Informatics Institute of Technology
In Collaboration with
University Of Westminster

Fibro-QuanNet

Pulmonary Fibrosis Prognosis Prediction using Quantum Machine Learning

Project Specification & Prototype Design documentation by

Mr Nuwin Hansitha Godakanda Arachchi

Supervised by

Mr Guhanathan Poravi

February 2023

Submitted in partial fulfilment of the requirements for the
BEng (Hons) in Software Engineering degree at the University of Westminster, UK

ABSTRACT

Pulmonary fibrosis is a progressive lung condition caused by damaged or scarred lung tissue obstructing the exchange of carbon dioxide and oxygen gasses in the alveoli Thereby, leaving the body deprived of the oxygen required for blood oxygenation and less lung volume. As per state-of-the-art medical practice, the deterioration/ scarring of the lung tissue is not entirely reversible or correctable, merely leaving patients with symptom management using therapy and clinical drug trials. An accurate judgment of the lung function decline is crucial for the management and trial treatment of the patient.

This research project endeavors to automate the prognosis prediction for fibrosis using a novel quantum machine-learning approach and re-constructing prognosis prediction to reach the quantum advantage. Although quantum computing is still in its formative years, research activities done in similar domains have proved to have immaculate success in both the correctness and speed of the results. The project explores the advantages one might gain by utilizing the developing quantum computing over the use of classical computational approaches, which will in return facilitate and encourage more optimization of machine learning using quantum computing.

Initial implementation of the core of the system has shown promising results at no optimization of the quantum mode with an accuracy level of 94.793, against highly optimized classical machine learning models (95.21), along with a Laplace loglikelihood of almost identical likelihood graphs, which in medical terms illustrates accurate diagnosis and predictions.

Keywords: Pulmonary Fibrosis, Prognosis Prediction, Quantum Computing, Quantum Machine Learning,

Subject Descriptors:

Computing methodologies → Machine learning → Machine learning approaches → Neural networks

Theory of computation → Models of computation → Quantum computation theory → Quantum complexity theory

Applied computing → Life and medical sciences → Consumer health

Human-centered computing → Human computer interaction (HCI) → Interaction paradigms → Graphical user interfaces

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Acronym	Description
AI	Artificial Intelligence
FVC	Forced Volume Capacity
GUI	Graphical User Interface
HRCT	High-Resolution Computed Tomography
IPF	Idiopathic Pulmonary Fibrosis
MONAI	Medical Open Network for Artificial Intelligence
MRI	Magnetic Resonance Imaging
PF	Pulmonary Fibrosis
QC	Quantum Computing
QML	Quantum Machine Learning
ROCc	Receiver Operating Characteristic curve

CHAPTER 1: INTRODUCTION

1.1 Prolegomena

This document illustrates the authors' attempt at gathering requirements for a pulmonary fibrosis prognosis prediction system while further investigating the stakeholders, different elicitation apparatuses, and use cases. The document further depicts the design concepts adhered to, along with the technical approach taken towards implementing the prediction system, and finally presents preliminary results of the model along with the work planning for the project conclusion.

This chapter establishes the problem and the research gap the author wishes to solve, along with the novelty and contributions made to the bodies of knowledge. Thereupon aims and objectives are defined alongside the challenges faced when developing the solution.

1.2 Problem Domain

1.2.1 Pulmonary Fibrosis (PF)

Pulmonary fibrosis (PF) is a progressive lung disease caused by damaged or scarred lung tissue, occasionally prefixed as idiopathic PF (IPF), when of unknown causality (Devaraj, 2014). The scarred/ damaged area will fibrous the pulmonic tissue, obstructing the exchange of carbon dioxide and oxygen gasses in the alveoli (tiny air-sacks at the end of the airflow branches (U.S. Department of Health and Human Services, 2011)), thereby leaving the body deprived of the oxygen required for blood oxygenation and less lung volume (American Lung Association, 2022).

Consequent to the significant loss of lung real estate and capacity, the amount of air the patient may respire declines considerably, thus, encountering persistent dyspnea (shortness of breath) (Pulmonary Fibrosis Foundation, 2022). As per state-of-the-art medical practice, the deterioration/ scarring of the lung tissue is not entirely reversible or correctable, merely leaving patients with symptom management using therapy and clinical drug trials (Mayo Foundation for Medical Education and Research, 2021).

Therefore, PF has received awareness and concern as one of the most common and lethal forms of idiopathic interstitial lung disease, with an associated survival median of just three years

(Das and Chakraborty, 2015). According to the Pulmonary Fibrosis Foundation, in the United States, one in every two-hundred adults' over 70 years of age may be affected by IPF, with 250,000 cases diagnosed and over 50,000 reported yearly (Pulmonary Fibrosis Foundation, 2022). Approximately 40,000 people lose their lives to PF/IPF in the United States alone (Schwartz, 2018).

Early diagnosis of PF plays a crucial part in early treatments to maintain the patient's lung capacity and function within an acceptable and healthy range while attempting to prevent further deterioration (Choi, 2021). The survival medians of mild, moderate, and severe PF categorized by their FVC percentages were 55.6, 38.7, and 27.4 months (about 2 and a half years) respectively (Nathan et al., 2011), indicating the necessity for early diagnosis of the disease.

1.2.2 Prognosis Prediction of Pulmonary Fibrosis

The deterioration of the lung capacity due to PF and IPF is impossible to be determined and may range from minimal or no degeneration over multiple years to immediate decline within a short time, often even weeks. Therefore, an accurate judgment of the lung function decline is crucial for the management and trial treatment of the patient. Prognosis also helps medical professionals determine the best course of treatment for the patient or where treatment may not be possible results will back the professionals' decision to suggest lung transplantation as the ultimate alternative (Kistler et al., 2014).

For conducting manual detection and progress prediction of PF, multiple institutions have published guidelines, including invasive techniques such as pulmonary tissue biopsy; as well as less invasive techniques such as video-assisted thoracoscopy (Richeldi et al., 2014). Similarly, several techniques are also used to access the lung function decline such as spirometry testing, a classical derivative of lung function and FVC (Wuyts et al., 2016). However, spirometry tests do not take into attention the underlying mechanisms nor the progression techniques.

High-Resolution Computed Tomography (HRCT) has thus come to wide-spread use as the most effective method of assessing lung function and had now become monotonous to conduct a HRCT/CT imaging to understand the underlying mechanism of the lung. In order to identify PF, radiologists and other medical professionals may look for radiologic features such as "honeycombing" or "lung architectural distortion" which indicates the presence of PF/IPF, then required to be verified using a biopsy sample of the lung wall tissue (Niknejad, 2022).

1.3 Problem Definition

Given the severity of PF, it is eminent that prognosis prediction of the diagnosis is crucial in both the patients' well-being as well as providing medical professionals with authoritative and reliable information to perform informed and objective decisions. However, for a manual prognosis prediction, the patient will be subjected to constant, expensive and rigorous testing such as spirometry testing/ HRCT imaging. This delays the prognosis prediction at a whole, taking away the competitive advantage of the prediction, thus taking away valuable time medical professionals may have to make profound and life-saving decisions regarding the patient.

As the National Health Service (2022) reports, IPF is as difficult to detect and prognose due to its quite similarity to other pulmonary deceases such as Chronic Obstructive Pulmonary Disease (COPD). Therefore, supporting the need for further and more thorough research into PF, along with computation intervention to quicken the process.

1.3.1 Problem Statement

Pulmonary fibrosis is a progressive pulmonary condition at which state-of-the-art medical practices are defenseless, thus, medical experts are required to yield critical decisions based on the predicted prognosis of lung functionality manually, which is time-consuming and prone to error.

1.4 Aim and Objectives

1.4.1 Research Aim

This research aims to design, develop, and evaluate a novel prediction model which is capable of providing accurate and efficient prognosis predictions of pulmonary fibrosis utilizing High-Resolution Computer Tomography data through quantum machine learning.

Elaborating on the aim, this research project will produce a system that can be used to automate the process of predicting the prospective, imminent deterioration in the pulmonary forced volume capacity of a patient diagnosed with PF/IPF with the use of pulmonary High-Resolution Computed Tomography (HRCT) imagery data. To achieve this, patient baseline pulmonary HRCT imagery data and pulmonary FVC data will be established along with further FVC results acquired from either pre- or post-baseline testing. With this research, the review and usage of present pre-processing techniques for HRCT imagery data in quantum computing, the model construct will be capable of handling raw HRCT imagery feed input and analyzing the characteristics to produce a prognosis prediction output in terms of FVC as per a defined period.

Prioritizing the validation or invalidation of the hypothesis, extensive study, and research along with the development of said components and evaluating its performance will be performed. The system is expected to be able to run locally or in hosted environments for personal or public use respectively. The machine learning models, and their codes will be made available as open source to facilitate future research. A research paper will be published documenting the outcome and the findings of the project.

1.4.2 Research Objectives

To ensure the research questions and aims are addressed and achieved effectively, the following objectives and milestones are expected to be met to ensure the successful completion of the project.

Table 1.1: Research Objectives

Objective	**Description**	**Learning Outcomes**	**Research Questions**
Problem Identification	Carry out the following tasks to identify the problem • **RO1**: Research interested domains and identify a potential problem which may be feasible to solve within the limited time constraints and technologies.	L01	
Literature Survey	Carry out an in-depth review of the following areas, • **RO2**: Analyze and understand fundamental concepts of quantum machine learning and	L01, L04, L05	RQ1, RQ2

	understanding training models and working with them. • **RO3**: Conduct preliminary studies into existing prognosis prediction systems. • **RO4**: Study on Computer Tomography to understand the need and technique of using HRCT in PF prognosis. • **RO5**: Analyzing existing models and identifying their limitations. • **RO6**: Critical review of the literature and elaborate on the research gap and methodologies		
Requirement Analysis	Carry out in-depth user requirement gathering in the following areas, • **RO7**: Understand and gather requirements users may expect from a prognosis prediction system for PF • **RO8**: Get insight and opinion from Pulmonologists, Radiologists, and quantum data scientists to build the system and mitigate any legal/ social/ ethical issues, • **RO9**: Identifying the tools and techniques (software requirements) and expected behavior for the system through questionnaires	L01, L02, L03, L06	RQ1, RQ3
Design	Develop the design architecture of the proposed system, capable of solving the gap.	L01	RQ3

	• **RO10**: Design a prognosis prediction system to demonstrate the FVC and HRCT data. • **RO11**: Design a data-preprocessing pipeline for HRCT imagery data feed. • **RO12**: Design the QML prediction model which is capable of producing the prognosis prediction		
Development	Implement a system that's capable of addressing the gap aimed to solve. • **RO13**: Develop a prognosis prediction system that can predict the prognosis of PF efficiently and quickly • **RO14**: Develop the QML model that can use quantum super-states to produce quick and efficient predictions • **RO15**: Develop a pre-processing pipeline for HRCT imagery feed data. • **RO16**: Develop the hyperparameter tuning component that improves the prediction system	L07, L05	RQ3
Testing & Evaluation	Testing and evaluating the prototype • **RO17**: Create a test plan and perform unit, integration and functional testing of the prediction system • **RO18**: Evaluate how efficient quantum models maybe be in comparison to classical models to perform predictions.	L04	RQ3

	• **RO19**: Perform requirement validation against all requirements and evaluate accuracy measures.		
Publish Findings	Produce and publish well-structured papers that will critically evaluate and review the research area. • **RO20**: Publishing evaluation and testing results of the project system. • **RO21**: Making code and models created during the project open-source and publicly available for future work.	L06, L08	

1.5 Novelty of the Research

The authors research novelties in this project can be identified as follows.

1.5.1 Problem Novelty

Pulmonary fibrosis prognosis prediction is an unyielding issue where even state-of-the-art medical capabilities are deemed defenseless. Thus, prognosis prediction is the only mechanism for palliative care and maintaining the fibroid level of the lung cavity (Glotov and Lyakhov, 2021). The author attempts to approach this exigent problem by producing a novel approach at prognosis prediction using quantum computing for the same data.

1.5.2 Solution Novelty

A quantum machine learning model for the prediction of the prognosis in pulmonary fibrosis has never been attempted before (Rachel, 2020). Although quantum computing still in its formative years, research activities done in similar domains have proved to have an immaculate success in both correctness and speed of the results, especially giving to its ability to train algorithms much faster than classical computing (Huang et al., 2021). Thus, it is hypothesized that quantum prognosis prediction models can also be immensely enhanced with the use of quantum computational architecture.

1.6 Research Gap

Based on previous studies concerning pulmonary fibrosis prognosis prediction, the literature implies minimal research activity utilizing quantum computing technologies as a scalable option to drive prognosis prediction for PF using machine learning to far more efficient and accurate outcomes (Rachel, 2020). This project proposes to address this gap by introducing a novel approach to pulmonary fibrosis prognosis prediction using quantum machine learning utilizing data imagery from medical apparatuses.

1.7 Contribution to the Body of Knowledge

The author's contribution to the problem and research domain bodies of knowledge can be summarized as follows:

- **Automated Prognosis Prediction**: Data Engineering + Quantum Machine Learning (QML) + Ensemble models
- **Pulmonary Fibrosis Prognosis**: Prediction automation + Artificial Intelligence (AI) + Data analysis

1.7.1 Contribution to the Problem Domain

This project proposes a novel quantum computing approach to PF prognosis prediction and explores QML algorithms, which have not been attempted or explored before, to facilitate the prognosis prediction of pulmonary fibrosis. Using the models produced through the processing of biomarkers, the project also explores the advantage one might gain by utilizing the developing quantum computing over the use of classical computational approaches, which in return will facilitate and encourage more optimization of machine learning using quantum computing.

1.7.2 Contribution to the Research Domain

Identifies the existing limitations in PF prognosis prediction using medical imaging and biomarkers, often prognosed after manual critical referencing and analysis of biomarkers of the pulmonary cavity of the patient and explores the possibility of automating the process in a more scalable path, in terms of accuracy, rather than using classical computation or manual prognosis.

1.8 Research Challenge

Considering the review of existing work done above, the following can be identified as the most challenging in creating a pulmonary fibrosis prognosis prediction QML model:

- Prediction model selection – Reviewing and selecting suitable quantum machine learning algorithms for predictions and classification.
- Feature identification – Reviewing methodologies available to extract input data parameters from HRCT data.
- Quantum computing access – Exploring available quantum computers and computational tools allowing optimal quantum performance

Usage of quantum-based data – Exploring the usage of quantum-based feature data extracted from HRCT imagery data and fine-tuning them towards PF prognosis prediction.

1.9 Chapter Summary

This chapter provided an overview of the problem the author attempts to solve through this research project. Why PF is such a pressing matter on a global scale has been evidently proven along with the gap the author has identified. The aims and objectives of this project were also discussed parallel to the learning outcomes each of those objectives are targeted to achieve. The contributions this project would make to the have also been discusses in detail, along with the novelty of the project, in both solution and the problem. The challenges faced when attempting this research were covered finally.

CHAPTER 2: SOFTWARE REQUIREMENTS SPECIFICATION

2.1 Chapter Overview

This software requirements specifications chapter focuses on identifying and analyzing all the potential stakeholders of the project and the system developed through it by identifying potential beneficiaries through a rich picture diagram. Further, mechanisms to collect and analyze perspective to produce the context diagram, use cases and requirements have also been discussed.

2.2 Rich Picture Diagram

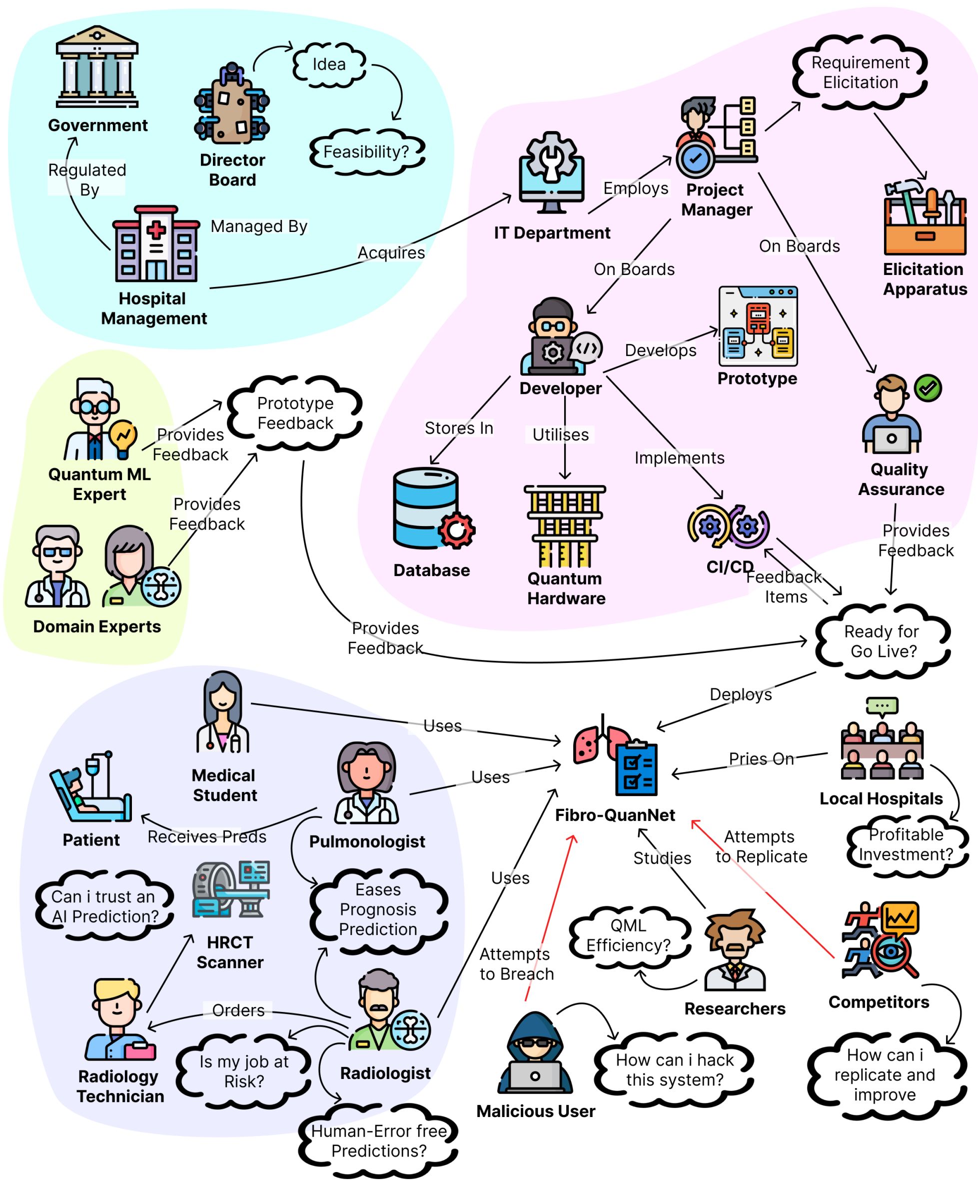

Figure 2.1: Rich Picture Diagram

The rich picture diagram above provides a bird view of all the potential stakeholders/ beneficiaries of Fibro-QuanNet. The diagram outlines most possible implications and how the expected behavior of the system would be of benefit.

2.3 Stakeholder Analysis

The stakeholder analysis has been illustrated using the stakeholder onion model (Figure 2.2) which represents the stakeholders in association with the system, in respective environments followed by each stake holder viewpoint being discussed in the Shareholder Viewpoints.

2.3.1 Stakeholder Onion Model

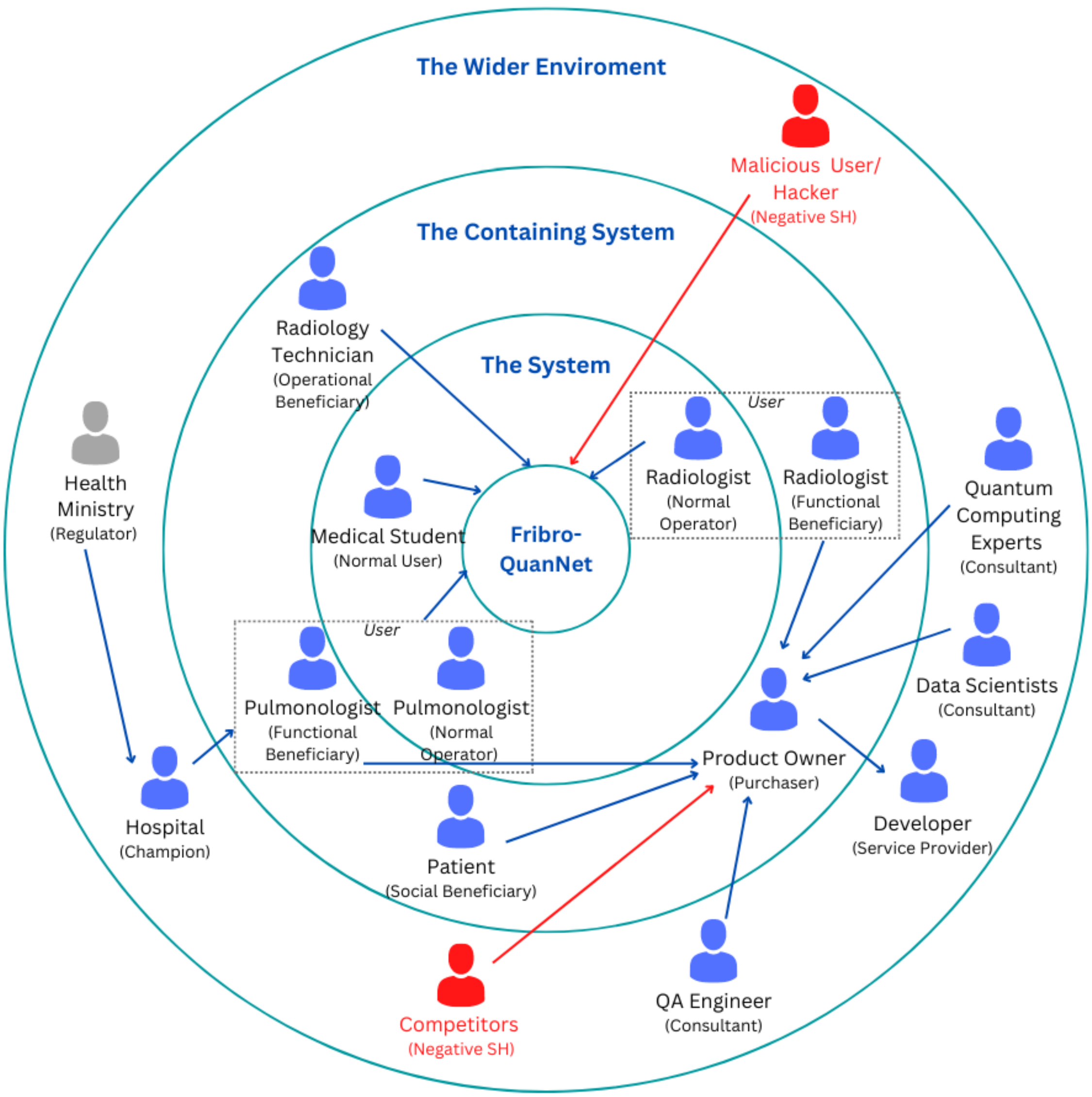

Figure 2.2: Stakeholder Onion Model

2.3.2 Stakeholder Viewpoints

Stake Holder	Role	Benefits/ Role Description
Hospital Admin/ Technology Officer (Product Owner)	Purchaser	Makes sure the application is functional and organizes while gathering and communicating requirements from the functional beneficiaries to the financial beneficiaries and the developer.
Pulmonologist	Functional Beneficiary & Normal User (Hybrid User)	Can use the system to produce more accurate prediction of the progression/ prognosis of pulmonary fibrosis in the patients the pulmonologist gets.
Radiologist		Can use the system to produce accurate reports of the prognosis when referred to by pulmonologists after performing the initial HRCT.
Patient	Social Beneficiary	Will receive accurate prognosis predictions much faster and accurately than usually would manually, which may provide time for proper clinical drug trials.
Medical Students	Normal User	Can use the application to produce accurate prognosis predictions of PF as well as communicate requirements and changes directly to the normal operators received from higher functional beneficiaries.
Radiology Technician	Operational Beneficiary	Will provide the application with necessary data for training models and other requirements of HRCT imagery

Quantum Machine Learning Expert	Consultant	Aids and guides to improve performance of the quantum machine learning models and algorithms.
Pulmonary Disease Expert		Provides advice and guides the developers into the domain to improve the system accuracy and performance.
Dev-ops Engineer		Advice & aids web-app accessibility.
QA Engineer		Advice on quality aspect of the application.
Developer	Hybrid Service Provider	Develops and maintains the system.
Hospital	Product Champion	Reduced opportunity cost for prognosis prediction manually. May attract potential competitive customers due to accurate automated predictions.
Competitor	Negative Stakeholder	May observe and attempt to produce similar systems to manipulate and turn over existing and potential purchasers and product champions.
Hacker	Hostile Agent & Negative Stakeholder	May observe and attempt to breach system security and be hazardous to the system and data integrity of the application laws and conduct.
Health Ministry	Regulator	Will be politically influencing the system to comply all laws and control any loopholes.

Table 2.1: Stakeholder Viewpoints

2.4 Selection of Requirement Elicitation Methods

Gathering requirements from related and significant stakeholders have been using a variety of requirement elicitation tools in this project. This section discusses the tools utilized and the reasoning for usage. Questions and other material used in the preparation of the tools have been attached to the appendix.

Stakeholder Role	**Target Stakeholder**	**Gathering Apparatus**
Negative Stakeholder	Competitors	***Apparatus 1:*** **Literature Review** The author has reviewed and critically evaluated existing applications with capabilities to successfully preform prognosis predictions of pulmonary fibrosis and other similar domains (pulmonary and HRCT based) grounded on the last five years to identify a potential gap in the field. Therefore, using the features studied from the existing system, the author can gather key requirements for the system.
Functional Beneficiary	Pulmonologists, Radiologists & Medical Students	***Apparatus 2:*** **Structured Interviews** As means of collecting requirements from the users of the system, pulmonologists, radiologists, and medical students, structured interviews were performed. Due to the lack of availability of many medical professionals, interviews were deemed the most effective path to collect requirements, additionally giving to its ability to gather qualitative feedback than through text feedback.
Regulator & Consultant	QML Experts, Pulmonology Experts	***Apparatus 3:*** **Structured Interviews** As means of collecting expert-opinion of the system and its models, interviews were conducted with QML and

		Pulmonology experts. Due to the restricted amount of availability of medical experts and the quantum computing domain being new, conducting interviews were deemed efficient.
Developer	Developer	***Apparatus 4:*** **Experimenting/ Prototyping** Prototyping allows the developer the ability to attempt different methodologies progressively and repetitively for the implementation of the system, giving to the agile development life-cycle method the project follow.

Table 2.2: Requirement Elicitation Apparatus & Reasoning

2.5 Discussion of Results

The result-data acquired through the requirement elicitation apparatus versed above have been presented and discussed below.

2.5.1 Findings from Literature Review

Citation	Findings
(Goldin et al., 2019)	Firstly, the authors explore the use of a wrapper method for feature selection over the use of the Filter methods, more popularly used, due to their tendency to misalign the feature selection with the classification steps, resulting in a lesser performance in the classification model. Demonstrates how the use of Evolutionary Computation (EC) to improve the overall performance of wrapper methods. Combines the feature selection prediction algorithms using a wrapper method which combines a quantum particle swarm optimization algorithm to select a few features with random forest for classification.

(Amin et al., 2022)	The authors have identified and utilized a two-stage DP paradigm which includes the use of a CGAN for synthetic data generation, as well as the utilization of both classical and quantum machine learning for the classification stage, using classical convolutional neural networks and quantum Quanvolutional neural networks for the other. The CGAN takes use of three different layers, such as ReLU, for the generative and LeakyReLU for the discriminative networks. For the CNN, ReLU and Softmax activation techniques are used and based on the selected hyper-parameters such as the batch-size, Epoch, the model is trained. The QNN uses 3 layers including quantum-layers, dense layers of specified activation and drop-out layers. The experiments performed on both classical and quantum algorithms, distinguishes the increased performance and accuracy of the CNN and the QNN, and a Confusion Matrix demonstrates 0.96 accuracy in the QML models where the classical algorithm only reached an accuracy of 0.86.
(Sengupta and Srivastava, 2021)	Uses the OpenCV library to de-noice the HRCT imagery as pre-processing techniques. The study also discusses the use of other QML algorithms such as QCNN and hybrid CNN, which utilizes both classical and quantum machine learning to produce a hybrid model in both environments. The authors also use hinge loss and adaptive learning rate optimization (ADAM) over the use of the typical gradient descent-based optimizer giving to its lesser cost computationally and implementation.
(Mandal et al., 2020)	Attempts an ensemble of classical machine learning models as listed under techniques used. Laplace Log technique has been used here for evaluation requirements, which is convenient to measure the accuracy which in return is essential for medical practices to derive a confidence score.

(Wong et al., 2021)	Attempts to utilize GSInquire deep neural networks to validate the performance of the deep neural network using explainability driven performance validation. This ensures transparency in the decision-making process using the clinical imagery data. This also helps identify gaps that exist in classical ML due to the models making "right decisions for wrong reasons". Further, this builds a sense of trust on the system, with the decisions made clearly viewable.

Table 2.3: Findings from Literature Review

2.5.2 Findings from Functional Beneficiary Interview

Pulmonologists, radiologists, and medical students were identified as the core functional beneficiary of Fibro-QuanNet as depicted in the onion model above. Structured interviews were carried out with experts in both the pulmonology and radiology domains to collection opinions and establish the requirements of the system. A consultant pulmonologist from the Sri Lankan Council of Pulmonologists (SLCP), two radiologists from a private hospital, a senior house officer from the Karapitiya National Hospital of SL, and four medical students both pre-intern and MD/ MBBS readers were interviewed and produced below is the thematic analysis of the results and opinions gathered during the interviews. The complete set of questions directed and the expected outcomes of is are attached in the appendix.

Code	**Theme**	**Conclusion**
Biomarkers for Prognosis Prediction	Identification and analysis of available and widely used biomarkers for the prediction of the prognosis in PF.	As imagery data inputs from the user plays a key role in Fibro-QuanNet, this question attempts to recognize biomarkers are widely accepted in the industry. Answers received includes HRCT imagery, MRI imagery and X-Ray imagery, of which all medical professionals agreed, HRCT would be take prominence.

Automating prognosis prediction	Identifies the need for an automated system to predict the prognosis of PF.	As this project deals with a very time sensitive lung condition, which, by the most subtle of detail can make a mass difference in the well-being of the patient, and given the errors a manual system tends to pose, all medical professionals agreed to the need and use of an automated system for PP.
Detection of Fibrosis	Identifies the indications and potential symptoms of PF in the lung cavity through imagery	Training the model to identify the presence of PF is one of the first steps in the process of PP. Thus, information such as honeycombing in the lung cavity, sectional darkening, abnormality in cavities, etc., was collected from pulmonologists of ways to identify and properly the analyze the presence of PF in the lung cavity using HRCT imagery.
Performing prognosis	Identifies the methodologies adhered to when performing prognosis.	Fibro-QuanNet attempts at automating the prognosis prediction of PF, where learning and knowing how to effectively prognose PF obviously plays a crucial role. The medical professionals directed the author with resources to refer, as well as giving a high-level idea and a step-by-step walkthrough of the process.

Table 2.4: Thematic analysis of findings from interviews with functional beneficiaries

2.5.3 Findings from Consultant Interview

Data scientists and quantum computing experts have been identified as the key technology consultants of Fibro-QuanNet system. Structured interviews were carried with classical and quantum machine learning experts in-order to collect their opinions and suggestions regarding the development of the system. A Ph.D. quantum machine learning expert at the Microsoft Quantum Research Center, India; a masters (reading) machine learning expert; and four other classical ML experts have been interviewed to establish a solid foundation to the understanding of the system development. Produced below is the thematic analysis of the results gathered from the interviews. The full set of questions directed is attached in the appendix.

Code	**Theme**	**Conclusion**
Public Quantum Computers	Identifies publicly available quantum computers usable during the project	Since this project deals with utilizing quantum computational capabilities, publicly available quantum computers had to be identified and initialized to get maximum computational capabilities for the development process, as well as for risk mitigation. Systems such as IBM Quantum, Microsoft QML Stimulator, Google Quantum, etc. was suggested by the experts.
Quantum Algorithms implemented	Identifies quantum machine learning algorithms the technical experts have implemented before to get pointers	Since this project revolves around being able to utilize quantum computing capabilities to run quantum machine learning algorithms, the technical experts were asked what algorithms they've implemented to identify and get potential pointers for the implementation of Fibro-QuanNet.
Medical pre-processing libraries	Identified potential libraries that can be used for the processing of biomarkers effectively	The first step of the development process as suggested involves pre-processing the input HRCT imagery. Thus, most experts suggested the use of third-party libraries to ensure valid pre-processing and to effectively manage time within the short time

		constraints. Libraries such as OpenCV, MONAI, etc. were suggested.
Feature extraction/ selection techniques	Identifies possible feature extraction and feature selection techniques for the implementation of the system	Experts suggested, to keep the application as efficient as possible and to make the predictions as accurate as possible, feature extraction and selection techniques would be helpful for fine-tuning. Techniques such as sampling, noise reduction, contour sampling were suggested.
Evaluation Techniques	Identifies possible evaluation techniques for evaluating Fibro-QuanNet	Since evaluation of the model plays a crucial role in the project, possible and the best evaluation techniques were queried from the technical experts for smooth and acceptable evaluation of the models built in the project. Techniques such as AUC, PR curve, error rate mappings, etc., were suggested by the experts.

Table 2.5: Thematic analysis of findings from interviews with the Consultants

2.5.4 Findings from Prototyping

Findings through reiterated prototyping, many challenges and problems were faced, listed under criterion below.

Criteria	**Conclusion**
Quantum model development	During the repetitive building of the quantum models involved in Fibro-QuanNet, problems such as the allocation of qubits and dataset integrations were faced. This had to be addressed by using multiple

	quantum computers from time to time to give interval in utilizing the same resource over.
Quantum computational resource allocation	One of the key issues faced during the prototyping was the availability of quantum resources for the building and run-time of the models. Quantum resources were re-located amid or after idle time of the resources. To address this issue, again, multiple quantum computers were made available during the development process.
Dataset availability	Finding a dataset for this project was a challenging process due to the information the system requires. Even though HRCT images could be collected and analyzed, patient metadata such as the age, genetic history, etc. which are also key factors deciding the prognosis of the patient, had to also be identified and evaluated. To overcome this challenge, the author reached out to other researchers that have worked on classical approaches of the PF prognosis prediction for data collection and try to get datasets they may have created or utilized, where he was able to identify the OSIC pulmonary fibrosis dataset created for a ML competition, which includes all the information required for the project development.

Table 2.6: Criterion of the findings from prototyping

2.6 Summary of Findings

Id	Finding	Literature Review	Interview (FB)	Interview (consultant)	Prototyping
1	PF prognosis prediction would benefit and save lives of PF patients who require urgent predictions for treatments.	✓	✓		

2	Prediction systems may reach a physical boundary for further research in classical ML, making way for alternative pathways.			✓	
3	Achieving the quantum advantage would be of benefit for prediction systems of the prognosis for PF by ramping its speed and efficiency.	✓		✓	✓
4	The gap identified will contribute to both the Medical/ health domain and the quantum computing domain alike.	✓	✓	✓	
5	Building and maintaining custom and use-case specific models for Prediction systems are beneficial and preferred over the use of typical libraries	✓		✓	✓
6	Having a method to effectively and swiftly prediction the prognosis of PF will be a benefit to the patients and users.		✓	✓	
7	Using Quantum Neural Networks/ Quanvolutional Neural Networks over other quantum algorithms will be effective and easier to work within the set time frame	✓		✓	✓
8	Using HRCT imagery data as the input data set for PF prognosis would be deemed superior to using other biomarkers such as MRI etc.	✓	✓		
9	The application should be usable and maintainable due to the fast-improving quantum machine learning field and to incorporate its changes	✓		✓	

10	Having an adaptable and variable prediction model would be preferable to allow for future use of the same models and datasets.	✓		✓	✓
11	Opinions of quantum machine learning experts, pulmonologists and radiologists would be a major influence in the application usage.		✓	✓	
12	Having a sufficient set of well-cleaned & pre-processed data would be vital for the performance of the system.	✓		✓	

Table 2.7: Summary of Findings

2.7 Context Diagram

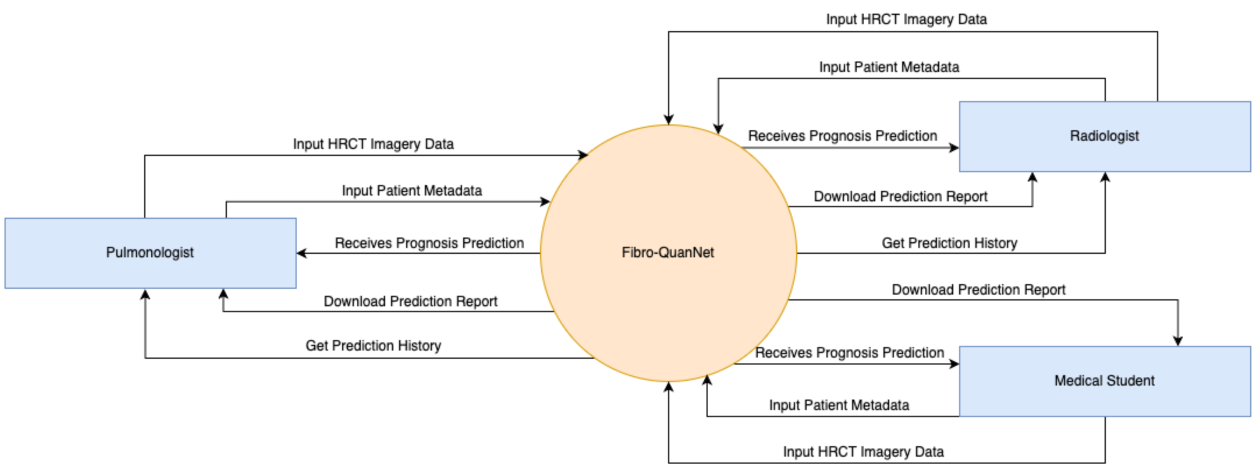

Figure 2.3: Context Diagram

2.8 Use Case Diagram

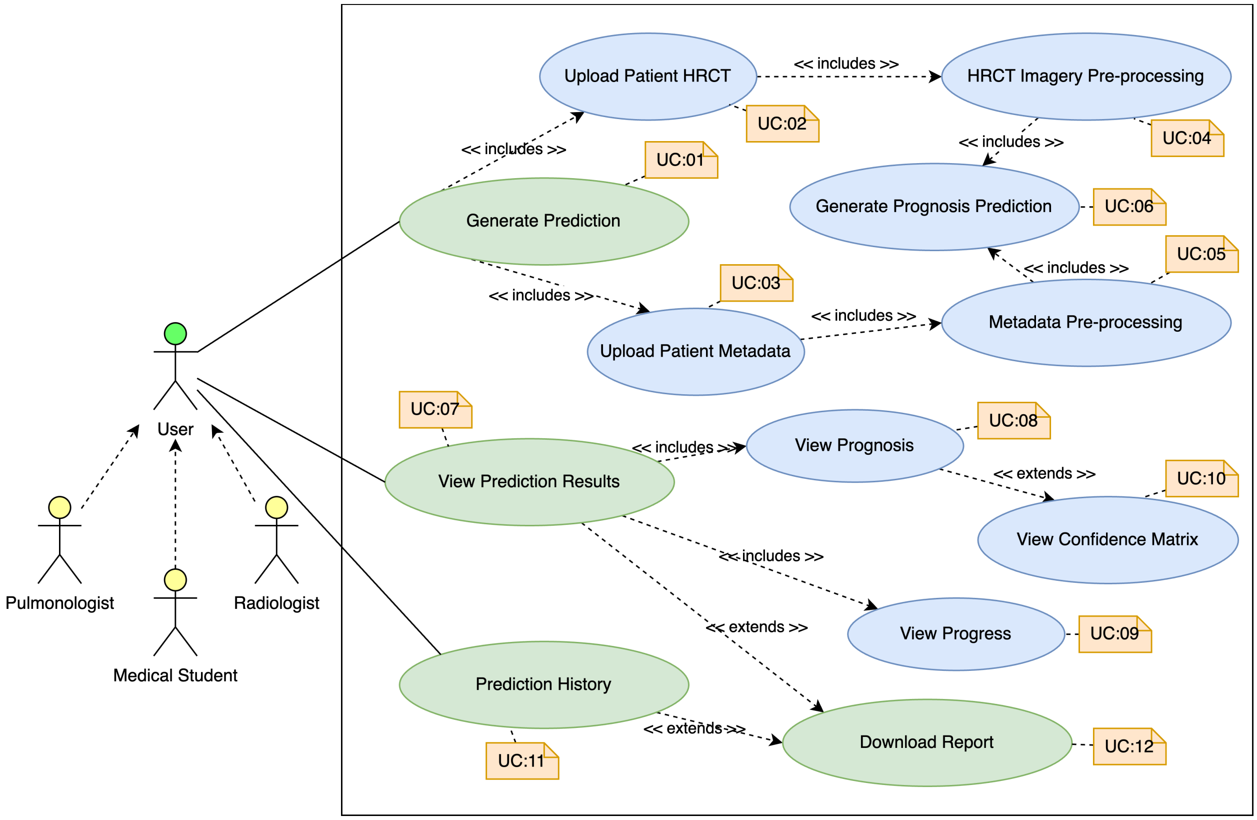

Figure 2.4: Use Case Diagram

2.9 Use Case Description

Use Case	Generate Prediction
Id	UC:01
Description	This functionality allows the use to enter the HRCT imagery and metadata to the system which triggers the pre-processing and PM.
Primary Actor	User
Supporting Actors	None
Stakeholders	Pulmonologist, Radiologist, Medical Student
Pre-condition	The user needs to have uploaded HRCT (UC:01) and the metadata (UC:03) data to the system

Post Condition	Success End: Results produced with the FVC as expected with the confidence matrix.	
Trigger	Success end of the UC:05	
Main Flow	**Actor**	**System**
	1. User opens the upload wizard 3. 30-100 slides of “. dicom” files selected 6. Inputs and uploads Metadata	2. Prompts to upload HRCT 4. dicom type files uploaded 5. Prompts to upload metadata 7. Starts the pre-processing 8. Quantum Module runs
Alternative Flows	-	
Exception Flow	**Actor**	**System**
	3a. Select less than 30 dicom images 3b. Selects over 100 dicom images	4a./ 5a. uploading failed (network) 7a. models starting failed
Included Use Cases	UC:02, UC:03, UC:04, UC:05, UC:06	

Table 2.8: Use Case Description UC:01

Use Case	View Prediction Results
Id	UC:07
Description	View the prognosis prediction results
Primary Actor	User
Supporting Actors	None
Stakeholders	Pulmonologist, Radiologist, Medical Student
Pre-condition	Successful UC:01 and attaches
Post Condition	Success End: The prediction model successfully concludes the prediction

Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Table 2.9: Use Case Description UC:07

2.10 Requirements

2.10.1 Functional Requirements

The MoSCoW principle has been followed to regulate the priority levels given to each requirement in par with their significance.

Priority	**Description**
Must have **(M)**	The core functionality of Fibro-QuanNet and must be implemented.
Should have **(S)**	These features are not core requirements of the application, however, would add value if produced.
Could have **(C)**	These features are desirable requirements that will be considered luxury features and are not essential for the product.
Will not have **(W)**	These features are excluded from the product implementation.

Table 2.10: Requirements Priority Key Map

FR-ID	Requirement	Use Case	Priority Level
FR01	User must be able to upload HRCT imagery to the system	UC:01	M
FR02	The user must be able to enter prevailing FVC to the system	UC:03	M
FR03	The system must be able to pre-process the HRCT	UC:02	M
FR04	The system should choose and borrow most efficient quantum-processing time from the fleet of available quantum machines	UC:05	S
FR05	The system must be able to detect and predict the prognosis of PF in the input HRCT	UC:05	M
FR06	The system should output the confidence matrix of the prognosis prediction	UC:09	S
FR07	The user should be able to view past prognosis predictions	UC:10	S
FR08	The user should be able to download prognosis prediction report	UC:11	C
FR09	The system should have a user-friendly graphical user interface	UC:06	S
FR10	The system should be able to view the current progression of PF	UC:08	C
FR11	The system could produce, and print reporting based on the prognosis prediction	UC:11	C
FR12	The user should be able to elect to contribute HRCT data	UC:07	C
FR13	The system will not show the reasoning behind the prognosis prediction	NA	W
FR14	The system will not take in any other biomarkers as input	NA	W

Table 2.11: Functional Requirements

2.10.2 Non-Functional Requirements

NFR-ID	Requirement	Description	Priority
NFR01	Quality of Output	The accuracy of the prognosis predictions must be of highest quality possible due to its key-role in patient care. The system must be able to produce a quality attribute along with the PP describing the accuracy and quality of the prediction.	Crucial
NFR02	Performance	The system must be able to produce prediction in a much faster and with higher precision using and achieving the quantum advantage for PF PP.	Crucial
NFR03	Usability	Since the system caters for pulmonologists and radiologists who can be of all ages and abilities, the system must incorporate an easy-to-use user interface, which will abide by all design guidelines for accessibility.	Crucial
NFR04	Scalability	Since QC is still at its early ages, and PF still being researched on heavily for treatments, much more functionalities can be added to the system in the future.	Desirable

Table 2.12: Non-Functional Requirements

2.11 Chapter Summary

Through this chapter, the author has brought forth illustrations and descriptions of all the stakeholders that may interact directly, will be benefited, or will take per-role in any external viewpoints using diagrams such as the rich-picture diagram and the stakeholder onion model. The author also discussed the tools used the in gathering the requirements of the system through various stakeholders identified and described. Lastly, the systems use cases, functional and non-functional requirements were also per tailed in detail.

CHAPTER 3: DESIGN

3.1 Chapter Overview

The design chapter aims to present the architecture of the system that was created based on the requirements gathered through the elicitation methodologies discussed in the SRS chapter. The author has provided detailed explanations for the decisions made during the design process in line with the design goals of the project. These justifications include the selection of the design paradigm, high-level architecture diagram, data flow diagrams, system process activity diagrams, and low-fidelity user interface designs.

3.2 Design Goals

Quality Attribute	**Description**
Correctness	The correctness of the predictions made by Fibro-QuanNet must be of highest accuracy possible by utilizing the latest and all available data. This has been addressed in the architecture by specially allocating a presentation tier for a confidence matrix of the prediction performed.
Performance Efficiency	The architecture of Fibro-QuanNet is built such that not all the models utilize the scarce availability of quantum allocation, reserving it for the more crucial models ensuring resource availability.
Scalability	Since quantum computing is still in its early development stages, scalability will be crucial. The architecture allows scalability at its core by segmenting the model components for separate quantum allocations.
Flexibility	Fibro-QuanNet can be identified as one of the most flexible applications due to its operability on any device or OS smoothly. This is achieved in the architecture by utilizing remote computational devices for the foreseeable future with QC still in its early ages.
Reusability	The architecture of the system must provide compartmentalized, reusable components which will enable the system to be updated and facilitate the addition of new functionality and use cases, as required through the

	adaptability goals which will be in-return benefit both patients and the product champion.

Table 3.1: Design Quality Attributes

3.3 System Architecture Design

The system follows the tiered architecture approach, with three distinguished layers, the presentation, logic, and data tiers giving to its ability to establish a logical separation within the system components. Figure 3.1 below illustrates the three-tier architecture of Fibro-QuanNet.

3.3.1 Architecture Diagram

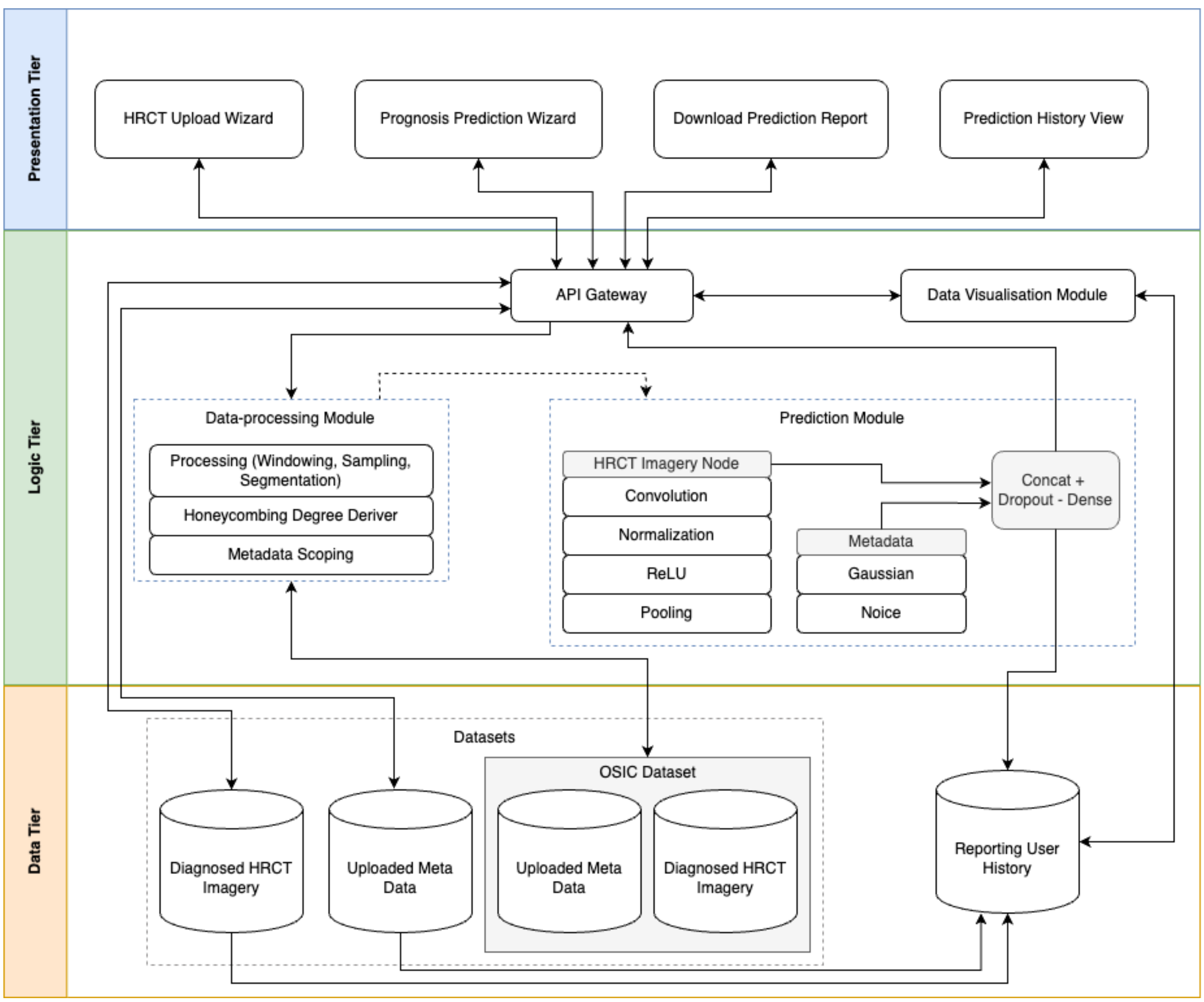

Figure 3.1: High-Level Tiered Architecture Design

3.3.2 Layers of the Architecture

As illustrated above, Fibro-QuanNet follows a three-tier architecture which consists of the presentation layer, the business logic layer, and the data layer. The three layers play a vital role in keeping the architecture modular, and allows further plugins of microservices to the application, facilitating the design quality attributes.

3.3.2.1 Presentation Tier

- *HRCT Upload Wizard* – This UI allows the user to upload the essential input data such as the HRCT inputs and the metadata of the patient such as their age, smoking status, etc.
- *Prognosis Prediction Wizard* – This UI wizard allows the user to view the prognosis prediction output after the uploading data functionality is complete. This will give the user the exact prediction for a set period in FVC, along with a confidence score of the prediction performed.
- *Download Prediction Report* – This UI allows the user to download/ print the predicted prognosis as a medical report, where all the necessary attributes will be included to a reporting structure
- *Prediction History View* – This UI allows the user to view the past prognosis predictions he/ she may have done using the system. The application will allow a user sign-in functionality, where they will be able to track all the cases, they've worked on using Fibro-QuanNet.

3.3.2.2 Logic Tier (Business Logic)

- *Data Pre-processing Module* – The pre-processing model required to pre-process the user inputs before sending them into the prediction module. The preprocessing model is broken down into three main steps:
    - *Processing (Windowing, Sampling, Segmentation)* – This segment is responsible for the initial processing of the data by performing actions such as windowing, sampling and finally segmenting the HRCT sections.
    - *Honeycombing Degree Deriver* – The processed data is scanned for attributes of honeycombing, which is a key indicator for pulmonary fibrosis.
    - *Metadata Scoping* – Input metadata is taken into the scope here and added to prepared for the input to the prediction model.

- *Prediction Module* – The core model of the system which is has been specifically trained to predicting the prognosis of pulmonary fibrosis. The core model also includes several nodes and a concatenation node.
    - *HRCT Imagery Node* – The HRCT node is the node which deals with the pre-processed imagery data. This node involved four other sub-processes:
        - Convolution Filtering
        - Batch Normalization
        - ReLu (rectified linear unit) Activation Function
        - 2D Average Pooling

      Each of these layers are repeatedly performed until the desired size is achieved.
    - *Metadata Node* – The metadata node includes of a Gaussian Noise normalization model which aims to mimic the randomization of the processes in nature.
    - *Concatenation and Drop-out* – Finally, both these branches are concatenated, followed by the final dense and drop-out layers.
- *Data Visualization Module* – This module is responsible for creating visual components of the application such as the report generating and prediction result visualization.
- *API Gateway* – The API gateway acts like the router which exposes the endpoints to the public. API calls will be routed towards each controller to perform the necessary modeling.

3.3.2.3 Data Tier

- *Datasets Database* – This database is responsible of storing and serving the datasets used for the system. The initial dataset of the system as well as HRCT imagery that the user had opted to save in the systems will also be saved in these NoSQL databases. The datasets database can be broken down into two further parts
    - *OSIC Dataset* – This database holds the initial database used for the training of the system.
    - *HRCT & Metadata Input Database* – This database keeps track of all the HRCT and metadata that had been uploaded to the system (given the user had opted for the contributing to the system development)
- *User History* – This database will store the information and the reports generated by users of the system that has opted for keeping their history tracked on the application.

3.4 System Design

3.4.1 Choice of Design Paradigm

After thorough consideration of the requirements for Fibro-QuanNet, the author had seen **Structured System Analysis and Design Method (SSADM)** design paradigm as the preferred design paradigm that aligns with all the requirements of this system over the use of its rival, Object Oriented Analysis and Design (OOAD). Several key decisions that lead to this choice are explored and justified below.

- The research project aims to develop a machine learning based solution which does not require development concerned to object orientation.
- Due to the tight time constraints of the research, it was deemed convenient to develop the application without complex objectification such as object orientation.
- Most quantum programming languages are not object-oriented languages.
- It is also proven that following SSADM assists in future additions to the system.

3.4.2 Design Diagrams

3.4.2.1 Level One Data Flow Diagram

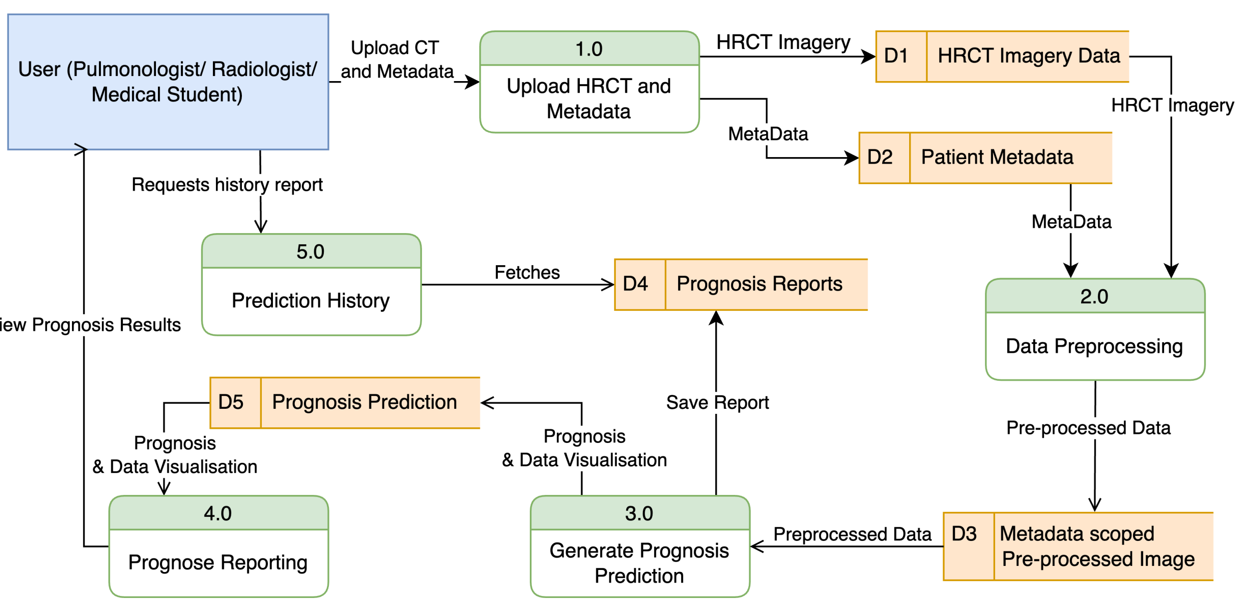

Figure 3.2: Level 1 Data Flow Diagram

The Level 1 data flow diagram, presented as Figure 3.2, illustrates the sequential flow of each action segment that is processed at the system's server end. The diagram is composed of several components that work together to meet the functional requirements of the system. Additionally, the diagram portrays the essential outputs that are to be provided to the user after certain processes have been completed.

3.4.3 Level Two Data Flow Diagrams

Data Preprocessing Level Two Data Flow Diagram

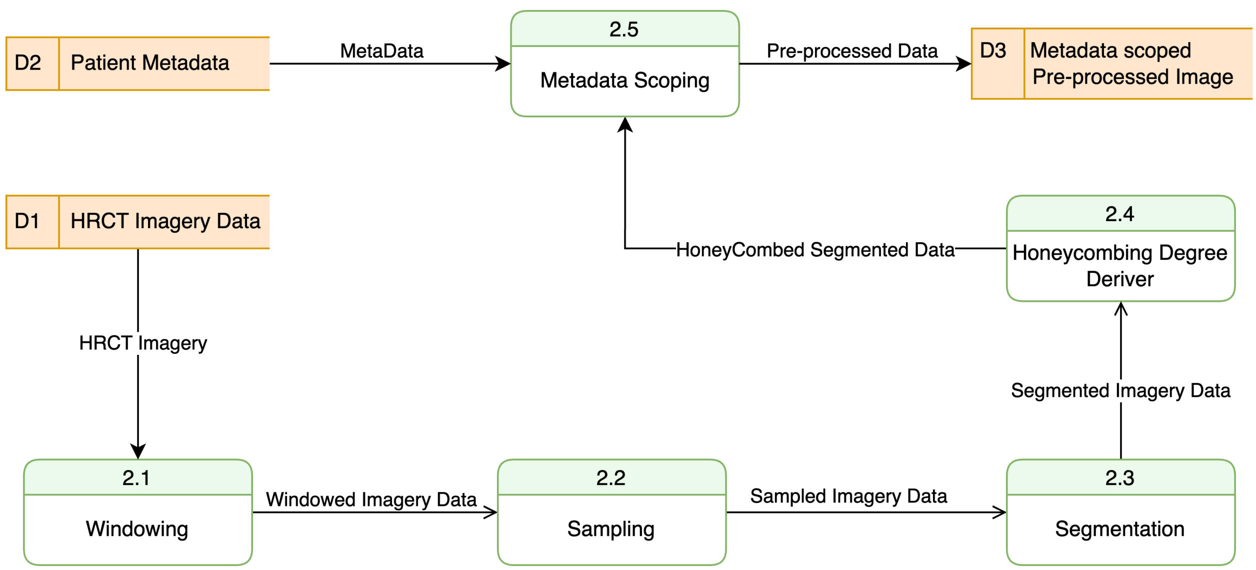

Figure 3.3: Data Preprocessing Level 2 Data Flow Diagram

Generate Prognosis Prediction Level Two Data Flow Diagram

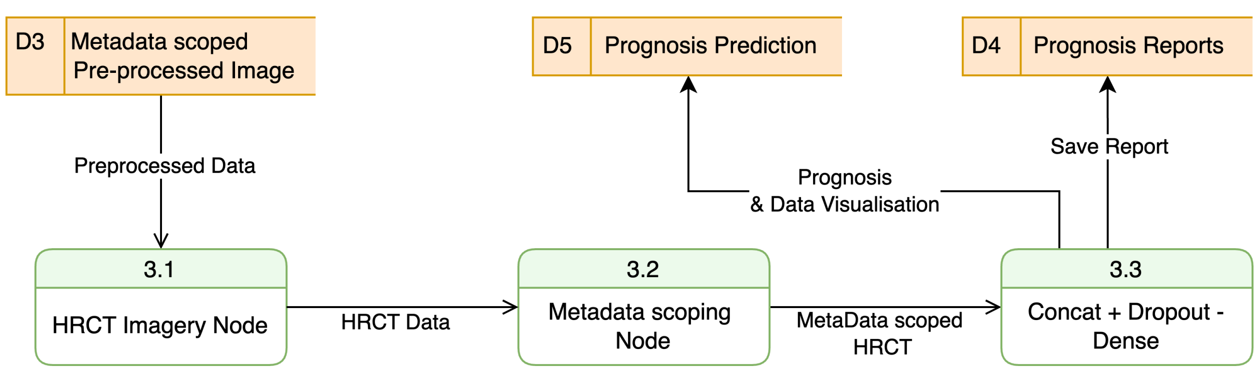

Figure 3.4: Generate Prognosis Prediction Level 2 Data Flow Diagram

3.4.4 System Process Flow Diagram

This diagram is a depiction of the entire system process flow of Fibro-QuanNet with all the system components and validations where necessary.

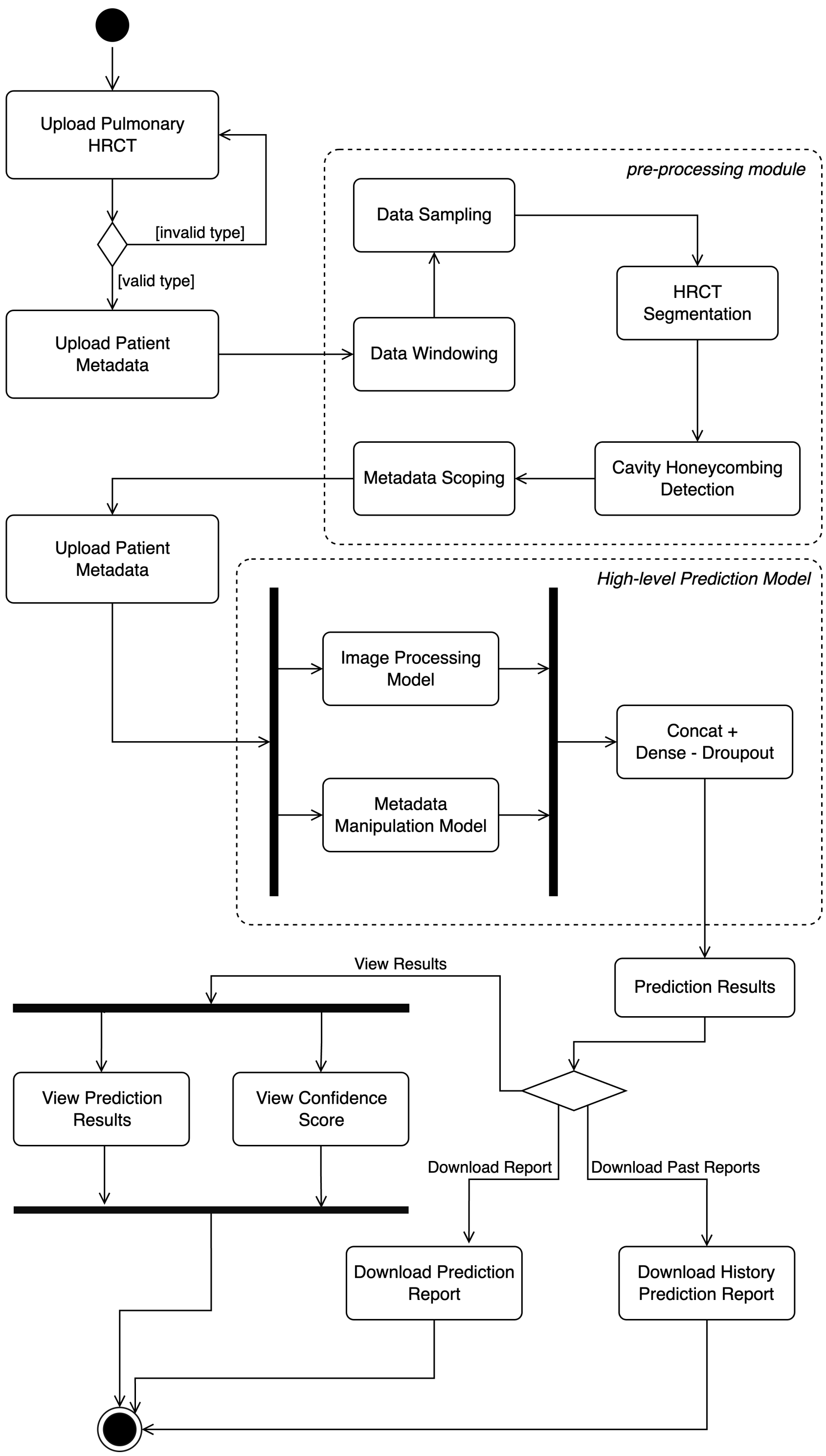

Figure 3.5: System Process Flow Diagram

3.4.5 User Interface Design

As a representation of how the interface will look like for a regular user interacting with the system, a low fidelity prototype is presented below. The complete collection of wireframes can be found in the appendix section.

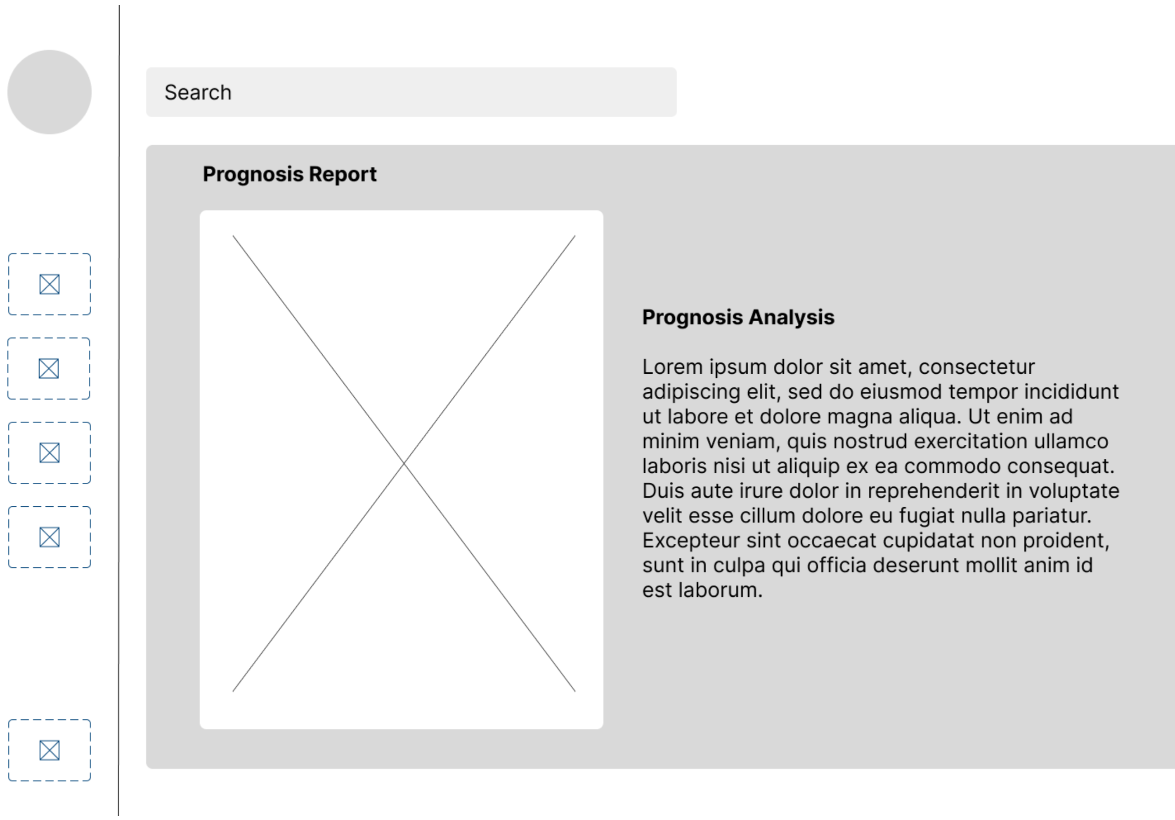

Figure 3.6: Wireframe 1.0 - Dashboard

3.5 Chapter Summary

In this chapter, the design decisions for the system were derived based on the requirements collected during the requirement gathering and elicitation stage. To represent the system, a 3-Tier architecture was chosen, and each of its components was extensively examined. The design paradigm used was explained and supported by reasons, and diagrams were created to illustrate the system flow. The user interface component was then designed according to the stated design goals. Finally, a user experience process was presented to represent the entire system.

CHAPTER 4: INITIAL IMPLEMENTATION

4.1 Chapter Overview

This chapter provides and overview of the initial implementation of the core of Fibro-QuanNet. Decisions made regarding the preferred dataset used, programming language, libraries and other tools used have be justified here.

4.2 Technology Selection

4.2.1 Technology Stack

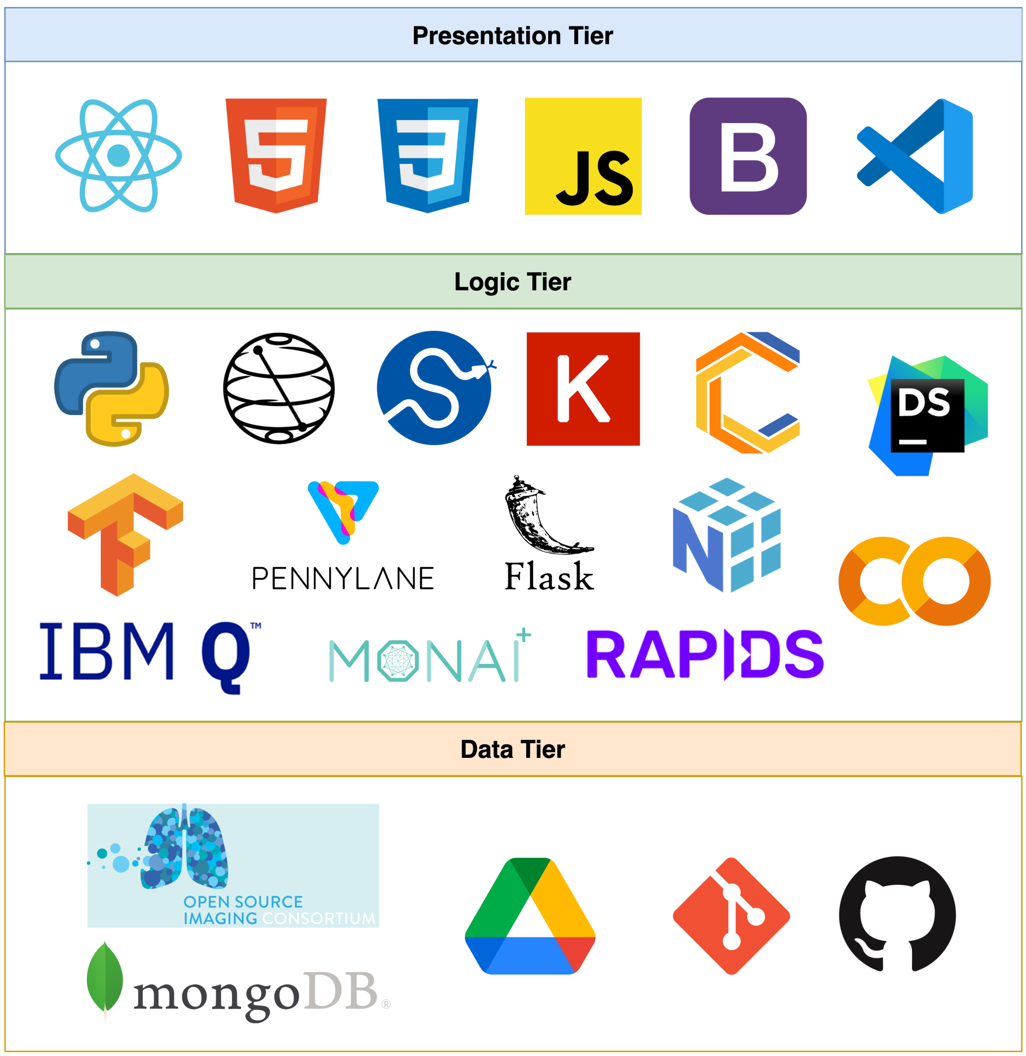

Figure 4.1: Technology Stack

4.2.2 Data-set Selection

The core of Fibro-QuanNet is a data science project, thus requiring a dataset with pulmonary fibrosis prognosis predictions, to train and test the models. Additionally, since this project focuses on medical data, the data must be coming from a valid data point for accuracy. The data required would be:

- Pulmonary cavity HRCT imagery data
- Patient Metadata – Smoking status, age, FVC, etc.

The data requirement was fulfilled using a dataset from the **open-source imaging consortium** which is has made available a pulmonology prognosis prediction dataset which includes data such as the HRCT imagery on baseline, and the patient meta data including the smoking status, age, FVC and lung deterioration percentage.

4.2.3 Development Frameworks

Giving to the application being a python-based application, the author has chosen a python-based development framework to implement the connection to its backend applications. Thus, **Flask** has been selected as the preferred development framework giving to its lightweight and ease in the deployment of the application. Flask also supports the use of different libraries and applications used in developing the system, thus making it ideal for the use of Fibro-QuanNet.

4.2.4 Programming Languages

Since Fibro-QuanNet is a machine learning application as discussed above, **Python** has been selected as the preferred programming language for the development of the models. Python is a general-purpose programming language and can be used for a wide range of tasks beyond machine learning, including web development, scientific computing, and automation.

It is a popular choice for machine learning projects giving to its ease of use, rich libraries such as NumPy, Pandas, Matplotlib, and TensorFlow which make it easier to preprocess and analyze data, build models, and visualize results; and frameworks, versatility, and support for parallel and distributed computing make it a popular choice for machine learning projects.

4.2.5 Libraries Utilized

Library	**Justification**
PennyLane	PennyLane is a robust library capable of which consists of the capacity to cater a wide range of quantum computation engines. PennyLane allows the development and training of quantum machine learning models, and then easily run them on different quantum platforms, and compare results across different quantum systems.
Cirq	Cirq is also a robuts QML library which on the other hand allows the use and practicing of the use of inbuilt quantum machines for creating and running quantum models in googles hardware itself. One of the key reasons to use Cirq is its ease of use, which allows you to define quantum circuits and algorithms using a simple and intuitive syntax.
Quiskit	Quiskit provides a range of tools and libraries for quantum computing tasks, including support for quantum algorithms, quantum circuit design, and quantum machine learning.
MONAI	MONAI (Medical Open Network for AI) is a deep learning framework for medical imaging and related tasks to develop, train, and evaluate deep learning models for a range of tasks, such as image segmentation, classification, and generation. MONAI provides a comprehensive and user-friendly framework for developing and training deep learning models for medical imaging tasks.
RAPIDS	RAPIDS Acceleration of Diagnostics Deployment is a software library that provides GPU acceleration for data science tasks. It is designed to provide high-performance computing capabilities for data processing and analysis, allowing data scientists to work more efficiently and quickly. By leveraging the power of GPUs, RAPID can help to significantly speed up data processing and analysis, allowing data scientists to focus on creating meaningful insights from their data.

Table 4.1: Libraries with justifications

4.2.6 Integrated Development Environments

IDE	Justification
DataSpell	DatasSpell provides a collaborative workspace and user-friendly interface, making it easy for teams to work together on projects and share insights. This can help facilitate better communication and collaboration within your team, leading to more effective and efficient development processes.
Google Collaboratory	Colab provides a valuable platform for people who want to run data science and machine learning workflows without having to worry about managing their own computing resources.
VSCode	VSCode provides a clean and easy-to-use interface that makes it easy for programmers to navigate their code and find the tools needed.
Jupyter	Jupyter is a easy to use application which provides the entire use of a notebook locally, and also eases the use of quantum libraries.

Table 4.2: Integrated Development Environments Selected

4.2.7 Summary of Technology Selection

Component	Technologies Selected
Programming Languages	Python, JavaScript
Development Framework	Flask
Quantum Computing Library	PennyLane with IBM Quantum
UI Framework	React Bootstrap
Libraries	Cirq, Quiskit, MONAI, RAPIDS, NumPy, TensorFlow, Keras
IDE – Research	Google Collaboratory
IDE – Product	VSCode, PyCharm
Version Controlling	GitHub, Git

Table 4.3: Summary of technologies selected

4.3 Implementation of the Core Functionality

4.3.1 Data Pre-processing (HRCT and Metadata)

4.3.1.1 Linear Decay

```
def get_tab(df):
    vector = [(df.Age.values[0] - 30) / 30]

    if df.Sex.values[0] == 'male':
       vector.append(0)
    else:
       vector.append(1)

    if df.SmokingStatus.values[0] == 'Never smoked':
        vector.extend([0,0])
    elif df.SmokingStatus.values[0] == 'Ex-smoker':
        vector.extend([1,1])
    elif df.SmokingStatus.values[0] == 'Currently smokes':
        vector.extend([0,1])
    else:
        vector.extend([1,0])
    return np.array(vector)
```

```
[11]
A = {}
TAB = {}
P = []
for i, p in tqdm(enumerate(train.Patient.unique())):
    sub = train.loc[train.Patient == p, :]
    fvc = sub.FVC.values
    weeks = sub.Weeks.values
    c = np.vstack([weeks, np.ones(len(weeks))]).T
    a, b = np.linalg.lstsq(c, fvc)[0]

    A[p] = a
    TAB[p] = get_tab(sub)
    P.append(p)
```

Figure 4.2: Linear Regression Pre-processing Model

This code segment above performs a linear decay function on the input HRCT imagery and the metadata based on their status of smoking, age of the patient, lung capacity etc.

4.3.1.2 Quantum Preprocessing

```
if PREPROCESS == True:
    q_train_images = []
    print("Quantum pre-processing of train images:")
    for idx, img in enumerate(train_images):
        print("{}/{}        ".format(idx + 1, n_train), end="\r")
        q_train_images.append(quanv(img))
    q_train_images = np.asarray(q_train_images)

    q_test_images = []
    print("\nQuantum pre-processing of test images:")
    for idx, img in enumerate(test_images):
        print("{}/{}        ".format(idx + 1, n_test), end="\r")
        q_test_images.append(quanv(img))
    q_test_images = np.asarray(q_test_images)

    # Save pre-processed images
    np.save(SAVE_PATH + "q_train_images.npy", q_train_images)
    np.save(SAVE_PATH + "q_test_images.npy", q_test_images)


# Load pre-processed images
q_train_images = np.load(SAVE_PATH + "q_train_images.npy")
q_test_images = np.load(SAVE_PATH + "q_test_images.npy")
```

Figure 4.3: Quantum Pre-processing Model

4.3.2 EfficientNet Algorithm

```python
def get_efficientnet(model, shape):
    models_dict = {
        'b0': efn.EfficientNetB0(input_shape=shape,weights=None,include_top=False),
        'b1': efn.EfficientNetB1(input_shape=shape,weights=None,include_top=False),
        'b2': efn.EfficientNetB2(input_shape=shape,weights=None,include_top=False),
        'b3': efn.EfficientNetB3(input_shape=shape,weights=None,include_top=False),
        'b4': efn.EfficientNetB4(input_shape=shape,weights=None,include_top=False),
        'b5': efn.EfficientNetB5(input_shape=shape,weights=None,include_top=False),
        'b6': efn.EfficientNetB6(input_shape=shape,weights=None,include_top=False),
        'b7': efn.EfficientNetB7(input_shape=shape,weights=None,include_top=False)
    }
    return models_dict[model]

def build_model(shape=(512, 512, 1), model_class=None):
    inp = Input(shape=shape)
    base = get_efficientnet(model_class, shape)
    x = base(inp)
    x = GlobalAveragePooling2D()(x)
    inp2 = Input(shape=(4,))
    x2 = tf.keras.layers.GaussianNoise(0.2)(inp2)
    x = Concatenate()([x, x2])
    x = Dropout(0.5)(x)
    x = Dense(1)(x)
    model = Model([inp, inp2] , x)

    weights = [w for w in os.listdir('osic-model-weights') if model_class in w][0]
#     model.load_weights('../input/osic-model-weights/' + weights)
    model.load_weights('effnet-b5-30epochs-1/effnet_30.h5')
    return model

model_classes = ['b5'] #['b0','b1','b2','b3',b4','b5','b6','b7']
models = [build_model(shape=(512, 512, 1), model_class=m) for m in model_classes]
print('Number of models: ' + str(len(models)))
```

Figure 4.4: EfficientNet Algorithm Model

This the core segment of code in the system which utilizes the EfficientNet model to train the model of Fibro-QuanNet. This, however, only partially runs on the quantum space due to restrictions of quantum bits availability.

4.3.3 Multiple Quantile Regression Algorithm

```python
C1, C2 = tf.constant(70, dtype='float32'), tf.constant(1000, dtype="float32")

def score(y_true, y_pred):
    tf.dtypes.cast(y_true, tf.float32)
    tf.dtypes.cast(y_pred, tf.float32)
    sigma = y_pred[:, 2] - y_pred[:, 0]
    fvc_pred = y_pred[:, 1]

    #sigma_clip = sigma + C1
    sigma_clip = tf.maximum(sigma, C1)
    delta = tf.abs(y_true[:, 0] - fvc_pred)
    delta = tf.minimum(delta, C2)
    sq2 = tf.sqrt( tf.dtypes.cast(2, dtype=tf.float32) )
    metric = (delta / sigma_clip)*sq2 + tf.math.log(sigma_clip* sq2)
    return K.mean(metric)

def qloss(y_true, y_pred):
    # Pinball loss for multiple quantiles
    qs = [0.2, 0.50, 0.8]
    q = tf.constant(np.array([qs]), dtype=tf.float32)
    e = y_true - y_pred
    v = tf.maximum(q*e, (q-1)*e)
    return K.mean(v)

def mloss(_lambda):
    def loss(y_true, y_pred):
        return _lambda * qloss(y_true, y_pred) + (1 - _lambda)*score(y_true, y_pred)
    return loss

def make_model(nh):
    z = L.Input((nh,), name="Patient")
    x = L.Dense(100, activation="relu", name="d1")(z)
    x = L.Dense(100, activation="relu", name="d2")(x)
    #x = L.Dense(100, activation="relu", name="d3")(x)
    p1 = L.Dense(3, activation="linear", name="p1")(x)
    p2 = L.Dense(3, activation="relu", name="p2")(x)
    preds = L.Lambda(lambda x: x[0] + tf.cumsum(x[1], axis=1),
                     name="preds")([p1, p2])

    model = M.Model(z, preds, name="CNN")
    #model.compile(loss=qloss, optimizer="adam", metrics=[score])
    model.compile(loss=mloss(0.8), optimizer=tf.keras.optimizers.Adam(lr=0.1, beta_1=0.9, beta_2=0.999, epsilon=None, decay=0.01, amsgrad=False), metrics=[score])
    return model
```

Figure 4.5: Multiple Quantile Regression Algorithm

To minimize the difference between the observed response variable and the anticipated response variable, the quantile regression model is calculated in the conventional implementation of multiple quantile regression using an optimization-based methodology. The choice of technique relies on the needs of the current situation. Iterative algorithms, such as gradient descent or coordinate descent, are frequently used to address optimization problems.

The selection of the algorithm, the decision regarding the quantiles to be estimated, and the decision regarding the predictors must all be carefully considered when implementing multiple quantile regression. It is crucial to thoroughly validate the model to make sure that it is calibrated correctly and makes accurate predictions.

4.4 Chapter Summary

The implementation chapter outlined the tools and technologies that are used by Fibro-QuanNet along with its libraries and frameworks utilized with reasoning for each of these selections. Further, a step-by-step implementation of the codebase has also been attached with its explanations to match each code snippet. The use of both quantum and classical implementations have been notified and clearly defined where necessary and required.

CHAPTER 5: CONCLUSION

5.1 Chapter Overview

This chapter outlines the initial testing results of the prototype developed. The quantum models have shown promising results and are discussed in detail in this chapter, along with the improvements required to develop a successful minimum viable product minimum viable product. Further, the deviations from the scope and the time schedule initially planned for has been reported.

5.2 Deviations

5.2.1 Scope related deviations

In the proposal of this project, it was mentioned that the design paradigm the project would take is OOAD. However, following careful consideration of the system and the project time constrains, the design paradigm has been switched to SSADM with justifications provided in chapter 4.

5.2.2 Schedule related deviations

As per the initial schedule of the implementation, no core deviations have been faced so far with the project components. However, personal targets to finish he MVP level quantum algorithms have not been met so far in regards to the application. Thus, the next couple of months will be dedicated to getting the quantum advantage met, along with having the UI implementation of the system completed.

5.3 Initial Test Results

A statistical method called Laplace Log Likelihood is used to analyze and estimate a model's parameters. It uses a penalty term known as a Laplace prior to account for uncertainty in the parameter estimations and is a type of maximum likelihood estimation.

A variety of models, including logistic regression, generalized linear models, and linear regression, use the Laplace Log Likelihood to estimate their parameters. By encouraging parameter estimates to be tiny and close to zero, this penalty term lessens the effect of outliers and extreme data values.

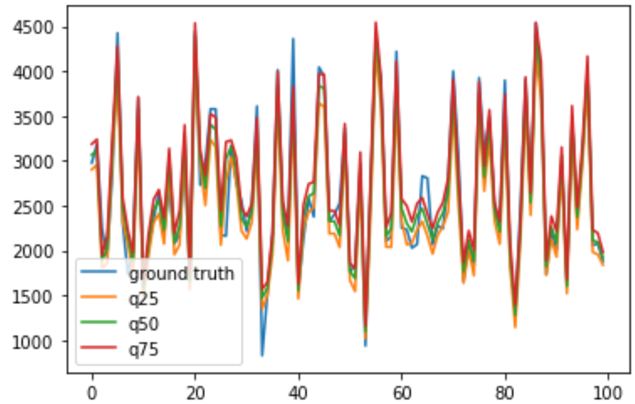

Figure 5.1: Laplace Loglikelihood Result

The penalty term, which is the absolute value of the parameters multiplied by a constant, is added to the log likelihood function to produce the Laplace log likelihood.

5.4 Required Improvements

The machine learning models created in the system are mostly facing inference from the quantum space due to its unprecedented unavailability of quantum bits and allocation for the systems. Thus, the components with the most number of changes have been identified as follows:

- Move and make the models into their quantum processing space with adequate quantum allocation.
- Fine Tune the algorithms to achieve a substantial quantum advantage to overcome and increase the existing accuracy and advantage.
- Develop a User Interface for the system for the normal users.

5.5 Demo of the Prototype

Link to the demo video: https://youtu.be/nwwj0VhgCqw

Link to the code base: https://tinyurl.com/4w85bad9

5.6 Chapter Summary

This chapter covered initial test results of the system using Laplace loglikelihood and the deviations of the system from the originally planned data.

REFERENCES

American Lung Association, N.H.C. (2022). Each Breath: Facts about Pulmonary Fibrosis. *7 Things Everyone Should Know about Pulmonary Fibrosis*. Available from https://www.lung.org/blog/7-things-know-pulmonary-fibrosis [Accessed 6 September 2022].

Amin, J. et al. (2022). Quantum Machine Learning Architecture for COVID-19 Classification Based on Synthetic Data Generation Using Conditional Adversarial Neural Network. *Cognitive Computation*, 14 (5), 1677–1688. Available from https://doi.org/10.1007/s12559-021-09926-6.

Choi, W.-I. (2021). Current and future treatment for idiopathic pulmonary fibrosis. *Journal of the Korean Medical Association*, 64, 256–263. Available from https://doi.org/10.5124/jkma.2021.64.4.256.

Das, S. and Chakraborty, K. (2015). Review paper of Idiopathic Pulmonary Fibrosis. *International Journal of Advanced Research*, 3, 1565–1570.

Devaraj, A. (2014). Imaging: How to recognise Idiopathic Pulmonary Fibrosis. *European Respiratory Review*, 23 (132), 215–219. Available from https://doi.org/10.1183/09059180.00001514.

Glotov, A. and Lyakhov, P. (2021). Pulmonary Fibrosis Progression Prognosis Using Machine Learning. *2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*. May 2021. 0327–0329. Available from https://doi.org/10.1109/USBEREIT51232.2021.9455070.

Goldin, J.G. et al. (2019). Prediction of progression in idiopathic pulmonary fibrosis using CT scans at baseline: A quantum particle swarm optimization - Random forest approach. *Artificial Intelligence in Medicine*, 100, 101709. Available from https://doi.org/10.1016/j.artmed.2019.101709.

Huang, H.-Y. et al. (2021). Power of data in quantum machine learning. *Nature Communications*, 12 (1), 2631. Available from https://doi.org/10.1038/s41467-021-22539-9.

Kistler, K.D. et al. (2014). Lung transplantation in idiopathic pulmonary fibrosis: a systematic review of the literature. *BMC Pulmonary Medicine*, 14 (1), 139. Available from https://doi.org/10.1186/1471-2466-14-139.

Mandal, S. et al. (2020). Prediction Analysis of Idiopathic Pulmonary Fibrosis Progression from OSIC Dataset. *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*. October 2020. 861–865. Available from https://doi.org/10.1109/GUCON48875.2020.9231239.

Mayo Foundation for Medical Education and Research. (2021). Pulmonary fibrosis - Symptoms and causes. *Mayo Clinic*. Available from https://www.mayoclinic.org/diseases-conditions/pulmonary-fibrosis/symptoms-causes/syc-20353690 [Accessed 6 September 2022].

Nathan, S.D. et al. (2011). Long-term Course and Prognosis of Idiopathic Pulmonary Fibrosis in the New Millennium. *Chest*, 140 (1), 221–229. Available from https://doi.org/10.1378/chest.10-2572.

National Health Service. (2022). Idiopathic Pulmonary Fibrosis - Diagnosis. *nhs.uk*. Available from https://www.nhs.uk/conditions/idiopathic-pulmonary-fibrosis/diagnosis/ [Accessed 18 October 2022].

Niknejad, M.T. (2022). Pulmonary fibrosis | Radiology Reference Article. *Radiopedia*. Available from https://radiopaedia.org/articles/pulmonary-fibrosis [Accessed 18 October 2022].

Pulmonary Fibrosis Foundation. (2022). What Is Pulmonary Fibrosis? *Default*. Available from https://www.pulmonaryfibrosis.org/understanding-pff/about-pulmonary-fibrosis/what-is-pulmonary-fibrosis [Accessed 31 August 2022].

Rachel, M. (2020). Detection of Cystic Fibrosis Symptoms Based on X-Ray Images Using Machine Learning- Pilot Study. *Biomedical Journal of Scientific & Technical Research*, 24 (2). Available from https://doi.org/10.26717/BJSTR.2020.24.004019 [Accessed 6 August 2022].

Richeldi, L. et al. (2014). Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis. *New England Journal of Medicine*, 370 (22), 2071–2082. Available from https://doi.org/10.1056/NEJMoa1402584.

Schwartz, D.A. (2018). Idiopathic Pulmonary Fibrosis. *National Organization for Rare Disorders (NORD)*. Available from https://rarediseases.org/rare-diseases/idiopathic-pulmonary-fibrosis/ [Accessed 18 October 2022].

Sengupta, K. and Srivastava, P.R. (2021). Quantum algorithm for quicker clinical prognostic analysis: an application and experimental study using CT scan images of COVID-19 patients. *BMC Medical Informatics and Decision Making*, 21, 227. Available from https://doi.org/10.1186/s12911-021-01588-6.

U.S. Department of Health and Human Services. (2011). Definition of alveoli - NCI Dictionary of Cancer Terms - NCI. Available from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/alveoli [Accessed 6 September 2022].

Wong, A. et al. (2021). Fibrosis-Net: A Tailored Deep Convolutional Neural Network Design for Prediction of Pulmonary Fibrosis Progression From Chest CT Images. *Frontiers in Artificial Intelligence*, 4, 764047. Available from https://doi.org/10.3389/frai.2021.764047.

Wuyts, W.A. et al. (2016). Daily Home Spirometry: A New Milestone in the Field of Pulmonary Fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 194 (8), 1033–1034. Available from https://doi.org/10.1164/rccm.201605-1090LE.

APPENDIX A: REQUIREMENT ELICITATION

A.1: Interview Questions to Consultant Technical Experts

Question	**Relevance**	**Research Question**
Have you worked with cases regarding pulmonary fibrosis?	To identify whether the target audience has worked with the domain disease.	**RQ2**
What medical biomarkers have you used to predict PF?	To identify the potential biomarkers to predict PF	**RQ1**
Which of those biomarkers would you suggest being the best to predict the prognosis of PF?	To identify the requirement for a reliable system	**RQ4**
Have you performed a prognosis prediction of a PF case manually?	To identify whether the target audience has performed the manual prediction of the automated system.	**RQ6**
What features do you look for in the lung cavity for effective PF diagnosis?	To identify the technique used to identify PF and recognize effective ways to do so.	**RQ3**
What features in the lung cavity do you look for when predicting the prognosis of a PF patient?	The identify the technique used to predict the prognosis of patient.	**RQ3**
How crucial do you think the prediction of PF prognosis is to the medical outcome of a patient?	To identify the crucialness of the research project, emphasizing the significance.	**RQ7**

How would you rate the difficulty of performing a prognosis prediction?	To identify the difficulty in predicting manually the prognosis.	**RQ3**
How would you rate the likeliness of a manual prognosis prediction being incorrect?	To identify the requirement for a reliable system	**RQ9**
Do you think automating the process of prognosis prediction of PF would be worthwhile and benefit patients while saving time for medical professionals?	To identify the requirement for a reliable system	**RQ10**
Would you be interested in using an automated tool to predict the prognosis of PF automatically?	Validating the need for the system.	**RQ8**
Have you used or heard about any other applications that can predict the prognosis of PF?	To identify potential competitors	**RQ2**
What features did you like about while using those existing applications?	To identify the difficulty in predicting manually the prognosis.	**RQ4**
What downsides did you face when using these existing applications?	Validating the need for the system.	**RQ6**
What other features would you like to see in an application that can predict the prognosis of pulmonary fibrosis? This may not always be the best option for pulmonary fribrosis preeiction however.	To identify the difficulty in predicting manually the prognosis.	**RQ5**

A.2: Questionnaire to Domain Experts

Question	Relevance	Research Question
Have you worked with quantum computing?	To identify the requirement for a reliable system	**RQ3**
What quantum machines have you used which are open for public use?	Validating the need for the system.	**RQ9**
Have you worked with any quantum machine learning algorithms?	To identify potential competitors	**RQ10**
Have you worked with any quantum image recognition-based machine learning?	To identify the difficulty in predicting manually the prognosis.	**RQ8**
Have you worked with any transfer learning for quantum machine learning?	Validating the need for the system.	**RQ2**
Have you worked on QML for the MedTech industry?	To identify the difficulty in predicting manually the prognosis.	**RQ4**
What quantum ML algorithms/ models for image recognition do you think would help PF prognosis prediction achieve the quantum advantage?	To identify the requirement for a reliable system	**RQ6**
Have you worked with pre-processing models for HRCT medical imagery?	To identify the requirement for a reliable system	**RQ3**

What pre-processing library would you say is best for HRCT/ any medical biomarking?	Validating the need for the system.	**RQ9**
What do you think is the best approach is for developing a prediction model using image recognition?	To identify potential competitors	**RQ10**
What evaluation methodologies have you used in your ML projects?	To identify the difficulty in predicting manually the prognosis.	**RQ8**
Which evaluation method do you think will be best for a prediction model?	To identify the requirement for a reliable system	**RQ3**
Have you developed/ come across any quantum ML models capable of predicting the prognosis of pulmonary fibrosis?	Validating the need for the system.	**RQ9**

APPENDIX B: USE CASES

Use Case	View Prediction Results	
Id	UC:02	
Description	View the prognosis prediction results	
Primary Actor	User	
Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results
Id	UC:03
Description	View the prognosis prediction results

Primary Actor	User	
Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results
Id	UC:04
Description	View the prognosis prediction results
Primary Actor	User
Supporting Actors	None
Stakeholders	Pulmonologist, Radiologist, Medical Student
Pre-condition	Successful UC:01 and attaches

Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results	
Id	UC:05	
Description	View the prognosis prediction results	
Primary Actor	User	
Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view	2. Checks for percentage in quantum model

	5. Selects to download the report	3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results	
Id	UC:06	
Description	View the prognosis prediction results	
Primary Actor	User	
Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction	2. Extracts data from the database

	4. Can download the older report	3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results	
Id	UC:08	
Description	View the prognosis prediction results	
Primary Actor	User	
Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results	
Id	UC:09	
Description	View the prognosis prediction results	
Primary Actor	User	
Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results
Id	UC:10
Description	View the prognosis prediction results
Primary Actor	User

Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results
Id	UC:11
Description	View the prognosis prediction results
Primary Actor	User
Supporting Actors	None
Stakeholders	Pulmonologist, Radiologist, Medical Student
Pre-condition	Successful UC:01 and attaches
Post Condition	Success End: The prediction model successfully concludes the prediction

Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result 4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

Use Case	View Prediction Results	
Id	UC:13	
Description	View the prognosis prediction results	
Primary Actor	User	
Supporting Actors	None	
Stakeholders	Pulmonologist, Radiologist, Medical Student	
Pre-condition	Successful UC:01 and attaches	
Post Condition	Success End: The prediction model successfully concludes the prediction	
Trigger	User Action	
Main Flow	**Actor**	**System**
	1. User selects results view 5. Selects to download the report	2. Checks for percentage in quantum model 3. Output the model result

		4. Output the data visualization
Alternate Flows	**Actor**	**System**
	1. User selects past prediction 4. Can download the older report	2. Extracts data from the database 3. Produces the visualizations
Included Use Cases	UC:08, UC:09, UC:10	

APPENDIX C: DESIGN WIREFRAMES