KidniFy – A mobile based patient care application for Kidney Patients

2023-032

Project Proposal Report

Samarawila D.R.N.

B.Sc. (Hons) Degree in Information Technology specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

March 2023

KidniFy – A mobile based patient care application for Kidney Patients

2023-032

Project Proposal Report

B.Sc. (Hons) Degree in Information Technology specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

March 2023

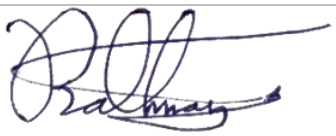
DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Samarawila D.R.N.	IT20235260	

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the Supervisor:	2023.04.06	
Signature of the Co-Supervisor:	2023.04.06	

ABSTRACT

Kidney disease is a significant health issue in Sri Lanka, affecting a substantial portion of the population. Early detection and treatment of the disease are essential to prevent complications and improve patient outcomes. Machine learning (ML) models have shown great potential in predicting kidney disease based on patient data and facilitating early intervention. However, there is a lack of research on the effectiveness of patient-specific ML-based models in the Sri Lankan context.

This undergraduate research project aims to investigate the effectiveness of patient-specific ML-based kidney disease prediction models in the Sri Lankan population and identify the factors that may impact their performance. The study will develop and evaluate ML models that are tailored to Sri Lankan patients' characteristics and clinical data to predict the risk of kidney disease. Additionally, the project will develop a treatment recommendation system based on the predicted risk and patient-specific data to provide personalized treatment options.

The research will contribute to the growing body of knowledge on the use of ML in healthcare and specifically on patient-specific kidney disease prediction and treatment in Sri Lanka. The findings of this project may help improve the accuracy of diagnosis and treatment of kidney disease in Sri Lanka, ultimately reducing the burden of the disease on the healthcare system and improving patient outcomes.

The project will utilize existing datasets, as well as conduct surveys and interviews with healthcare providers and patients, to gather insights into the factors that may influence the performance of these models. The results of the study will be presented in a research report, providing a valuable learning opportunity for the undergraduate researcher. The findings may also inform the development of evidence-based policies and practices to combat kidney disease in Sri Lanka and serve as a foundation for further research in this area.

Keywords: kidney disease, machine learning, prediction models, patient-specific, Sri Lanka, early detection, treatment, healthcare providers, patients, diagnosis, burden, treatment recommendation system.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

AI - Artificial Intelligence

AUC - Area Under the Curve

App - Application

AWS - Amazon Web Services

CKD - chronic kidney disease

CR - Conduct Sprint Reviews

CS - Conduct Sprint Retrospectives

DB - MongoDB

DT - Decision Trees

GLU - Glucose Fasting

JS - JavaScript

LR - Logistic Regression

LOC - Localization

ML - Machine Learning

NodeJS - Node JavaScript

PB - Prioritized Product Backlog

PI - Project Objectives and Scope

PRO - Proteinuria

QA - Quality Assurance

RBC - Red Blood Cells

RN - React Native

RNative - React Native

SL - Sri Lanka

SVM - Support Vector Machines

T-CHO - Total Cholesterol

TF - TensorFlow

TG - Triglycerides

UPCR - Urine Protein and Creatinine Ratio

1 INTRODUCTION

1.1 Background and Literature Survey

Kidney disease is a prevalent health issue in Sri Lanka, affecting approximately 16% of the population. Early detection and management of kidney disease are crucial to prevent further damage and maintain overall health. To predict whether a patient might have kidney disease, doctors in Sri Lanka traditionally use a method that involves asking the patient questions about their symptoms and other factors.

The doctor will first obtain the patient's medical history when predicting kidney disease using this method. They will ask about any symptoms the patient has been experiencing, such as leg swelling, fatigue, changes in urine output, or difficulty passing urine. Additionally, the doctor will inquire about any risk factors the patient may have, such as medical history and medication use. The doctor will also ask about the patient's lifestyle habits, including alcohol or tobacco use, as these can increase the risk of kidney disease [1].

Based on the patient's responses, the doctor may be able to determine whether the patient has kidney disease or is at risk for developing the disease. If the patient exhibits symptoms or has risk factors, the doctor may order further testing, such as blood and urine tests, to confirm the diagnosis.

Over the years, researchers have been exploring more advanced technologies to improve the accuracy of kidney disease prediction. Despite the progress made in using machine learning to identify kidney patients, issues still need to be addressed. Accuracy is one of the main concerns when using these technologies in predicting kidney disease. While machine learning algorithms are effective at identifying patterns in large amounts of data, they may not always accurately predict a patient's risk of developing kidney disease.

To address this issue, we are exploring the development of a mobile application that can accurately predict kidney patients by utilizing machine learning algorithms. Such an application would provide an easy-to-use, accessible tool for patients and healthcare

professionals, potentially revolutionizing how we detect and treat kidney disease. Below are some of the research studies that have been conducted in this area.

Elias Dritsas and Maria Trigka, researchers affiliated with the University of Patras in Greece, studied to predict chronic kidney disease (CKD) using machine learning (ML) techniques. CKD is a condition that causes a gradual decline in kidney function and can ultimately lead to end-stage renal disease and death if not correctly diagnosed and treated. The researchers utilized a class-balancing approach to address the non-uniform distribution of instances in the two classes, performed feature ranking and analysis, and trained and evaluated several ML models based on various performance metrics. The study's results demonstrated that the Rotation Forest (RotF) model outperformed the other models, achieving an Area Under the Curve (AUC) of 100% and high levels of precision, recall, F-measure, and accuracy, with a rate of 99.2% [2].

Chin-Chuan Shih and the team conducted a research study in Taiwan in 2020 to develop efficient risk prediction models for complications and mortality rates associated with chronic kidney disease (CKD). To predict early CKD, the study utilized four data mining algorithms, which included a classification and regression tree, a C4.5 decision tree, a linear discriminant analysis, and an extreme learning machine. The study collected data from an adult health examination program conducted between 2015 and 2019, including information from 19,270 patients from 32 chain clinics and three special physical examination centers. The predictive variable used was the glomerular filtration rate (GFR), and 11 independent variables were considered. Out of the four models, the C4.5 decision tree algorithm demonstrated superior performance, with higher accuracy, sensitivity, specificity, and area under the curve metrics. The study identified several significant risk factors for early CKD, including Urine protein and creatinine ratio (UPCR), Proteinuria (PRO), Red blood cells (RBC), Glucose Fasting (GLU), Triglycerides (TG), Total Cholesterol (T-CHO), age, and gender. The proposed risk prediction models could assist in predicting early CKD and provide valuable insights into managing this healthcare priority by considering personality and health examination representations [3].

In 2023, Piyawat Kantagowit, Fangyue Chen, Tanawin Nopsopon, Arisa Chuklin, and Krit Pongpirul prepared a systematic review protocol to use ML to diagnose CKD. CKD is a significant contributor to morbidity and mortality worldwide, and ML-based decision-support tools have been developed to aid in various aspects of CKD care. The systematic review protocol aims to assess the performance and report the quality of prognostic and diagnostic ML models in CKD diagnosis and prediction. The review will include a systematic search of various databases and compare ML-based models' performance with non-ML-based models as the primary outcome. The secondary analysis will include model use cases, construct, and reporting quality. The results of this systematic review will offer clinicians and technical specialists valuable insights into the current development and potential standardization of ML in CKD care [4].

In the year 2023, a research study was carried out by Ariful Islam, Ziaul Hasan Majumder, and Alomgeer Hussein to examine the capability of using machine-learning approaches for the early diagnosis of chronic kidney disease (CKD). Early detection and appropriate therapy can slow down or halt the progression of the disease, and the use of machine-learning approaches can aid in achieving an early diagnosis. The study investigated the relationship between data factors and the characteristics of the target class and developed a collection of prediction models using machine learning and predictive analytics. At the outset, 25 different variables were taken into consideration, however, only the top 30% of those parameters were found to be most effective for identifying CKD. Twelve classifiers based on machine learning were tested using a supervised learning environment. The XgBoost classifier emerged as the most efficient, with an accuracy rate of 0.983, precision of 0.98, recall of 0.98, and F1-score of 0.98. The study demonstrates the potential of machine learning and predictive modeling to discover new solutions for early diagnosis of CKD and other diseases. Recent advancements in machine learning have demonstrated promising outcomes in the early detection of kidney disease and beyond [5].

In conclusion, Machine Learning (ML) has shown great potential in predicting and diagnosing chronic kidney disease (CKD), but some limitations still need to be addressed. One of the main limitations of ML-based models in predicting and diagnosing CKD is their accuracy. Although these models have shown promising

results, they still need to be improved to achieve higher levels of accuracy. Another area for improvement is the availability of these models for day-to-day users. These models are often complex and require specialized knowledge, limiting their accessibility.

Overall, ML-based models have the potential to revolutionize the diagnosis and treatment of CKD. Still, more work needs to be done to ensure their accuracy and accessibility for day-to-day users. By addressing these limitations, we aim to improve the quality of care for patients.

1.2 Research Gap

The current applications of machine learning in identifying kidney disease patients lack accuracy, and there is a lack of usable prototypes that patients can try out on app stores. Moreover, the absence of localization in local languages makes it difficult for patients to access and understand the content. This research gap highlights the need for a comprehensive solution that is accurate, easily accessible, and localized.

Our mobile application aims to address this research gap by providing an accurate prediction of a user's kidney health status using multiple machine learning models. By using a combination of various machine learning models, we can increase the accuracy of predictions and provide patients with more reliable results.

In addition to accuracy, the application will also increase the availability of kidney disease prediction tools for patients in Sri Lanka. With the majority of the population using mobile phones as their primary means of accessing the internet, a mobile application is the most effective way to reach a large number of people. Our user-friendly app will prompt patients to complete a questionnaire, and based on their responses, provide a relative frequency representation indicating their risk of developing kidney diseases.

Based on the patient's location, the application will display a list of the nearest medical facilities where they can go through the necessary tests. This feature will increase the accessibility of medical facilities for patients, especially those in remote areas. By

guiding patients to nearby medical facilities, we hope to increase the likelihood of early detection and timely treatment.

Localizing the application in Sinhala and Tamil languages will also play a critical role in increasing the awareness of kidney disease among the general public. By providing content in local languages, patients can access and understand the information more easily, empowering them to take proactive steps towards maintaining their kidney health.

In conclusion, our mobile application addresses the research gap in the accuracy, availability, and localization of kidney disease prediction tools. By using multiple machine learning models, providing a user-friendly interface, and localizing content in local languages, we aim to increase the accuracy and awareness of kidney disease among patients in Sri Lanka.

2 RESEARCH PROBLEM

Kidney disease is a major health concern in Sri Lanka, with a prevalence rate of approximately 17.7% among the population. The disease is characterized by the loss of kidney function over time, and can lead to a range of complications, including high blood pressure, anemia, and bone disease. Detecting and treating kidney disease early is essential for preventing complications and improving patient outcomes.

Despite the importance of early detection and treatment, there are significant challenges in identifying individuals at risk of kidney disease in Sri Lanka. One major challenge is the lack of access to timely and accurate diagnostic tools, particularly in rural areas where healthcare resources are often limited. Another challenge is the lack of awareness among the general public about the risks and symptoms of kidney disease, which can lead to delayed diagnosis and treatment.

To address these challenges, there is a growing interest in developing machine learning (ML) models for kidney disease prediction, which can help identify individuals at risk

of the disease and facilitate early intervention. However, there is a need for research to determine the effectiveness of these models in the Sri Lankan context, and to identify the factors that may impact their performance.

Therefore, the research problem for this proposal is to investigate the effectiveness of ML-based kidney disease prediction models in the Sri Lankan population, and to identify the factors that may influence their performance. Specifically, the proposed research will address the following research questions:

1. How accurate are existing ML-based models for kidney disease prediction in the Sri Lankan population?
2. What factors may influence the performance of these models, such as age, gender, socioeconomic status, and geographic location?
3. What is the level of awareness among the general public in Sri Lanka about the risks and symptoms of kidney disease, and how can this be improved to facilitate early detection and treatment?
4. What are the current barriers to accessing timely and accurate diagnostic tools for kidney disease in Sri Lanka, and how can these be addressed?

To address these research questions, the proposed study will employ a mixed-methods approach, combining quantitative analysis of existing datasets with qualitative interviews and surveys of healthcare providers and patients. The study will also involve collaboration with key stakeholders in the healthcare sector, including government agencies, non-governmental organizations, and healthcare providers, to ensure that the findings are relevant and actionable.

The proposed research has the potential to make a significant contribution to the field of kidney disease prediction and treatment in Sri Lanka, by identifying the factors that may impact the performance of ML-based models and by proposing strategies to improve early detection and treatment of the disease. Ultimately, the findings of this research may help reduce the burden of kidney disease in Sri Lanka and improve patient outcomes.

3 OBJECTIVES

3.1 Main Objective

Main goal of this component is developing an effective solution for predicting a user's kidney health condition. Our mobile application will prompt patients to complete a questionnaire, and based on their answers, the app will provide a prediction as to whether the patient is at risk of contracting any kidney diseases.

3.2 Specific Objectives

- Based on the patient's answers, display a graphical representation indicating the percentage of risk they have of developing any type of kidney disease.
- Provide information on the available treatment options and testing procedures for the type of kidney disease with the highest percentage risk.
- Based on patient’s location, display a list of the nearest medical facilities where they can go through the necessary tests.

4 METHODOLOGY

4.1 System Architecture

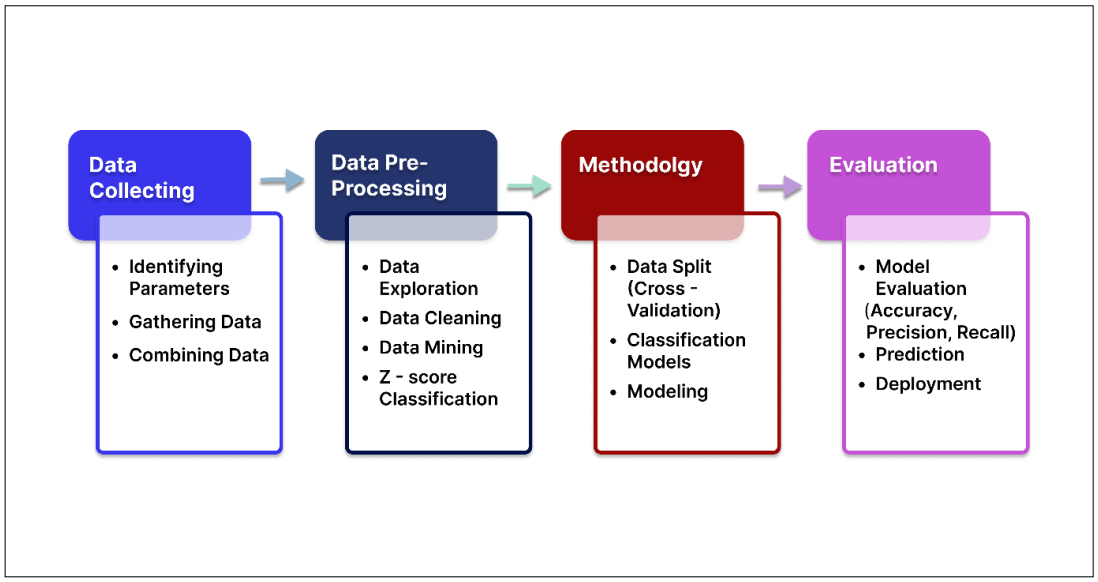


Figure 1: The System Architecture for the component

As a mobile application for predicting chronic kidney diseases, our architecture has several key components that work together seamlessly. Our system is designed to collect data from patients, preprocess that data, train several models, evaluate the models' accuracy, and then use the best-performing model to make predictions based on user inputs within the mobile app. First, we collect data from patients. This can include a variety of factors that may influence a patient's likelihood of having a chronic kidney disease, such as their age, medical history, and lifestyle habits. Once we've collected this data, we preprocess it to ensure it's ready for analysis. Next, we train several models on this preprocessed data. We use a variety of machine learning algorithms to do this. By training multiple models, we're able to compare their accuracy and select the best-performing one. Once we've trained our models, we evaluate their accuracy. This allows us to determine which model is most accurate and should be used for making predictions. Once we've identified the best model, we integrate it into the mobile app. Within the mobile app, we ask patients questions that are related to the data we collected earlier. By analyzing these answers, the app is able to make a prediction about the patient's likelihood of having a chronic kidney disease.

We use a relative frequency representation to communicate this prediction to the patient.

4.2 Commercialization of the Product

1. Develop a comprehensive marketing strategy: The first step would be to develop a marketing strategy that targets kidney patients and those at risk of kidney disease. This could involve promoting the app on social media, targeted advertising, and outreach to health organizations and medical professionals.
2. Launch the mobile app: Once the marketing strategy is in place, launch the mobile app. Ensure that the app is user-friendly and that features such as the risk prediction tool, image analysis, and water quality measurement device are easy to use.
3. Offer a free trial: To encourage users to try out the app and see its value, offer a 30-day free trial with access to all features. During the trial period, users can evaluate the app and decide whether to continue using it by purchasing a subscription. Consider offering incentives to users who subscribe after the free trial, such as a discounted subscription price or additional features not included in the trial version.
4. Partner with healthcare providers: Partner with healthcare providers such as hospitals and clinics to offer our app as a resource for their patients. We could also consider partnering with healthcare insurance providers to offer the app as part of their member benefits.
5. Collect and analyze data: Collect data from app users to identify patterns and trends in kidney disease prevalence and risk factors. Use this data to improve the app's features and functionality and to develop targeted marketing campaigns.

5 SOFTWARE / HARDWARE METHODOLOGY

5.1 Software Methodology

We decided to use Agile methodology in our development process. Agile methodology is an iterative and collaborative approach to software development that emphasizes flexibility, feedback, and continuous improvement. It is particularly well-suited to our scenario because it allows you to work in short sprints, gather feedback from stakeholders, and adjust the project's scope and priorities based on their input.

1. **Plan and define the project scope**: Define the project objectives and scope. Identify the stakeholders involved, the desired outcomes, and the constraints affecting the project's development. Identify and prioritize the features of the application and the data to be collected from the patients, such as their symptoms, medical history, lifestyle factors and other necessary health information.
2. **Create the backlog**: Create a prioritized product backlog. Each item should be clear and concise and represent a tangible functionality. Consider integrating the machine learning algorithms and the other technologies in the backlog as separate stories.
3. **Plan sprints**: Using the product backlog, plan a series of sprints that each deliver a subset of the backlog. Sprints should be short and iterative (two to four weeks) to allow for adjustments in requirements and feedback from stakeholders.
4. **Conduct sprint reviews**: At the end of each sprint, conduct a sprint review with stakeholders to assess progress and demonstrate completed work. Gather feedback that can be used to adjust priorities in the next sprint.
5. **Conduct sprint retrospectives**: Conduct a retrospective after each sprint review to identify areas for improvement and potential changes in the project approach. Take the feedback from the retrospective into account when planning the next sprint.
6. **Develop machine learning algorithms**: Use TensorFlow to develop the machine learning algorithms that will be used to analyze the patient data and

categorize patients into Safe, Cautious, or Danger zone. Train the models on a large set of data representative of the patient population.

7. **Build the application**: Develop the React Native application, integrating the machine learning algorithms and other features. Build the user interface, data analysis, and recommendations engine. Use NodeJS to develop the backend API that interacts with the application and stores the patient data in MongoDB.
8. **Test and QA**: Test the application thoroughly to ensure that it meets the functional and non-functional requirements of the product backlog. This includes testing usability, performance, security, and accessibility. Use automated tests where possible to improve efficiency.
9. **Deploy and maintain**: Deploy the application to AWS, ensuring it is scalable and secure. Monitor the application for bugs and issues and regularly update it with new features and enhancements. Use AWS services such as Elastic Beanstalk, Lambda, and API Gateway to manage the deployment and maintenance of the application.
10. **Iterate and improve**: Use feedback from users and stakeholders to improve the application continuously. This may involve adding new features, improving the machine learning algorithms, or updating the user interface.

5.2 Tools and Technologies

- **React Native**: React Native can be used to develop a mobile app that can run on both major mobile operating systems. It will be used to develop the front end of the application.
- **NodeJS**: NodeJS can be used to develop the backend APIs that the mobile app will interact with. The APIs can be used to process patient data, retrieve the GFR value, run machine learning algorithms, and provide the recommended health advice and treatment options.
- **TensorFlow**: TensorFlow can be used to develop machine learning models that analyze patient data and categorize them into safe, cautious, or danger zones. The models can be trained to provide personalized recommendations based on the patient's specific health conditions and medical history.

- **MongoDB**: MongoDB can be used to store patient information and health data. The database can be used to securely store patient information, and the data can be accessed by machine learning algorithms and backend APIs.
- **AWS**: AWS can be used to host the backend APIs and machine learning models. AWS provides a secure and scalable platform for hosting these components, which ensures that the app can handle many users and requests.
- **Payhere**: Payhere can be used to integrate various payment methods into the app, including Visa and Mastercard, digital wallets, and bank transfers. This makes it easier for patients to pay for medical help and treatment options.

6 DESCRIPTIONS OF PERSONAL AND FACILITIES

Member	Component	Task
Samarawila D.R.N.	Predictive analysis for risks of having a kidney disease	• Identify kidney diseases and the types of tests used to diagnose them. • Collect data from real patients. • Pre-process the collected data through data cleaning, normalization, and feature collection. • Develop a machine learning model to recognize and categorize different types of kidney diseases. • Train the machine learning model using pre-processed data. • Create an API that allows the mobile app to communicate with the machine learning model.

		• Integrate the API with the frontend of the mobile app. • Develop a list of recommended tests and steps for proper diagnosis and treatment based on the predicted kidney disease. • Develop a system to suggest nearby medical facilities for treatment.

Table 1:Description of Personal Tasks

7 BUDGET AND BUDGET JUSTIFICATION

Resources	Estimated Price (LKR)
Cloud server host	25000
Travelling	10000
Internet	5000
Stationery	2000
Miscellaneous	5000
Total	**47,000**

Table 2: Estimated expenditure

The proposed budget total cost amount is LKR 47000. To cover this expenditure, our group plans to collect funds from group members. The budget table should detail all the project expenses, including any necessary equipment, materials, or services required to complete the project. These costs might change in the future due to unforeseen circumstances or unexpected expenses, but with a clear budget plan and contributions from group members, the project can be completed successfully.

8 GANTT CHART

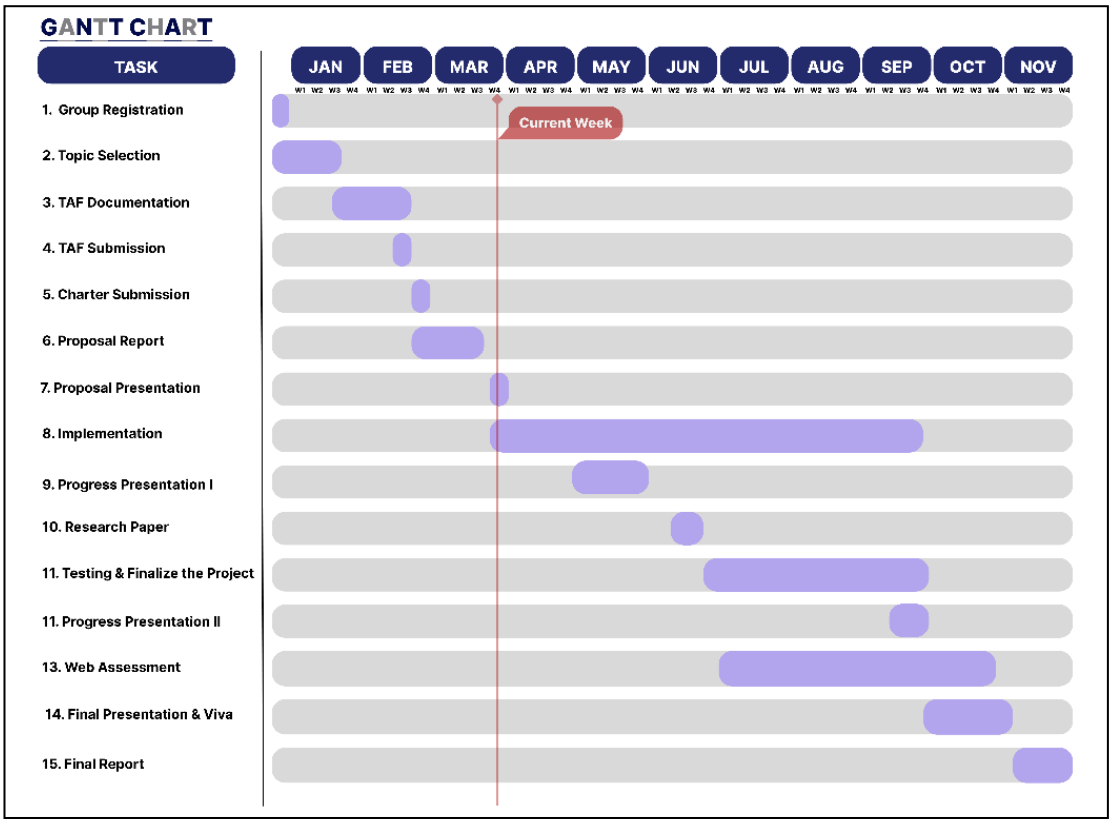


Figure 2:Gantt chart for the component

The Gantt chart above represents our proposed plan for the research project, with a focus on my component. We have made progress from January until this week, which is indicated by the green color. The remaining tasks are in purple, which we plan to complete in the coming months.

This Gantt chart is an essential tool for our research plan, as it helps us manage our time and resources effectively. It shows the timelines for each task, the dependencies between them, and the overall project schedule.

9 WORK BREAKDOWN CHART

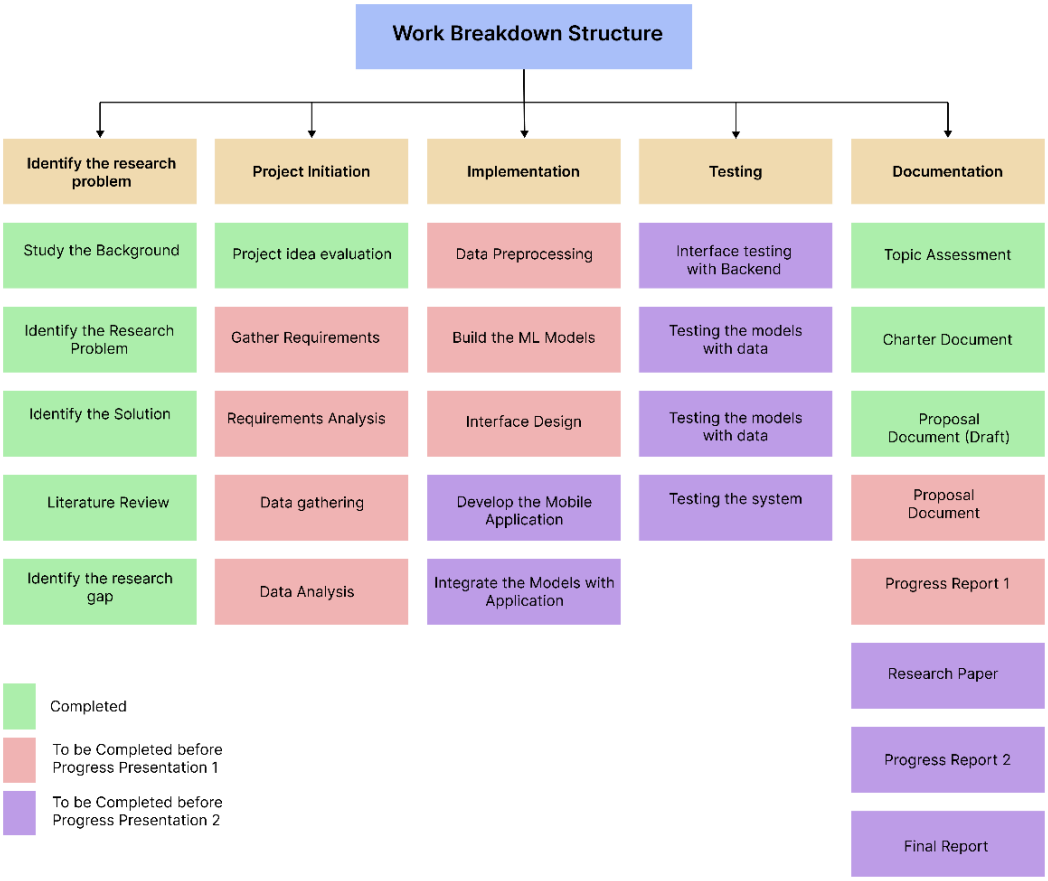


Figure 3:Work Breakdown Chart for the component

The work breakdown structure diagram for this project comprises five stages: the identification of the research problem, the initial stage, implementation, testing, and documentation. In the first stage, the research problem will be identified and analyzed thoroughly to develop a comprehensive understanding of the project's objectives and requirements. The initial phase will involve designing the software components of the mobile app-based prediction system for chronic kidney patients. The implementation stage will focus on developing machine learning algorithms that analyze patient data, including their symptoms, family history, medical history to predict about patient's kidney status. The testing stage will be used to evaluate the accuracy and effectiveness of the prediction using simulated patient data. Finally, the documentation stage will

involve preparing comprehensive documentation of the project, including technical specifications, user manuals, and the project report.

10 REFERENCES

[1] National Insitute of Diabetes and Digestive and Kidney Diseases," [Online]. Available: https://www.niddk.nih.gov/health-information/kidney-disease/chronic-kidney-disease-ckd/causes#:~:text=Diabetes%20and%20high%20blood%20pressure,type%20of%20treatment%20you%20receive..

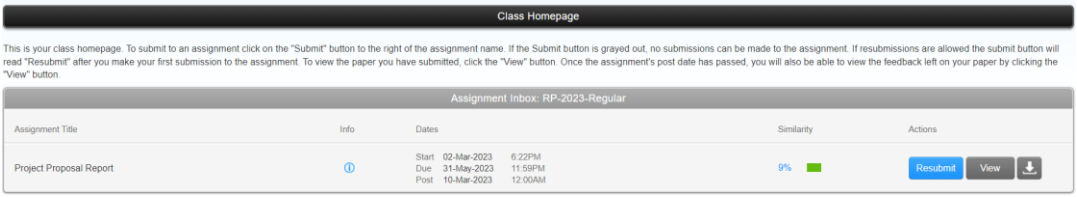
[2] M. T. Elias Dritsas, "MDPI," 2022. [Online]. Available: https://www.mdpi.com/2504-2289/6/3/98. [Accessed 9 April 2023].

[3] C.-J. L. G.-D. C. C.-C. C. Chin-Chuan Shih, "National Library of Medicine," July 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7399976/. [Accessed April 2023].

[4] F. C. T. N. A. C. P. Piyawat Kantagowit, "PLOS ONE," February 2023. [Online]. Available: https://journals.plos.org/plosone/article/authors?id=10.1371/journal.pone.0278729.

[5] Z. H. M. A. H. Ariful Islam, "Science Direct," 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2153353923000032.

11 APPENDIX


Class Homepage
This is your class homepage. To submit to an assignment click on the "Submit" button to the right of the assignment name. If the Submit button is grayed out, no submissions can be made to the assignment. If resubmissions are allowed the submit button will read "Resubmit" after you make your first submission to the assignment. To view the paper you have submitted, click the "View" button. Once the assignment's post date has passed, you will also be able to view the feedback left on your paper by clicking the "View" button.
Assignment Inbox: RP-2023-Regular
Assignment Title
Info
Dates
Similarity
Actions
Project Proposal Report
Start 02-Mar-2023 6:22PM
Due 31-May-2023 11:59PM
Post 10-Mar-2023 12:00AM
9%
Resubmit
View