Informatics Institute of Technology
In Collaboration with
University Of Westminster

Pulmonary Fibrosis Prognosis Prediction using Quantum Machine Learning

Project Proposal by
Mr Nuwin Hansitha Godakanda Arachchi

Supervised by
Mr Guhanathan Poravi

November 2022

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

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

1 PROBLEM

1.1 Prolegomena

This proposal provides a synopsis of pulmonary fibrosis while emphasizing the efficacy of prognosis prediction to help in clinical drug trials and patient symptom management, stressing the research significance. The document further describes the authors' attempt at automating the process of prognosis prediction using a novel quantum machine-learning approach accompanied by critical analysis and discussion of existing related work, indicating the research gap while justifying the research challenges. Research methodologies the author wishes to adhere to, work plan, and deliverables are defined thereupon.

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. 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 accommodation (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 (about twice the seating capacity of Madison Square Garden) lose their lives to PF/IPF in the United States alone.

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 Research Motivation

Pulmonary fibrosis is one of the lethal pulmonary interstitial lung conditions known, claiming thousands of lives, which ultimately leaves patients at an extremely un-comfortable situation having thickening, stiffening and scarring the lung tissue (Schwartz, 2018). Accurate and efficient prediction of the likely prognosis of this disease will be able to give patients more time for clinical treatment trials, which might one day be able to find a potential cure for PF/ IPF.

The manual diagnosis process by itself costs a good fortune for rigorous and constant follow up testing and consumes a longer time span to derive an accurate prognosis using accumulated test FVC results and HRCT imagery. Getting to perform an accurate prognosis prediction using a single HRCT testing would be able to save up millions being spent on follow up, overwhelming testing and unnecessary hospital visits. Pulmonary fibrosis is also close to the author personally, with PF diagnosis cases in close home.

1.5 Related Work

In relation to pulmonary fibrosis, multiple attempts have been made using classical machine learning to attempt to build the most efficient and effective application that provides accurate prognosis predictions to users. As far as computational power drives human, PF prognosis attempts have been made using the following computation technologies:

i. Classical computing – Classical machine learning approach
ii. Quantum computing – Quantum machine learning approach

In this project, the author wishes to attempt the PF prognosis prediction using the growing and experimental quantum computation approach. Therefore, below covered are existing work related to PF and other pulmonary conditions covered.

1.5.1 Classical Machine Learning Approach

Citation	Technique Used	Improvements	Limitations
(Mandal et al., 2020)	Ensemble between Multiple Quantile Regression (MQR), Ridge Regression (RR) and ElasticNet	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.	Even though Mandal and the other authors discuss and prove the use of the ensemble improves accuracy, the speed of the ensemble takes to produce the desired output has not been discussed, which indicates, given the complexity of the models, longer times to

			process the desired outputs.
(Wong et al., 2021)	Fibrosis-Net – Deep Convolutional Neural Network using Computed Tomography imagery data	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.	The authors tends to focus more on the transparency of the deep neural networking rather than on the ML models. The GSInquire does provide a clear idea of the decision making, however the models are taken back. The evaluation method, Laplace Log also shows a minimum increase in the performance score and the confidence matrix.
(Stancioi et al., 2021)	Classical image processing using Hough Algorithm.	Explores the use and performance enhancement using the Hough algorithm. Transformations that could be applied to the processing of the CT imagery data such as the divergent (one-to-many) transformation projection and the	Paper is only a conceptual framework and has never been implemented to actually test the hypothesis in a functional environment rather than through calculations that may in return, not be able to

		convergent (many-to-one) transformation projection are discussed deeply and calculations are made to access the use of both techniques. Proposes an approach to derive a parameter space and weight based on the gradient and curvature of the CT imagery.	achieve the said advantage. Further, the Hough algorithm also has its own limitation, the transformation calculation, which consumes a few cycles of time, slowing down the system.

Table 1.1: Classical Machine Learning Related Work

1.5.2 Quantum Machine Learning Approach

Citation	**Technique Used**	**Improvements**	**Limitations**
(Goldin et al., 2019)	Quantum Particle Swarm Optimization – Random Forest (QPSO-RF) algorithm, as an optimizer for the infamous wrapper method, integrating feature selection and pattern recognition	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	The risk of premature convergence of the QPSO remains despite the authors use of probabilistic cross-over as well as mutation-operators. Further, the model is clear to be limited to a much smaller region of interest (ROI) owing to its optimization toward a single model criterion. The datasets was also

		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	marked and labeled with the use of a single reader, which future reducers the trustworthiness and the reliability of the models and compromises the generalizability of the models. Cross-validation of the datasets, which plays a crucial role in the reliability of models, has not been preformed.
(Amin et al., 2022)	Use of a modified Conditional Generative Adversarial Network (CGAN) and Quanvolutional Neural Networks	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	As the authors have also clearly mentioned, the datasets were biased during the training, due to the datasets being extracted through hospital CT imagery in Pakistan, thus, the QNN model has reached an accuracy on 1.00 in the use of the POF hospital dataset, which is unrealistic, and may take down confidence scores. Further, the models only classify to

		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.	check is Covid-19 is present in the pulmonary cavity, which again does not provide the severity or the survival rates of the patient, which in hand also plays a crucial role in the treatment process, similar to that of pulmonary fibrosis prognosis prediction.
(Sengupta and Srivastava, 2021)	Uses Quanvolutional Neutral Network, aka. Quantum Neural Network (QNN)	Uses the OpenCV library to de-noice the HRCT imagery as pre-processing techniques. The study also discusses the use of other	The authors have collected the data through sources such as google and Keras TensorFlow, which in

		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.	return resulted in a smaller sample size, which did not produce clinical level accuracy results, reducing the confidence scores. The study only curtails the use of Computed Tomography for the detection of patches/ abnormalities in the pulmonary cavity. The study also entails the MVP which may require development & accuracy tuning to bring to a clinical accuracy, which again the authors point to may reduce performance.

Table 1.2: Quantum Machine Learning Related Work

After critical analysis of the existing work and technologies, few of which has been summarized above, the author has been able to understand and identify a clear gap existing in the performance of the applications available. As discussed, PF is a condition that may cause and continue to cause considerable damage to the lung of the patient, and clinical trials are the only existing pathway to manage the fibrosis, which as Devaraj (2014) requires the use of the latest technologies to improve the efficiency and effectiveness of the applications. Moreover, as Amin et al. (2022) has proved by the use of quantum models to produce similar prognosis predictions, the reaching the quantum advantage for the prognosis prediction of PF will be a break-through for future predictions of PF prognosis.

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. 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 Research 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 Problem 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 Research Questions

RQ1: What are the existing technologies used in classical and quantum computing methodologies to predict the prognosis of PF?

RQ2: How much further can prognosis prediction of PF be developed and enhanced using classical computing methodologies till it reaches a physical research periphery?

RQ3: How can prognosis prediction of PF using HRCT imagery data be optimized and enhanced to achieve the quantum advantage?

1.10 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.

1.11 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.

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 maybe 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	

Table 1.3: Research Objectives

1.12 Project Scope

The scope of the project has been derived after considering the research objectives and aims, review of existing work and the limited time extent permitted for this project.

1.12.1 In-Scope

The following list of features shall be the scope of the project:

- **Prognosis Predicting QML model –** Develop a quantum machine learning model that's capable of making predictions of PF prognosis using HRCT imagery data.
- **Testing & evaluating the prediction model –** Benchmarking the QML model against the performance of existing classical machine learning models.
- **Prognosis Predicting Graphical User Interface (GUI) –** Develop a system capable of predicting the users' pulmonary FVC in milliliters (ml) for an amount of time-based on HRCT imagery.

1.12.2 Out-Scope

The following are features that will not be included in the project:

- **Inter-related pulmonary conditions –** This project will be solely focusing on pulmonary fibrosis, and not any other inter-related pulmonary conditions such as Cystic fibrosis.
- **Use of different biomarkers –** This project will only cover the use of Imaging testing for the prediction learning process and will not cover the use of lung function testing or tissue sampling (biopsy) which are also diagnosis pathways.
- **Usage of different imaging techniques –** The primary imaging used in this project will be HRCT imagery as described in the objectives. Other biomarkers such as chest X-Rays or Magnetic Resonance Imaging (MRI) will not be covered in the project.

1.13 Prototype Diagram

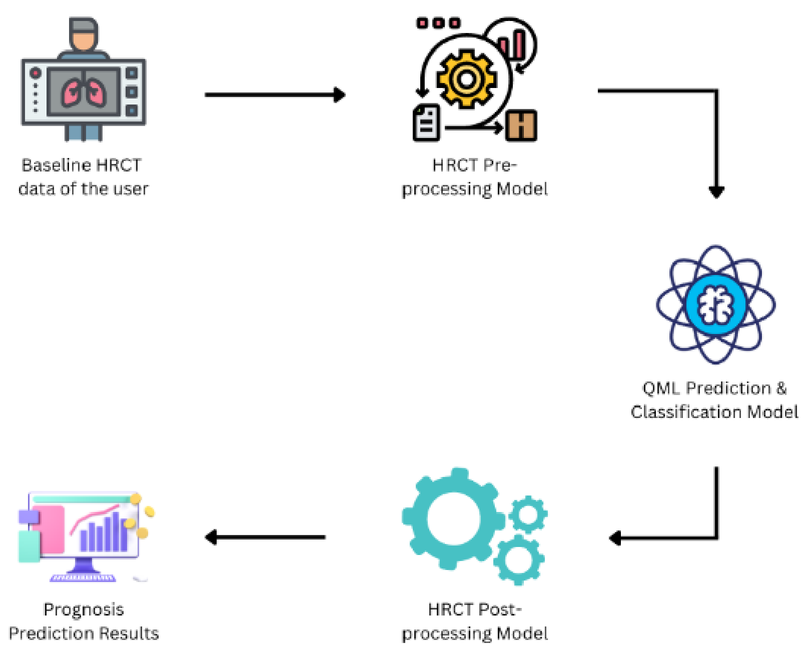

Figure 1.1: Prototype Diagram

2 PROPOSED METHODOLOGY

In order to efficiently manage cost, time and scope, the key factors governing the quality of any project, the Saunders Research Onion Model (Saunders, Lewis and Thornhill, 2007) has been used to construct the appropriate methodologies for this project.

2.1 Research Methodology

Research Philosophy	The research philosophy affects the nature in which data is collected and analyzed due to its relation to the investigation of reality. Among the different philosophical paradigms available and discussed, **Pragmatism** has been selected and used in this research, giving mostly to its involvement of both quantitative and qualitative data.
Research Approach	The research approach discusses the approach taken by the researcher when performing the research. For this research, the **Deductive** approach has been followed over the alternative Inductive approach since the research aims to test and prove the hypothesis and expected to be quantitative.
Research Strategy	The research strategy discusses the data collection strategies adhered to when collecting data to answer the research questions. Among the different strategies available, **Surveys** has been chosen as the primary focus since it shed light to a more quantitative approach. The target audience being in the medical profession, it would be quite beneficial to conduct interviews with them as well.
Research Choice	The research choice focuses on the choice of the methodologies selected and identifies the qualitative and quantitative characteristics of the research. The **Mixed method** has been selected as the research choice for this research due to its qualitative and quantitative character. Both surveys

	as well as interviews conducted with the target audience will be used to defend the hypothesis.
Time Horizons	The time horizon selected for this research will be **Longitudinal** since required data will be gathered and used prior and after the development of the system.

Table 2.1: Research Methodology

2.2 Development Methodology

2.2.1 Life-Cycle Model

The software development life cycle governs the continuous progression of the project from its initiation to the very end. Off the available life cycle methodologies, the **Agile** Software Development life cycle has been chosen for this research as the research development method giving to its adaptability to rapid changes in the requirements of the project.

2.2.2 Requirement Elicitation Methodology

Requirements of the project will be gathered through medical practitioners and PF patients using various methods of data collection such as questionnaires and surveys and will be analyzed to identify the requirements for this project as well as for the betterment of the domain application.

2.2.3 Design Methodology

Object-Oriented Analysis & Design (OOAD) has been selected as the design paradigm for this research due to its alignment and focus on data structures and tendency to relate to real-world objects, this object orientation, which may additionally also enable and help in the future scaling and re-usability of the models and applications built through this project.

2.2.4 Development Methodology

This project uses of **Object-Oriented Programming (OOP)** over its counterparts such as structured and functional programming due to its tendency to orient the application towards object orientation and thus be easier to handle and scale in the future.

2.2.5 Evaluation Methodology

Classification matrix evaluation techniques such as Confusion Matrix, ROCc (Receiver Operating Characteristic curve) and the Laplace Log will be used in this project. The ROC curve and the Laplace Log will play a key-role in producing a confidence score which is a crucial factor in medical predictions.

2.2.6 Solution Methodology

Similar to classical machine learning, quantum machine learning also requires several steps be taken in order for a successful model production. The architecture of the QML prediction model has been planned as follows.

For the data sources, the author will be utilizing the OSCI public dataset which includes a large sum of HRCT imagery for the training and the testing of the model. The application further aims to use the MONAI library to pre-process the HRCT imagery data, following which a QCNN will be utilized as the quantum net after a series of quantum gates, while version controlling using version controlling software discussed below.

Once the machine learning models are functional, CM and ROCcs will be utilized evaluate the performance of the models and an API will be developed to enable communication between the GUI application and the back-end model. The application will be deployed in a CI/CD pipeline to continuously integrate changes to the systems and keep up-to date.

2.3 Project Management Methodology

Adhering to a project timeline, scoping and costs plays a crucial part in any project. Therefore, selecting a suitable management methodology plays a key-role in the project.

Accounting for all opportunities and take-aways from methodologies available, **Prince2** has been selected to make sure the project is product oriented and will be flexible to any fluctuations that may occur and will need to be adjusted accordingly as well as to compartmentalize the work.

2.3.1 Schedule

Gantt Chart

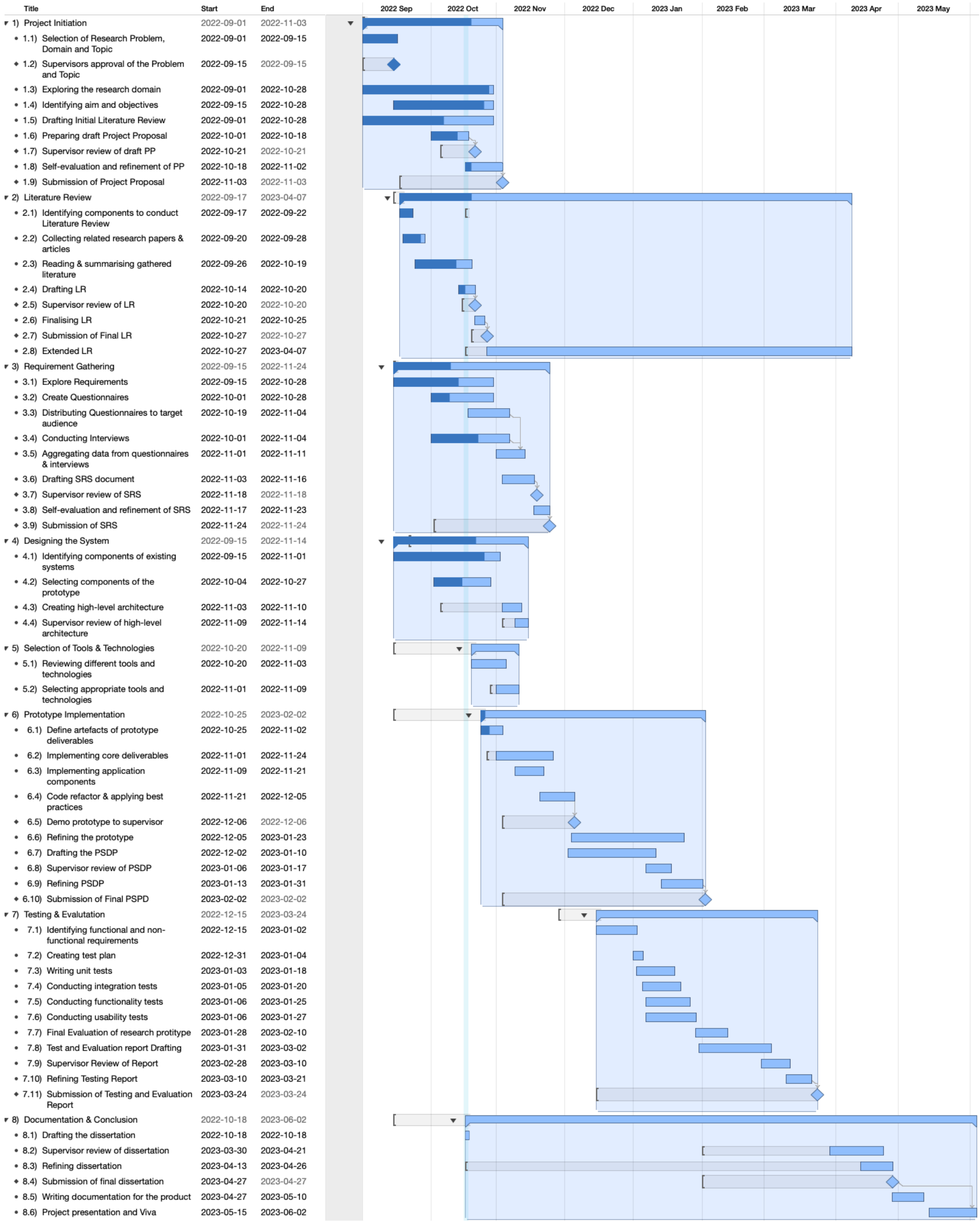

Figure 2.1: Gantt Chart

Deliverables

Deliverable	**Date**
Project Proposal Document The initial proposal of the project	3rd November 2022
Literature Review Document The Critical review of existing work and solutions	14th November 2022
Software Requirement Specification The document specifying requirements to be satisfied and developed as the final prototype and means of collecting data	24th November 2022
System Design Document The document specifies the design developed for the Prognosis Prediction system and overviews of the algorithms to be developed.	1st December 2022
Prototype The prototype with main core features functional	05th December 2022
Thesis The final report documenting the project and research process and decisions	23rd March 2023
Review Research Paper A research paper reviewing existing systems in the Prognosis Prediction domain published at a conference	30th March 2023
Final Research Paper A research paper introducing the Prognosis Prediction system developed at the end of this project	27th April 2023
Public Package / Library	30th April 2023

A public library to access and use the Prognosis Prediction system developed	

Table 2.2: Deliverables

2.3.2 Resource Requirements

The following are the software, hardware and data resources required in order to successfully navigate through the research project, as well as delivering expected solutions based on the objectives.

Software Requirements

- **Operating System (Windows/ macOS/ Linux) –** An operating system that is able to handle the computational demand of Quantum Computing and run/ support language and framework needed in the development process. Windows OS will be used by default.
- **IBM Quantum/ Google Quantum –** A publicly available remote quantum device is crucial to process the quantum machine learning project. IBM Quantum will be used by default for this project.
- **Python** – An all-purpose language widely used and compatible with many quantum machine learning projects. Python will be used to develop QML models
- **Qiskit/ Cirq Frameworks** – Open-source framework working for each IBM and Google quantum devices respectively. Qiskit will be used here by default.
- **PennyLane Python Library** – Python library used to support the development, training and testing of the QML models.
- **NodeJS –** The API that will be needed to communicate with QML backend and the frontend.
- **JavaScript (React)** – Front-end application where the prognosis prediction will be initiated, and results are to be shown.
- **PyCharm/ VSCode** – Integrated Development Environments used to support the development.
- **Google Colab** – Cloud IDE to support the training and testing of the QML models.

- **Zotero/ Mendeley** – Reference management tool used to keep track of research artifacts and manage references. Zotero is used in this project by default.
- **MS Office/ MS Sharepoint/ Google Docs/ Canva/ OmniPlan –** Services to store, design, develop and document the research process.
- **Google Drive/ GitHub –** File back-up for documents and source code.

Hardware Requirements

- **Intel Core i7 CPU or higher** – To perform heavy and intense QML and computational tasks.
- **16GB RAM or higher** – To multi-task smoothly with datasets and qml models running.
- **Disk space of 40GB or higher –** To store project related documentation, source code and data.

Skill Requirements

- Understand the fundamentals of quantum computing and quantum machine learning.
- Ability to successfully build, train and test QML models relevant to prognosis prediction.
- Ability to document technical and scientific writing required for the research.

Data Requirements

- Prognosed pulmonary HRCT imagery with their FVC capacity data – OSCI

2.3.3 Risk Management

Risk Item	Severity	Frequency	Mitigation Plan
Corruption/ access restriction to development source code	5	5	Keeping a back-up of all source code on Version Controlling (GitHub) and external physical drives

Deletion/ corruption of documentation	5	4	Using the cloud-first approach through tools such as Share-point and regular local back-ups
Inability to meet deliverable deadlines within the expected time span	5	4	Making sure personal deadlines are set and met as per the Gantt structure to make sure smooth and complete coverage of all deliverables
Lack of access to quantum hardware	5	5	Making sure multiple cloud quantum computational solutions are explored and ready to use when needed and essential.
Keeping in-par with the latest technological improvements in QML	4	5	Making sure constant literature survey is performed to ensure constant refresh of the latest advancement of the technologies applicable and hardware capacities.

Table 2.3: Risk Mitigation Plan

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.

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.

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].

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.

Saunders, M.N.K., Lewis, P. and Thornhill, A. (2007). *Research methods for business students*, 4th ed. Harlow, England ; New York: Financial Times/Prentice Hall.

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.

Stancioi, C.-M. et al. (2021). Implementing a System for Diagnosing Pulmonary Fibrosis using Hough Algorithm. *2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. Available from https://doi.org/10.1109/SACI51354.2021.9465554.

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.