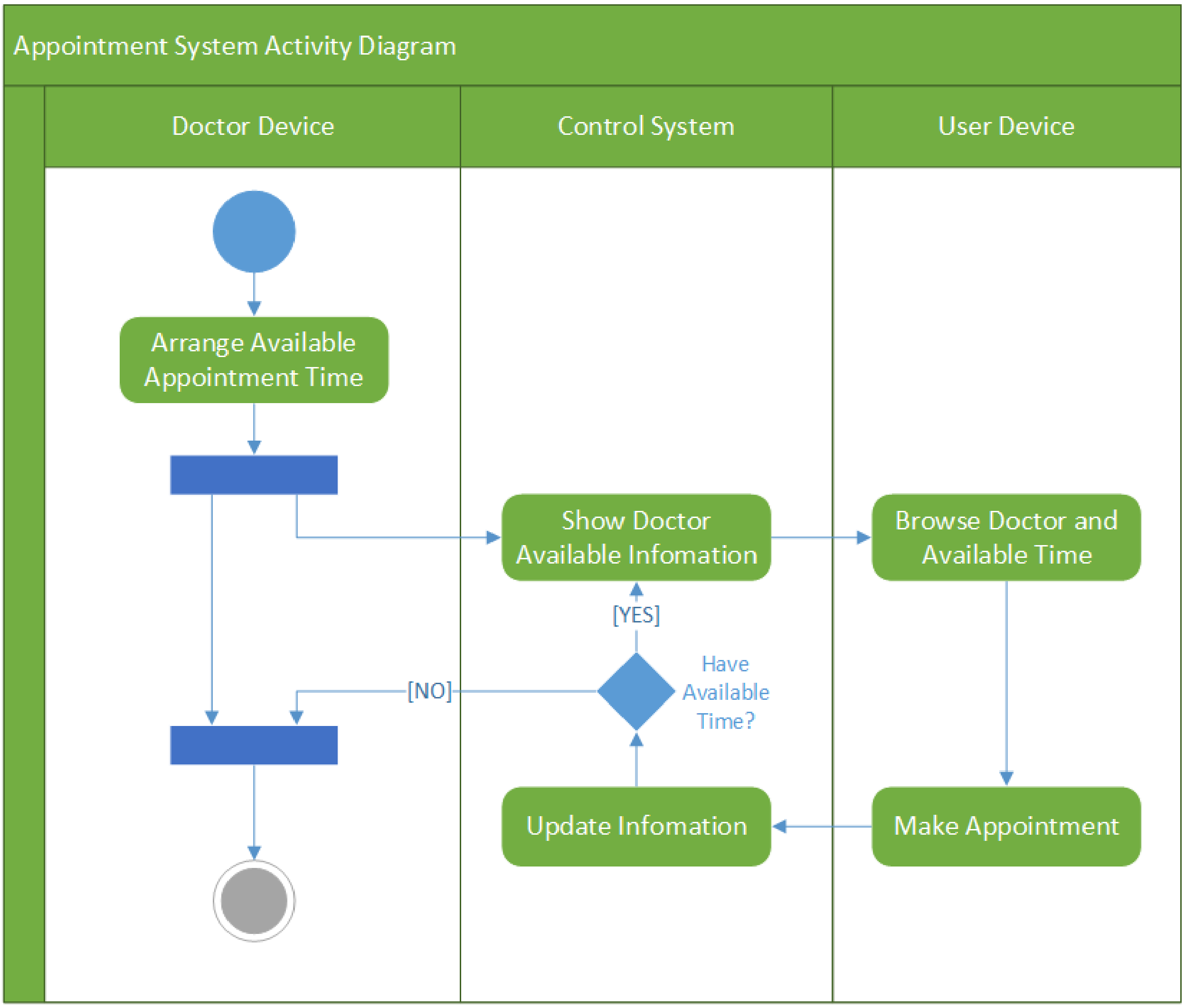

Figure 5 - Making Appointment Activity Diagram

- First, the doctor would arrange the available time for one day.
- Next, the system would store the information and show it to the customers.
- Then, the customers could browse the information and make appointments.
- Finally, the system would update information and check whether there is still have available time for appointment, if the answer is NO, this activity will go to the end.

5.3 Sequence Diagram

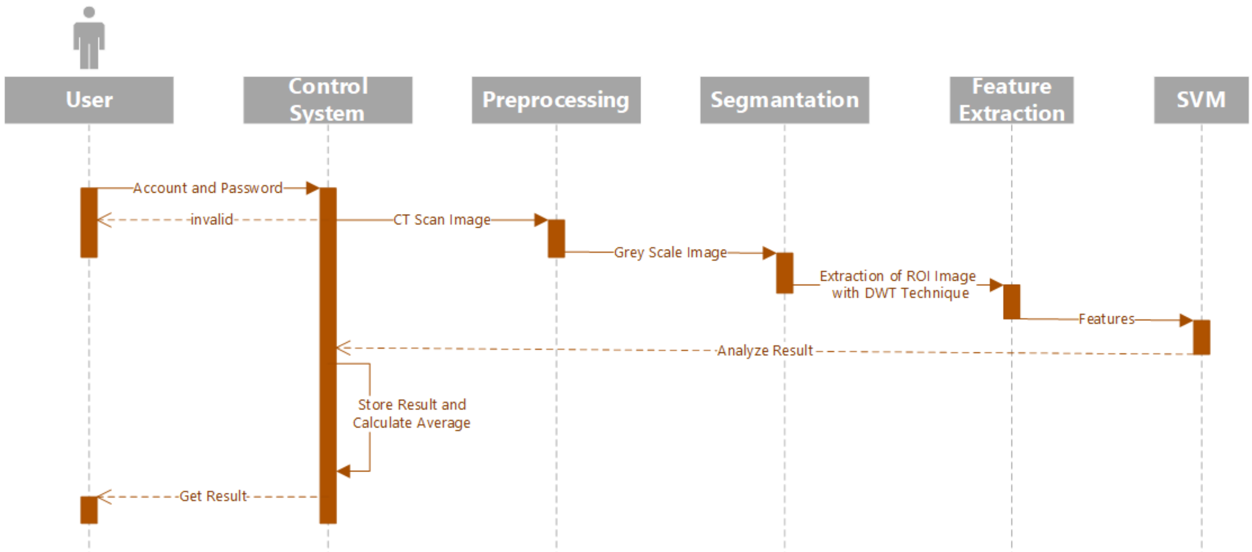

Figure 6 - Automatically Detection Sequence Diagram

The above sequence diagram represents the detailed process of automatically lung cancer detection.

- User need to use their valid account and password to login the system and uploaded one or several images.
- For the first stage of image **preprocessing**, the lung CT images which are in the DICOM image format are first converted to a grey scale image so that they can be easily processed using a digital processing technique and the quality of the image is improved by removing irregularities and noise present in the image.
- For **segmentation** stage, the ROI (Region of Interest) is selected after background edge removal to enable segmentation processing. The Discrete Waveform Transform (DWT) technique is used in the ROI image. The multi resolution analysis is carried out by separating low and high level frequency portions of the image using filters.
- For the stage of **feature extraction**, The GLCM (Gray Level Co-occurrence matrix) characterizes the texture of an image by calculating the value of pairs of pixels with specific value and the spatial relationship and extracts statistical measures from this matrix. It considers relation between the reference point and neighboring pixel. The GLCM has been used for the feature extraction. Feature such as entropy, correlation, energy, variance, contrast and dissimilarities are extracted from the image obtained by means of the 2-level DWT technique for the purpose of the classification.
- **Classification** is carried out using the SVM (Support Vector Machine), classifying whether the image is normal or tumor. For this algorithm, we plot data items in n dimensional space where n is the number of features with the value of the feature being equal to the value of the coordinate and then we perform classification by finding the hyper plane which can be used for classification and regression.

5.4 Class Diagram

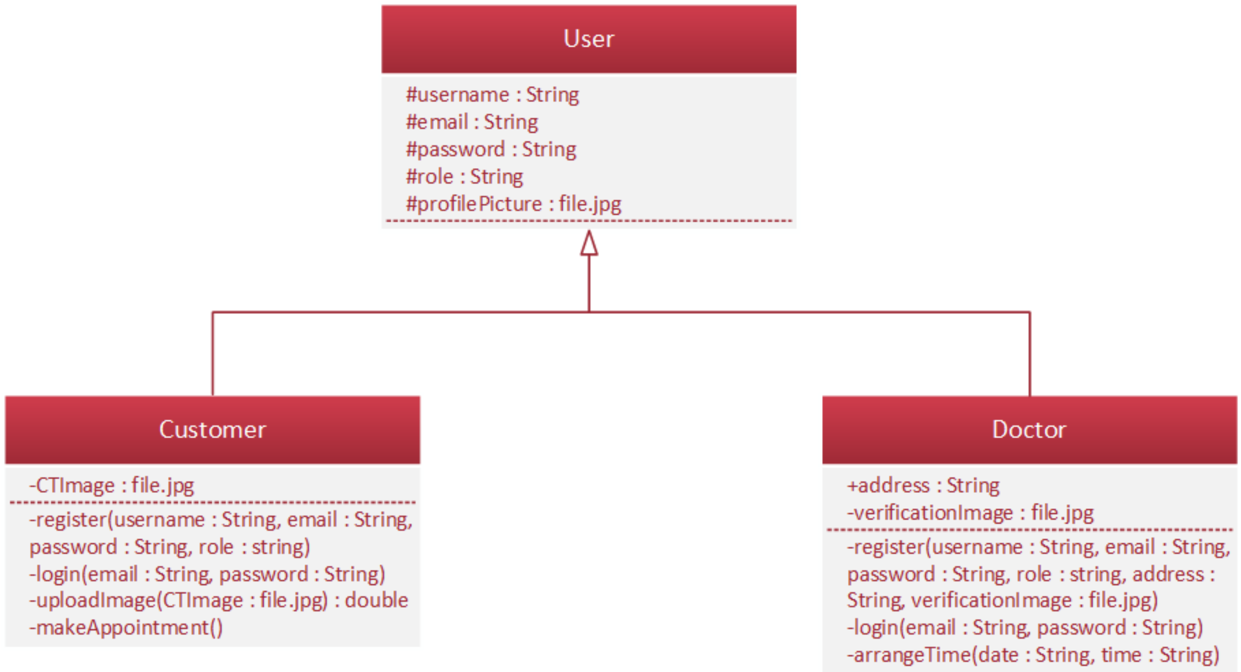

Figure 7 -Class Diagram

Since the two actors in this system have some same attributes like username, email, password and profile pictures, the above class diagram is representation of customers and doctors which extend the common class user.

5.5 State Diagram

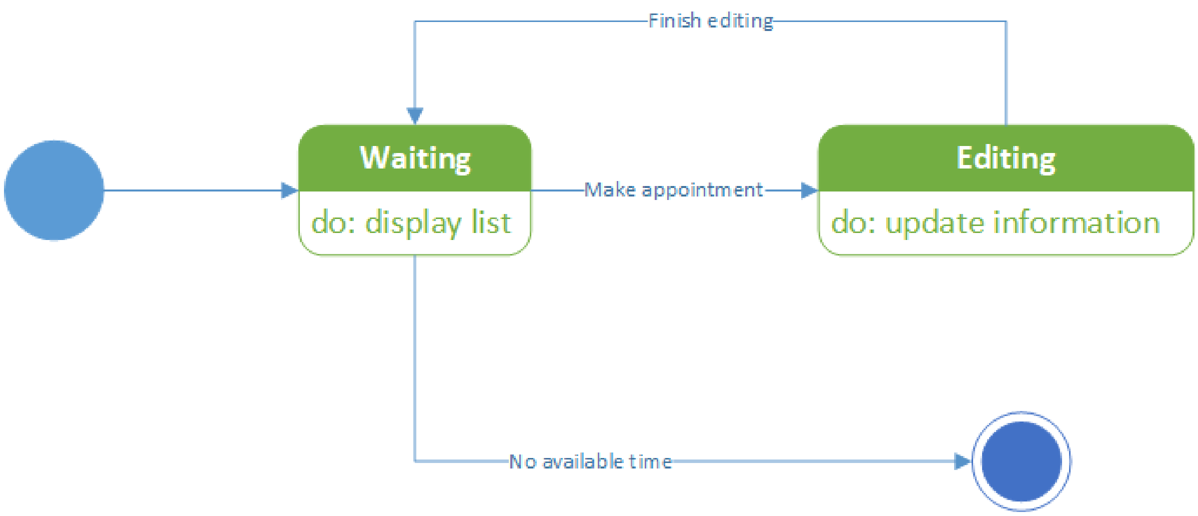

Figure 8 - Making Appointment State Diagram

The above state diagram represents the state of making appointment process.

- In the waiting state, the system will keep display the appointment list including the doctor's name, location and available time. If there are no available time in the list, it will end the serve.
- In the editing state, the system will update information according to the customers' choice and then transfer to the waiting state again.

6. SYSTEM ARCHITECTURE

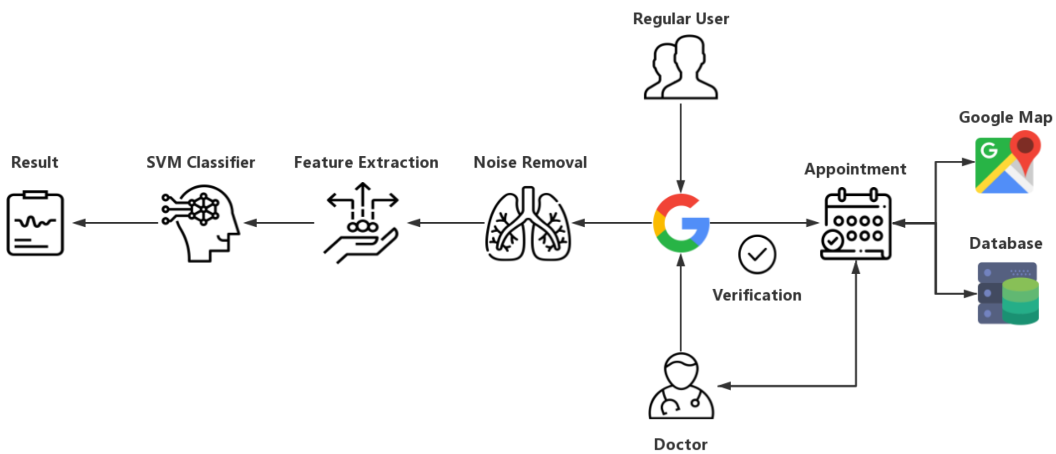

Figure 9. System Architecture

This system has two main functional parts, Lung Cancer Detection, and appointment scheduling. As seen above, all our users are free to access the lung cancer detection part. They shall upload their CT images(dcm format) to this system. Then, the system shall follow those four main steps to preprocess the upload images. After getting the result report, user could arrange an appointment with their most nearby or favorite doctor (if this doctor is the user of our system). Therefore, we would take the action of verifying doctor identity before they want to sign in this system. As long as having an effective supportive document, he/she could sign in this system as a doctor and arrange an appointment with a prospective patient, no matter where he/she is from, personal clinic or hospital. Of course, before using appointment scheduling, doctor need to update their location information. For the lung cancer detection part, here is the more specific version of design diagram.

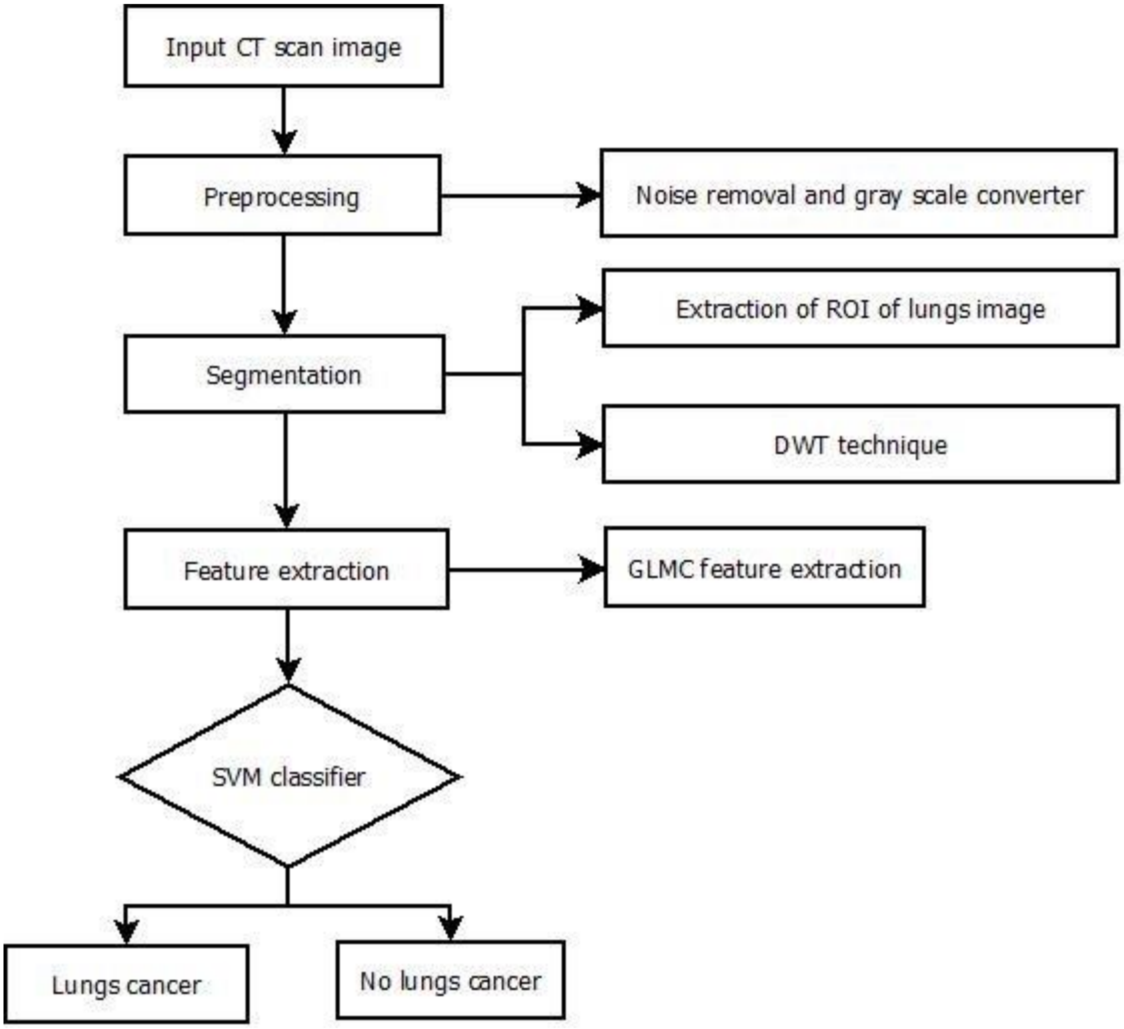

Figure 10. Lung Cancer Detection Design Diagram

This system uses a classifier to detect early lung cancer to improve the accuracy and sensitivity. In this part, we are going to detect early signs of lung cancer. At the same time, we don't ensure to measure the size of nodule and the stages of lung cancer.

6.1 LIDC-IDRI Dataset

This dataset contains standardized DICOM representation of the annotations and characterizations collected by the LIDC/IDRI initiative, originally stored in XML and available in the TCIA LIDC-IDRI collection. Only the nodules that were deemed to be greater or equal to 3 mm in the largest planar dimensions have been annotated and characterized by the expert radiologists performing the annotations. Only those nodules are included in the present dataset.

6.2 Image Preprocessing

The dataset we selected contains lung CT scan images which are in the "DICOM" image format. This first step is to convert the original image into a grey scale so that they can be easily processed by a digital processing technique. The two related tools are **Median Filter** and **Wiener Filter**. The core function of median filter is cv2.medianBlur(). The core function of wiener filter is make_blurred().

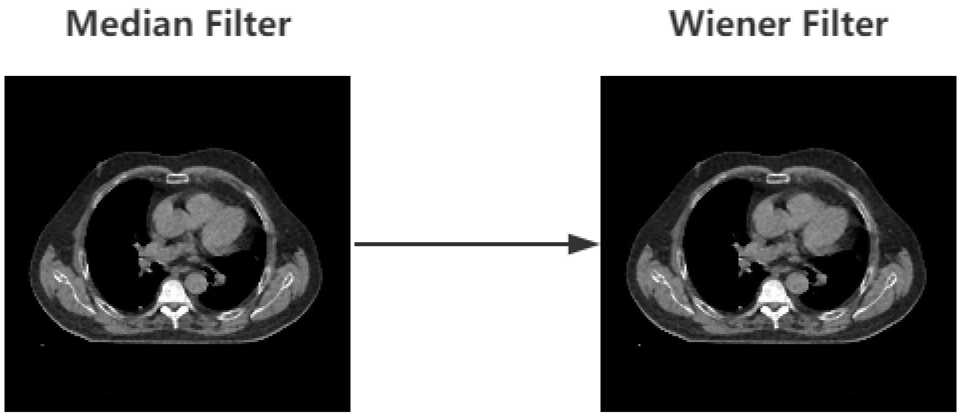

Figure 11. Two Steps of Image Preprocessing

6.3 Segmentation of Image

Segmentation involves detecting the edge and applying the threshold to the image. This part includes three steps, edge detection, thresholding, and applying DWT-2. The ROI based image is first decomposed into four sub bands LL, LH, HL and HH in the 1-level DWT technique. Further, the LL sub band is again decomposed into four additional sub bands using the 2-level DWT technique. The LL component has maximum information content and the other band contain edge in vertical, horizontal and diagonal directions.

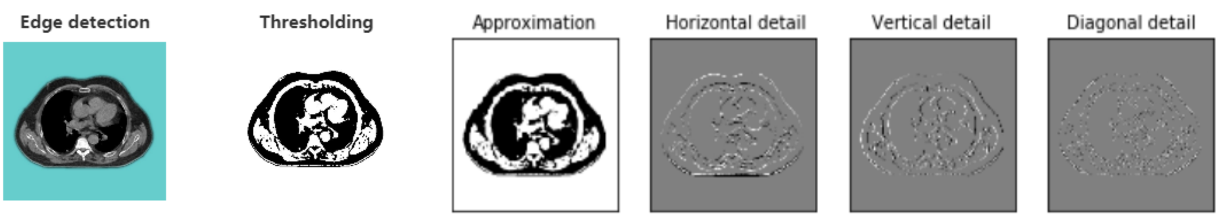

Figure 12. Several Steps of Segmentation

6.4 Feature Extraction

The GLCM (Gray Level Co-occurrence matrix) has been used for the feature extraction. Feature such as entropy, correlation, energy, variance, contrast and dissimilarities are extracted from the sampled image obtained by means of the 2-level DWT technique for the purpose of the classification.

6.5 SVM Classifier

In the final part, we use SVM (Support Vector Machine) to classify whether the image is normal or tumor. We plot data items in n dimensional space where n is the number of features with the value of the feature being equal to the value of the coordinate and then we perform classification by finding the hyper plane.