Central University of Technology, Free State

Development of a Predictive Analytics Dashboard for Early Risk Detection of Cardiovascular Disease in South African public hospital

Your Group Members' Names

Student Details

221001482 KM Ramela (Leader)

221013309 SR Letsoara (Responsible for creating the whole document for submission)

219001662 JP Mafanyolle

222028938 SK Mofokeng

221014333 VOP Luhlabo

Dissertation submitted in fulfilment of the requirements for the degree

ADVANCED DIPLOMA IN

INFORMATION TECHNOLOGY

in the

Department of Information Technology

Faculty of Engineering, Built Environment and Information Technology at the

Central University of Technology, Free State

Supervisor: Dr J Coetzer, Ph. D. (IT.)

Co-supervisor: ………………… (if applicable)

BLOEMFONTEIN

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
AUC	Area Under the Curve
BMI	Body Mass Index
CHD	Coronary Heart Disease
CPS	Cyber Physical Systems
CPPS	Cyber-Physical Production Systems
CVD	Cardiovascular Disease
EHR	Electronic Health Record
HCI	Human-Computer Interaction
HMI	Human-Machine Interaction

HTI	Human-Technology Interaction
ICT	Information and Communication Technologies
IoE	Internet of Everything
IoS	Internet of Services
IT	Information Technology
ML	Machine Learning
POPIA	Protection of Personal Information Act
RF	Random Forest
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
SSA	Sub-Saharan Africa
XGBoost	Extreme Gradient Boosting

ABSTRACT

Cardiovascular disease (CVD) accounts for approximately 17.3% of deaths in South Africa, with nearly 20% of high risk patients remaining undetected until complications arise in public hospitals. Existing healthcare information systems in South African public hospitals including those in Bloemfontein, predominantly rely on descriptive analytics that report historical trends rather than predicting future health risks. This reactive approach results in delayed interventions, poor patient outcomes and inefficient resource utilisation. Although machine learning models have demonstrated high predictive accuracy in experimental settings, their integration into real time clinical decision support systems within resource constrained South African public hospitals remains limited. This study aims to design, develop and evaluate a predictive analytics dashboard for early CVD risk detection using machine learning algorithms. A concurrent mixed methods design is employed, combining quantitative analysis of cardiovascular datasets with qualitative feedback from healthcare professionals. The research develops and validates four machine learning models Logistic Regression, Random Forest, XGBoost and Support Vector Machine to identify key risk factors and predict early cardiovascular disease risk. A lightweight interactive dashboard is designed to visualise patient risk profiles through colour-coded indicators, automated alerts and explainable AI features to support clinical decision making. Expected outcomes include a functional prototype that bridges the gap between predictive analytics research and practical clinical implementation, contributing to proactive healthcare interventions in South African public hospitals in Bloemfontein .

Research Topic

Development of a Predictive Analytics Dashboard for Early Risk Detection of Cardiovascular Disease in South African public hospitals.

Problem Statement

Cardiovascular disease (CVD) accounts for approximately 17.3% of deaths in South Africa, with nearly 20% of high-risk patients remaining undetected until complications arise (Statistics South Africa, 2022; Parker et al., 2021). Public hospitals in Bloemfontein (national district hospital) produce large amounts of digital health data including patient records, laboratory results and diagnostic information. Existing healthcare information systems mainly rely on descriptive analytics, which report historical trends rather than predicting future health risks (Goyal et al., 2023). As a result, hospitals often respond to cardiovascular conditions only after complications have developed, leading to delayed interventions, poor patient outcomes, and inefficient use of healthcare resources.

Although machine learning models(algorithms) have demonstrated high predictive accuracy (AUC up to 0.96) in experimental healthcare studies, their application within real time clinical decision support systems in South African public hospitals remains limited. This highlights a critical research gap in the lack of an integrated system capable of transforming complex predictive algorithms into practical, real time clinical tools for early cardiovascular risk detection. The potential of Big Data and predictive analytics to support proactive healthcare interventions remains largely underutilized in resource constrained environments. This study therefore seeks to develop, validate and evaluate a predictive analytics dashboard that converts raw clinical data into real time cardiovascular risk profiles to support proactive clinical decision making.

Research Aim

The aim of this study is to design and evaluate a predictive analytics dashboard for an early risk detection of cardiovascular disease in South African public hospitals using machine learning algorithms.

Research Objectives

1. To analyse cardiovascular disease related healthcare datasets in South African public hospitals to identify key risk factors for early CVD prediction.

2. To develop and validate machine learning models capable of predicting early cardiovascular disease risk

3. To design a predictive analytics dashboard that visualizes patient cardiovascular risk profiles to support clinical decision making

4. To evaluate the effectiveness and usability of the predictive analytics dashboard in supporting real time clinical decision making

Research Questions

1. What cardiovascular risk factors can be identified from healthcare datasets in South African public hospitals to support early CVD prediction?

2. How accurately will the machine learning models predict early cardiovascular disease risk using healthcare data from South African public hospitals?

3. How will a predictive analytics dashboard be designed to effectively visualize patient cardiovascular risk profiles for doctors?

4. To what extent does the predictive analytics dashboard improve real time clinical decision-making in detecting cardiovascular disease risk?

Literature Review

Introduction

Healthcare systems globally are shifting from reactive treatment models to proactive data driven approaches. This shift is driven by the rapid growth of Big Data, electronic health records (EHRs) and advanced analytics (Raghupathi & Raghupathi, 2014). While large amounts of patient data are generated through hospital systems laboratory reports and monitoring devices the effective use of this data for clinical decision making remains limited particularly in resource constrained environments (Obermeyer & Emanuel, 2016).Predictive analytics and machine learning offer significant potential to address this limitation by enabling early disease detection and improved decision making (Kosaraju, 2024; Rana & Shuford, 2024). This is especially important in South Africa where cardiovascular disease accounts for approximately 17.3% of deaths with many high risk patients remaining undiagnosed until severe complications occur (Statistics South Africa, 2022; Parker et al., 2021).This study critically examines existing research on predictive analytics for CVD detection identifies key gaps and establishes the foundation for developing a predictive analytics dashboard for early risk detection in South African public hospitals(Bloemfontein).

Cardiovascular Disease Burden and Detection Gaps

Cardiovascular disease is a major public health challenge in Sub Saharan Africa (SSA), driven by demographic transitions and increasing frequency of non-communicable diseases. (Mensah et al. 2015) reported that CVD accounted for approximately one million deaths in SSA in 2013 representing a significant proportion of total mortality. In South Africa this burden is particularly severe. CVD contributes to 17.3% of all deaths, with stroke and coronary heart disease among the leading causes (Statistics South Africa, 2022). Despite this early detection remains insufficient (Parker et al. 2021) highlight a critical detection gap where many individuals are only diagnosed after experiencing life threatening events such as heart attacks or strokes.

Evidence across SSA further demonstrates the high frequency of modifiable risk factors. (Mzunga et al.2025) found high rates of obesity, hypertension and dyslipidaemia in Malawi while (Tantchou Tchoumi and Butera 2013) identified hypertension as a dominant factor in cardiac patients in Cameroon. These findings indicate that risk factors are both widespread and measurable reinforcing the potential for predictive analytics to enable early identification. However, the variability in disease patterns across regions suggests that models developed in high income countries may not be directly applicable to South African populations. This highlights the need for context specific predictive solutions.

Machine Learning Applications for CVD Prediction

Machine learning has demonstrated high performance in predicting cardiovascular risk compared to traditional statistical models. (Weng et al. 2017) found that machine learning algorithms outperform established models such as the Framingham Risk Score in predictive accuracy. Similarly (Alaa et al. 2019) reported that automated machine learning approaches achieved high predictive performance with AUC values up to 0.96. Collective methods such as Random Forest and XGBoost are particularly effective due to their ability to model complex non-linear relationships (Daghistani & Alshammari, 2020. Ahmed et al.2026) further support this demonstrating that collective models achieve accuracy rates between 90–95% in healthcare prediction tasks. These models are also more resilient in handling missing and incomplete data, a common issue in public healthcare systems. Despite these advantages most healthcare systems still depend on descriptive analytics rather than predictive approaches (Goyal et al., 2023). This limits the ability of doctors to anticipate disease risk and intervene early. Predictive analytics therefore represents a critical progress toward proactive and personalised healthcare (Sidey-Gibbons & Sidey-Gibbons, 2019).

Limitations and contradictions in existing research

Despite promising results several limitations constrain the application of predictive analytics in hospitals. A key challenge is the lack of clarity in machine learning models. Rosli et al. (2024) argue that many models function as "black boxes," making it difficult for doctors to understand how predictions are generated. This reduces trust and limits adoption in practice. Data quality is another significant issue. (Mensah et al. 2015) highlight the limited availability of reliable health data in SSA which affects model development and validation. Additionally models developed in developed countries often fail when applied to diverse populations (Weng et al. 2017) note that traditional models such as the Framingham Risk Score perform poorly outside their original populations. While the World Health Organization (2004) suggests that CVD is rapidly increasing in Africa (Mensah et al. 2015) argue that the increase is more gradual and linked to demographic changes. Compatibility challenges further limit the integration of predictive systems into clinical workflows (Armoundas et al., 2024).

Critical Evaluation and Research gap

A major limitation in existing studies is the implementation gap (Kehinde 2025) argues that predictive models often operate independently of hospital systems limiting their practical usefulness Additionally many studies lack external validation and are based on datasets from developed countries reducing their applicability to South African contexts (Alaa et al., 2019; Ahmed et al., 2026). Another critical gap is the absence of integrated dashboards that combine predictive analytics with intuitive visualisation. Existing systems often lack user-friendly interfaces for Doctors (Shah et al., 2024). Effective dashboards must support real time instant decision making (Dowding et al., 2015; Few, 2013). This literature reveals three key gaps: a contextual gap, an implementation gap and a system gap. This study addresses these by developing a predictive analytics dashboard specialized to South African public hospitals. The study adopts a mixed methods design informed by Creswell (2014) and contextual considerations highlighted by Coetzer (2020). By integrating machine learning with an interactive dashboard this research contributes a practical solution that bridges predictive analytics and clinical decision making.

Methodology

This study adopts a concurrent mixed methods design (Creswell, 2014) combining quantitative analysis of cardiovascular disease datasets with qualitative feedback from healthcare professionals and lecturers to design and evaluate a predictive analytics dashboard for early CVD risk detection in South African public hospitals (Coetzer, 2020). The design is scaled to national district hospital in Bloemfontein focusing on a manageable set of data and four machine learning models and a lightweight dashboard prototype.

Research Setting and Data Sources

The study will not require direct access to live hospital information systems. Instead it will use:

- Publicly available CVD datasets, primarily the big cities health data inventor and Framingham heart study style dataset from Kaggle which both contains approximately 20 000 records with 15 variables related to age, blood pressure, cholesterol, BMI, smoking status, diabetes and 10 year coronary heart disease (CHD) risk (Kaggle, n.d.; Mensah et al., 2020).

These datasets will be used to simulate how a predictive dashboard could function in a South African setting without exposing real patient data.

Sampling Strategy

Because the main dataset is a fixed public Kaggle dataset sampling is predefined; the entire available dataset will be used for model development and validation. For model transfer experiments with a smaller national district hospital dataset and all available cases will be included.

For the qualitative component a small, targeted sample of participants might be selected. Participants and give feedback the system accuracy and impact on their decision making the percipients will include

- Random people and or nursing.
- Professional that might have to make a comment on the system

Sampling will continue until main themes about usability and trust are reasonably clear consistent with small scale qualitative studies (Creswell, 2014).

Quantitative Data Collection and Variables

Quantitative data will be collected from the Kaggle CCD, and big data dataset Key variables include:

- Demographics: age and sex.
- Clinical risk factors: systolic and diastolic blood pressure, total cholesterol, BMI, glucose level, smoking status, diabetes and hypertension.
- Outcome variable: 10-year risk of CHD (binary: 0 = no risk and 1 = risk).

Data will be imported into Python for cleaning and preprocessing. Missing values will be addressed using simple methods such as mean/median imputation or removal of records with excessive missing data. Variables will be scaled or normalised where necessary and the dataset will be split into training 70% and test 30% sets.

Model Development and Validation

The study will develop and validate machine learning models to predict early cardiovascular disease risk:

Algorithms used:

- Logistic Regression
- Random Forest
- XGBoost
- SVM

Model training and validation:

- Models will be trained on the training subset of the Kaggle dataset.
- Performance will be evaluated on the test set using accuracy, sensitivity, specificity and the area under the ROC curve (AUC).

This approach keeps the technical work manageable while still demonstrating the potential of predictive analytics for CVD risk detection.

Dashboard Design

A lightweight predictive analytics dashboard will be developed to visualise cardiovascular risk profiles. The dashboard will be implemented using a simple web application for basic dashboard tools such as Tableau. The design will draw on principles of information dashboard design (Few, 2013) and healthcare dashboard literature (Dowding et al., 2015; Shah et al., 2024).

Key dashboard components will include:

Individual patient view:

- A form where the user enters data like (age, blood pressure, cholesterol and others) to feed into the trained model.
- The dashboard returns a predicted risk category (eg. low) and highlights the top 3 contributing factors.

Summary view:

- Charts showing the proportion of low, medium and high risk patients in the dataset.
- Bar charts or contribution plots showing the most important risk factors.

The dashboard will use the Kaggle dataset as the backend; it will not require integration with a live hospital database, making it suitable for the project.

Evaluation of Effectiveness and Usability

The effectiveness and usability of the dashboard will be evaluated using a mixed approach:

Quantitative evaluation:

- Report standard performance metrics (accuracy, AUC, sensitivity and specificity) on the test set.
- Show how often the dashboard would correctly flag high risk patients versus those it misses.

Qualitative evaluation:

- Conduct short individual interviews or a short questionnaire with participants.
- Use a simple Usability and Trust Scale (Likert type questions) plus a few open-ended questions.
- The open responses will be thematically analysed (Creswell, 2014) to identify main themes that concerns clarity, usefulness and trust (in decision making) in the dashboard.

Ethical consideration

All data from Kaggle are open access and anonymised, the study will comply with Kaggle's terms of use by citing the dataset source. Any other data dataset will be for qualitative interviews participants will provide informed consent using a brief written form. Their responses will be anonymised in reports and only group level or de-identified quotations will be presented.

Consent form template

Participant Consent Form

Research Title: Development of a Predictive Analytics Dashboard for Early Risk Detection of Cardiovascular Disease in South African Public Hospitals

Researcher: Koena Mosa Ramela

Email: 221001482@stud.cut.ac.za

1. Purpose of the Study

The aim of this study is to design and evaluate a predictive analytics dashboard for an early risk detection of cardiovascular disease in South African public hospitals using machine learning algorithms.

2. What Participation Involves

- Participating voluntarily and having the right to withdraw.

3. Voluntary Participation

Taking part is voluntary and may withdraw at any given time.

4. Right to Withdraw

- Before submitting the questionnaire: You may withdraw at any time by closing the form without submitting it.
- After submitting the questionnaire: You may contact the researcher through email and request the data to be deleted.
- During the interview: You may withdraw anytime.

5. Confidentiality and Anonymity

- Questionnaire: Data will be kept anonymous and will not be shared. Only the researchers may/or know who submitted the questionnaire or completed the interview.
- Interview: Confidential, meaning your name will be stored separately from your answers in a password protected file and any quotes will be anonymous.

6. Data Storage – After processing the data will be removed.

7. Researcher Contact Details

Researcher: Koena M Ramela, Vinjiwe OP Luhlabo, Selloane Letsoara, Seruwe K Mofokeng, Josephine P Mafanyolle

Researcher Email: 221001482@stud.cut.ac.za
Ethics Committee: Central University of Technology

Ethics Email: ethics@cut.ac.za

8. Consent Statement

I understand that this study participation is voluntary, I can withdraw after submitting the questionnaire and I agree to take part.

I consent to the following:	Yes	No
I agree to complete the questionnaire		
I agree to the optional interview		

Name (strictly for interview purpose): __________________
Signature: __________________
Date: __________________

Researcher Signature: _________________ Date: __________________

- Questionnaire template

Project:

Predictive Analytics Dashboard for Early Risk Detection of Cardiovascular Disease (CVD) in South African Public Hospitals

Purpose:

Collect data from potential users regarding needs, challenges and expectations related to the proposed system

Target Respondents: Clinicians, nurses, hospital administrators in Bloemfontein national hospital.

You are invited to participate in a research study conducted by Koena Mosa Ramela(and other members) to design a Predictive Analytics Dashboard for Cardiovascular Disease (CVD) risk detection in Bloemfontein National district hospital.

Your feedback is crucial in helping us understand the current challenges in CVD management and the features that would be most valuable in a predictive analytics system. This questionnaire will take approximately 10-15 minutes to complete.

Your responses will be kept strictly confidential and anonymous. No identifying information will be collected. Data will be used solely for research purposes to inform the design of the proposed system.

Participation is voluntary and you may withdraw at any time without consequence.

By clicking 'Next' you consent to participate in this study.

For questions contact: KM Ramela 221001482@stud.cut.ac.za

SECTION A: DEMOGRAPHIC AND PROFESSIONAL PROFILE

Question 1: What is your primary role in the healthcare facility?

• Medical Doctor / Specialist

• Nurse (Registered / Enrolled)

• Other (please specify): ____________

Question 2: How many years have you worked in the South African public healthcare sector?

• Less than 1 year

• 1-5 years

• 6-10 years

• 11-20 years

• More than 20 years

Question 3: In which type of healthcare facility do you primarily work?

• District Hospital

• Regional Hospital

• Tertiary / Academic Hospital

• Community Health Centre

• Clinic

• Other (please specify): ____________

Question 4: What is your department or area of work? (Select all that apply)

• Cardiology / Heart Unit

• Emergency Department

• General Medicine

• Outpatient Department

• Administration / Management

• Information Technology

• Public Health / Epidemiology

• Other (please specify): ____________

SECTION B: CURRENT EXPERIENCE AND CHALLENGES

Question 5: How frequently do you encounter patients with or at risk of

cardiovascular disease in your work?

• Daily

• Several times per week

• Weekly

• Monthly

• Rarely / Never

Question 6: What methods do you currently use to assess CVD risk in patients? (Select all that apply)

• Manual calculation (e.g., Framingham Risk Score)

• Clinical judgment / Experience only

• Electronic Health Record (EHR) alerts

• Paper-based risk assessment forms

• Laboratory result reviews

• No formal risk assessment process

• Other (please specify): ____________

Question 7: How would you rate the effectiveness of current CVD risk

assessment methods in your facility?

Type: Likert Scale (1-5)

Scale: 1 = Very Ineffective | 2 = Ineffective | 3 = Neutral |

4 = Effective | 5 = Very Effective

Question 8: What are the main challenges you face in identifying high-risk CVD patients early? (Select all that apply)

• Lack of integrated patient data systems

• No automated risk calculation tools

• Limited access to patient history

• Time constraints during consultations

• Insufficient training on risk assessment tools

• Poor data quality / missing information

• No alerts or reminders for high-risk patients

• Limited resources / staff shortages

• Other (please specify): ____________

Question 9: How often do you think high-risk CVD patients are missed or identified too late in your facility?

• Very often (multiple times per week)

• Often (weekly)

• Sometimes (monthly)

• Rarely

• Never / Don't know

Question 10: Please describe a specific situation where early identification of CVD risk could have improved patient outcomes.

"Describe the situation, what happened and what could have been done differently."

SECTION C: USER NEEDS AND EXPECTATIONS

Question 11: How valuable would a predictive analytics dashboard be for CVD risk detection in your daily work?

Type: Likert Scale (1-5)

Scale: 1 = Not at all valuable or 2 = Slightly valuable or 3 = Moderately valuable

or 4 = Very valuable or 5 = Extremely valuable

Question 12: What features would you consider MOST IMPORTANT in a CVD predictive analytics dashboard? (Rank top 5)

- Real time risk score calculation for individual patients
- Visual risk indicators (colour-coded: low or medium or high risk)
- Automated alerts for high-risk patients
- Trend analysis showing risk changes over time
- Integration with existing hospital systems (EHR)
- Population level dashboards (hospital wide statistics)
- AI-generated clinical recommendations
- Mobile / tablet accessibility
- Offline functionality for low-connectivity areas
- Report generation (PDF/Excel export)
- Multi language support

Question 13: How quickly should a risk assessment be generated to be useful in clinical practice?

- Instantly (under 1 second)
- Within 3 seconds
- Within 10 seconds
- Within 30 seconds
- Within 1 minute
- Longer than 1 minute is acceptable

Question 14: What type of risk visualisation would be most helpful for clinical decision making?

• Numerical risk score (0-100)

• Colour coded categories (Green/Yellow/Red)

• Risk percentage (% likelihood)

• Comparative charts (patient vs. population average)

• Trend graphs (risk over time)

• Simple text descriptions (Low or Medium or High)

• Other (please specify): ____________

Question 15: Who should receive automated alerts when a patient is identified

as high risk? (Select all that apply)

• Attending doctor only

• Nursing staff

• Department head

• Patient themselves

• Emergency response team

• Referral specialists (cardiologists)

• No automated alerts needed

SECTION D: SYSTEM FEATURE PREFERENCES

Question 16: How important is integration with existing hospital Electronic Health Record (EHR) systems?

Scale: 1 = Not important or 2 = Slightly important or 3 = Moderately important or 4 = Very important or 5 = Critical

Question 17: What devices would you primarily use to access the dashboard? (Select all that apply)

• Desktop computer (office/workstation)

• Laptop

• Tablet (iPad/Android tablet)

• Smartphone

• Shared hospital terminals

• Other (please specify): ____________

Question 18: How important is offline functionality (ability to use without internet connection) for your work environment?

Scale: 1 = Not important or 2 = Slightly important or 3 = Moderately important or 4 = Very important or 5 = Critical

Question 19: Which clinical data points do you think are MOST IMPORTANT for accurate CVD risk prediction? (Select top 5)

• Age

• Gender

• Blood pressure (systolic/diastolic)

• Total cholesterol level

• HDL/LDL cholesterol levels

• Body Mass Index (BMI)

• Smoking status

• Diabetes status

• Family history of CVD

• Physical activity level

• Alcohol consumption

• Stress levels

• Previous CVD events

• Medication history

• Other (please specify): ____________

Question 20: Would you trust AI-generated clinical recommendations provided by the system?

Scale: 1 = Completely distrust or 2 = Somewhat distrust or 3 = Neutral or 4 = Somewhat trust or 5 = Completely trust

Question 21: What would increase your trust in AI-generated recommendations? (Select all that apply)

• Explanation of how the recommendation was derived

• Confidence level displayed with each recommendation

• References to clinical guidelines

• Ability to override/modify recommendations

• Validation by local clinical experts

• Transparency about system limitations

• Regular accuracy reports

• Other (please specify): ___________

SECTION E: DATA PRIVACY AND SECURITY

Question 22: How concerned are you about patient data privacy in a

predictive analytics system?

Scale: 1 = Not concerned or 2 = Slightly concerned or 3 = Moderately concerned or 4 = Very concerned or 5 = Extremely concerned

Question 23: What security measures are most important to you (Select all that apply)

• Role based access (only authorised staff see data)

• Audit logs (track who accessed what)

• Data encryption

• Automatic logout after inactivity

• Two factor authentication

• Patient consent tracking

• Data anonymisation for research

• Compliance with POPIA (Protection of Personal Information Act)

• Other (please specify): ___________

SECTION F: OPEN FEEDBACK

Question 24: What other features or capabilities would you like to see in a CVD predictive analytics dashboard that we have not mentioned?

"Please share any additional ideas, concerns, or suggestions..."

Question 25: Do you have any concerns about implementing such a system in South African public hospitals?

"Please describe any potential barriers, challenges or concerns."

Link to the google form questions - https://forms.gle/iPcZLCyuyNa1oesq9

Link to excel file for result -

https://1drv.ms/x/c/f6ded36322dfce35/IQB1RWOCCXRcSJvIUzFz1EwSAQZ8J_YnPDRfu5lHoMiso2A?e=YYCGHy

WBS AND GANTT TEMPLATE

Project Title

Development of a Predictive Analytics Dashboard for Early Risk Detection of Cardiovascular Disease in South African Public Hospitals

WBS Description

The Work Breakdown Structure (WBS) outlines all activities required to design, develop and evaluate a predictive analytics dashboard for cardiovascular disease risk detection. It ensures alignment with the research objectives including data analysis, model development and dashboard implementation.

1. Project Initiation and Planning

1.1 Identify research problem

1.2 Define research topic

1.3 Develop research proposal

1.4 Define research aim and objectives

1.5 Formulate research questions

1.6 Obtain supervisor feedback

1.7 Finalise proposal

2. Literature Review

2.1 Review cardiovascular disease (CVD) burden in South Africa

2.2 Review predictive analytics in healthcare

2.3 Review machine learning models (RF, XGBoost and SVM)

2.4 Analyse limitations of descriptive analytics systems

2.5 Identify research gaps (lack of real-time dashboards)

2.6 Write literature review

2.7 Edit and refine

3. Research Design and Ethics

3.1 Select research methodology (mixed methods)

3.2 Identify dataset (Kaggle healthcare dataset)

3.3 Define variables (age, BP, cholesterol etc.)

3.4 Design questionnaire (user needs & usability)

3.5 Develop consent form

3.6 Complete ethical clearance

3.7 Submit for approval

4. Data Collection and Preparation

4.1 Distribute questionnaire to users

4.2 Collect questionnaire responses

4.3 Download cardiovascular dataset

4.4 Clean dataset (handle missing values)

4.5 Preprocess data (normalisation and scaling)

4.6 Prepare dataset for modelling

5. Machine Learning Model Development

5.1 Split dataset (training and testing)

5.2 Develop Logistic Regression model

5.3 Develop Random Forest model

5.4 Develop XGBoost model

5.5 Develop Support Vector Machine (SVM)

5.6 Evaluate models (accuracy and AUC)

5.7 Compare model performance

5.8 Select best model

6. Predictive Dashboard Development

6.1 Define dashboard requirements (clinical use)

6.2 Design dashboard interface (user friendly)

6.3 Develop input form (patient data entry)

6.4 Integrate ML model into system

6.5 Build dashboard (web app)

6.6 Create risk visualisations (low/medium/high)

6.7 Test dashboard functionality

6.8 Improve dashboard usability

7. System Evaluation

7.1 Conduct usability testing

7.2 Collect feedback (questionnaire or interviews)

7.3 Analyse qualitative responses

7.4 Evaluate system effectiveness

7.5 Interpret findings

8. Final Report and Submission

8.1 Write methodology

8.2 Write results and discussion

8.3 Link findings to research questions

8.4 Compile full report

8.5 Proofread and format

8.6 Submit final report

- GANTT CHART

Project Timeline: 12 Feb 2026 – 30 Nov 2026

Gantt Chart Link to ClickUp - https://app.clickup.com/90152511908/v/g/2kyr4ad4-1115

1. PROJECT INITIATION & PLANNING

Task	Start	End	Duration	Dependency
Identify research problem	16 Feb	27 Feb	2 days	-
Define research topic	16 Feb	27 Feb	2 days	Problem
Develop proposal	16 Feb	27 Feb	2 days	Topic
Define objectives & questions	27 Feb	08 Mar	2 days	Proposal
Supervisor feedback	12 Mar	12Mar	2 days	Proposal

Table 1: Project Initiation

Milestone: Proposal Complete 12 March 2026

2. LITERATURE REVIEW

Task	Start	End	Duration	Dependency
Review CVD studies	08 Apr	09 Apr	2 days	Proposal
Review predictive analytics	09 Apr	10 Apr	2 days	
Review ML models	10 Apr	11 Apr	2 days	
Identify research gaps	11 Apr	12 Apr	2 days	Reviews
Write literature review	12 Apr	13 Apr	2 days	Gaps
Edit & refine	13 Apr	14 Apr	2 days	Writing
Submission of Literature Review	14 April	14 April		Wait for feedback

Table 2: Literature review

Milestone: Literature Review Complete 14 April

3. RESEARCH DESIGN & ETHICS

Task	Start	End	Duration	Dependency
Select methodology	11 Apr	20 Apr	2 days	Literature
Identify dataset	12 Apr	20 Apr	2 days	Method
Define variables	13 Apr	20 Apr	2 days	Dataset
Design questionnaire	13 Apr	20 Apr	3 days	Method
Create consent form	14 Apr	20 Apr	2 days	Questionnaire
Ethical clearance	14 Apr	20 Apr	2 days	Consent

Table 3: Research design & Ethics

Milestone: Ethics Complete 20 April

4. DATA COLLECTION & PREPARATION (CURRENT PHASE)

Task	Start	End	Duration	Dependency
Distribute questionnaire	14 Apr	20 Apr	2 days	Ethics
Collect responses	15 Apr	20 Apr	5 days	Distribution
Download dataset	15 Apr	20 Apr	2 days	
Data cleaning	16 Apr	20 Apr	5 days	Dataset
Data preprocessing	18 Apr	20 Apr	3 days	Cleaning
Data collection submission	19 Apr	20 Apr	2 days	

Table 4: Data Collection & Preparation

Milestone: Data Collection Complete 20 April

5. MACHINE LEARNING MODEL DEVELOPMENT

Task	Start	End	Duration	Dependency
Split dataset	20 Jul	23Jul	2 days	Data Prep
Logistic Regression	24 Jul	5 Aug	3 days	Split
Random Forest	24 Jul	5 Aug	3 days	Split
XGBoost	20 Jul	5 Aug	3 days	Split
SVM	20 Jul	5 Aug	3 days	Split
Model evaluation	20 Jul	5 Aug	3 days	Models
Select best model	20 Jul	5 Aug	2 days	Evaluation
Test prototype	6 Aug	10 Aug	2 days	Build

Table 5: Machine Learning Model Development

Milestone: Model Complete 10 August

6. DASHBOARD PROTOTYPE

Task	Start	End	Duration	Dependency
Define requirements	28 Apr	20 Apr	2 days	Models
Design interface	29 Apr	20 Apr	2 days	Requirements
Create prototype dashboard	01 May	20 May	4 days	Design

Simulate predictions	02 May	20 May	4 days	Model
Create visualisations	04 May	20 May	3 days	Dashboard
Test prototype	20 Jul	Jul	2 days	Build

Table 6: Dashboard Prototype

Milestone: Prototype Complete 20 July

7. SYSTEM EVALUATION

Task	Start	End	Duration	Dependency
Usability testing	Aug	Aug	2 days	Prototype
Collect feedback	Aug	Aug	3 days	Testing
Analyse data	Aug	Aug	2 days	Feedback
Interpret results	Aug	Aug	2 days	Analysis

Table 7: System Evaluation

8. FINAL REPORT & SUBMISSION

Task	Start	End	Duration	Dependency
Write methodology	Nov	Nov	6 days	Data Prep
Write results & discussion	Nov	Nov	5 days	Evaluation
Compile report	Nov	Nov	2 days	Writing
Proofreading	Nov	Nov	2 days	Compile
Final submission	Nov	Nov	1 day	All

Table 8: Final Report & Submission

Final Milestone: Submission 30 November 2026

System design planning section

The proposed system is a Predictive Analytics Dashboard for Early Detection of Cardiovascular Disease (CVD) designed for use in South African public hospitals particularly in Bloemfontein. The system will primarily be used by doctors and nurses and to identify patients at high risk of developing cardiovascular conditions before complications occur. This directly addresses the research problem of delayed diagnosis and reliance on reactive healthcare approaches. The system is designed based on insights gathered from the questionnaire which highlighted key challenges such as lack of integrated patient data, absence of automated risk assessment tools and time constraints during consultations. These findings inform the need for a system that is fast, simple and clinically useful in real time environments.

The main feature of the system is a real time risk prediction tool. Healthcare professionals will input patient data such as age, blood pressure, cholesterol levels and diabetes status and the system will instantly generate a CVD risk score. This responds directly to user feedback indicating that risk assessments must be produced within seconds to be useful in practice. The system will also include colour coded risk indicators (low, medium, high) making it easier for doctors to interpret results quickly during patient consultations.

Another key feature is automated alerts for high risk patients ensuring that critical cases are not missed. This aligns with questionnaire responses showing that many high risk patients are currently identified too late. The dashboard will also provide visual summaries such as charts showing risk distribution and key contributing factors helping healthcare staff make informed decisions. To improve usability and trust the system will include explainable AI features showing why a patient is classified as high risk. This directly addresses user concerns about trusting AI generated recommendations. Additionally the system will support mobile and tablet access and reflect user needs in environments with limited connectivity.

In conclusion the proposed system connects the research problem (late detection of CVD) user needs (speed, clarity, integration and trust), and the solution (a predictive user friendly dashboard). By grounding the design in real questionnaire data the system ensures practical relevance and supports improved clinical decision making in South African public hospitals.

USE CASE FOR CVD

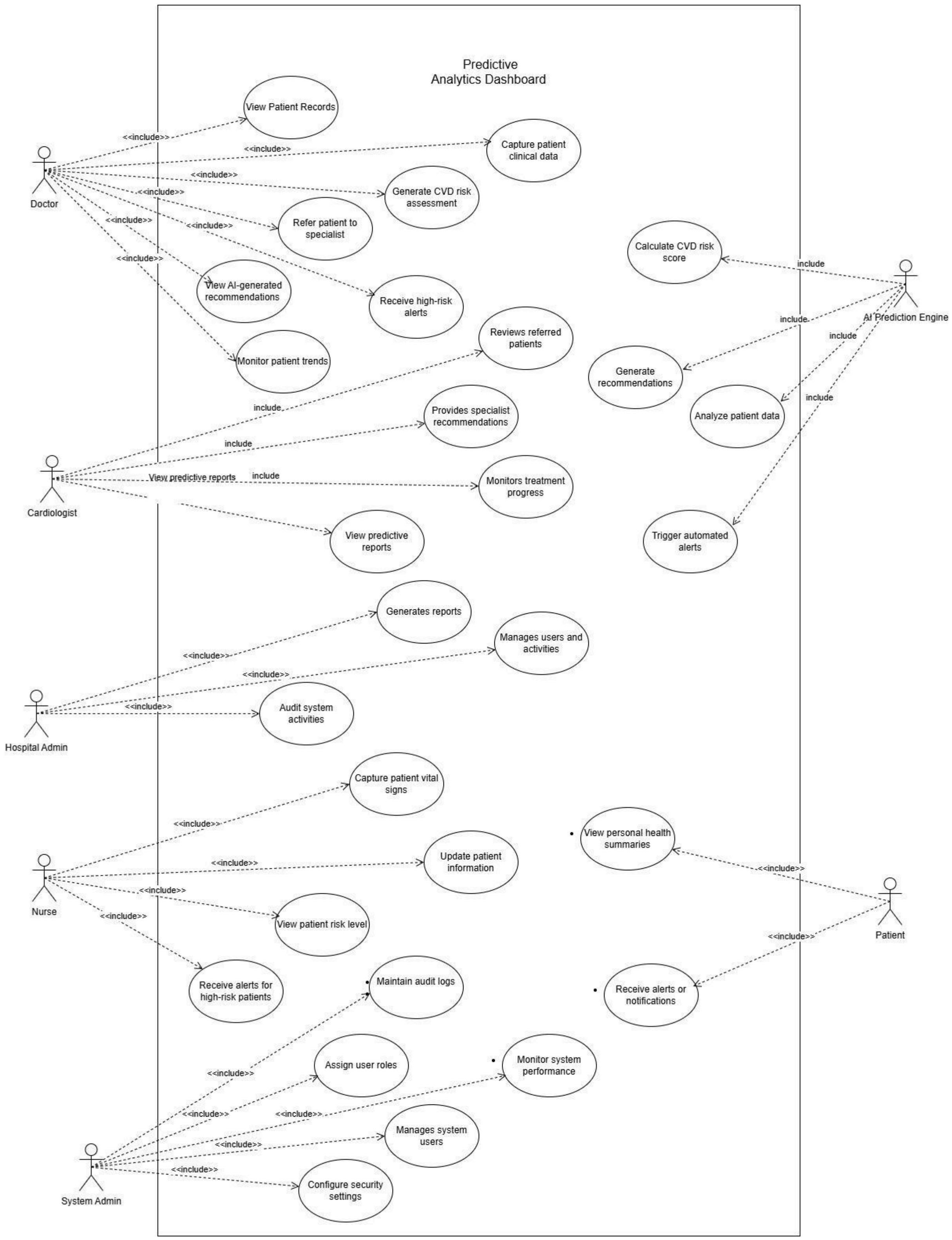

Figure 1 use case and actors of CVD

Use Case Documentation: Predictive Analytics Dashboard for CVD Risk Detection

SYSTEM OVERVIEW

The Predictive Analytics Dashboard helps South African public hospitals detect cardiovascular disease (CVD) early. Different users interact with the system based on their roles. The AI Prediction Engine works in the background to analyse data, calculate risk scores and trigger alerts.

Actors and Their Use Cases

1. Doctor

The doctor is the main clinical user. The doctor performs these actions:

View Patient Records – includes accessing full patient history.

Capture patient clinical data – includes entering lab results, symptoms, and risk factors.

Generate CVD risk assessment – includes requesting a risk evaluation.

Referring to a specialist – includes sending high-risk patients to a cardiologist.

Receive high-risk alerts – includes getting notifications when a patient's risk level rises.

View AI-generated recommendations – includes reading suggested next steps from the AI.

Monitor patient trends – includes tracking changes in risk over time.

2. Cardiologist

The cardiologist handles referred cases and long term care. Actions:

Reviews referred patients – includes examining cases sent by doctors.

Provides specialist recommendations – includes giving expert advice on treatment.

Monitor treatment progress – includes tracking patient outcomes after referral.

3. Hospital Admin

The hospital admin manages the system from an operational view.

Actions:

Generates reports – includes creating summary and performance reports.

Manages users and activities – includes overseeing who uses the system and what they do.

Audit system activities – includes reviewing logs for compliance and security.

4. Nurse

Nurses support doctors by collecting and updating patient data. Actions:

Capture patient vital signs – includes recording blood pressure, heart rate, etc.

Update patient information – includes editing demographics or contact details.

View patient risk level – includes checking a patient's current CVD risk category.

Receive alerts for high risk patients – includes getting notifications to prioritise care.

5. System Admin

The system admin maintains the technical environment.

Actions:

Maintain audit logs – includes ensuring all system events are recorded.

Assign user roles – includes granting permissions to doctors and nurses.

Monitor system performance – includes checking uptime, response speed and errors.

Manages system users – includes adding editing or removing user accounts.

Configure security settings – includes setting passwords, encryption and access controls.

6. AI Prediction Engine

The AI engine runs in the background without direct user interaction. Actions:

Calculate CVD risk score – includes using machine learning models on patient data.

Generate recommendations – includes producing clinical suggestions based on risk.

Analyse patient data – includes processing clinical records, vitals and history.

Trigger automated alerts – includes sending notifications to doctors and nurses.

7. Patient

Patients receive information from the system.

Actions:

View personal health summaries – includes seeing their own risk level and trends.

Receive alerts or notifications – includes getting messages about high risk status or follow up if needed.

ENTITY RELATION DIAGRAM FOR CVD

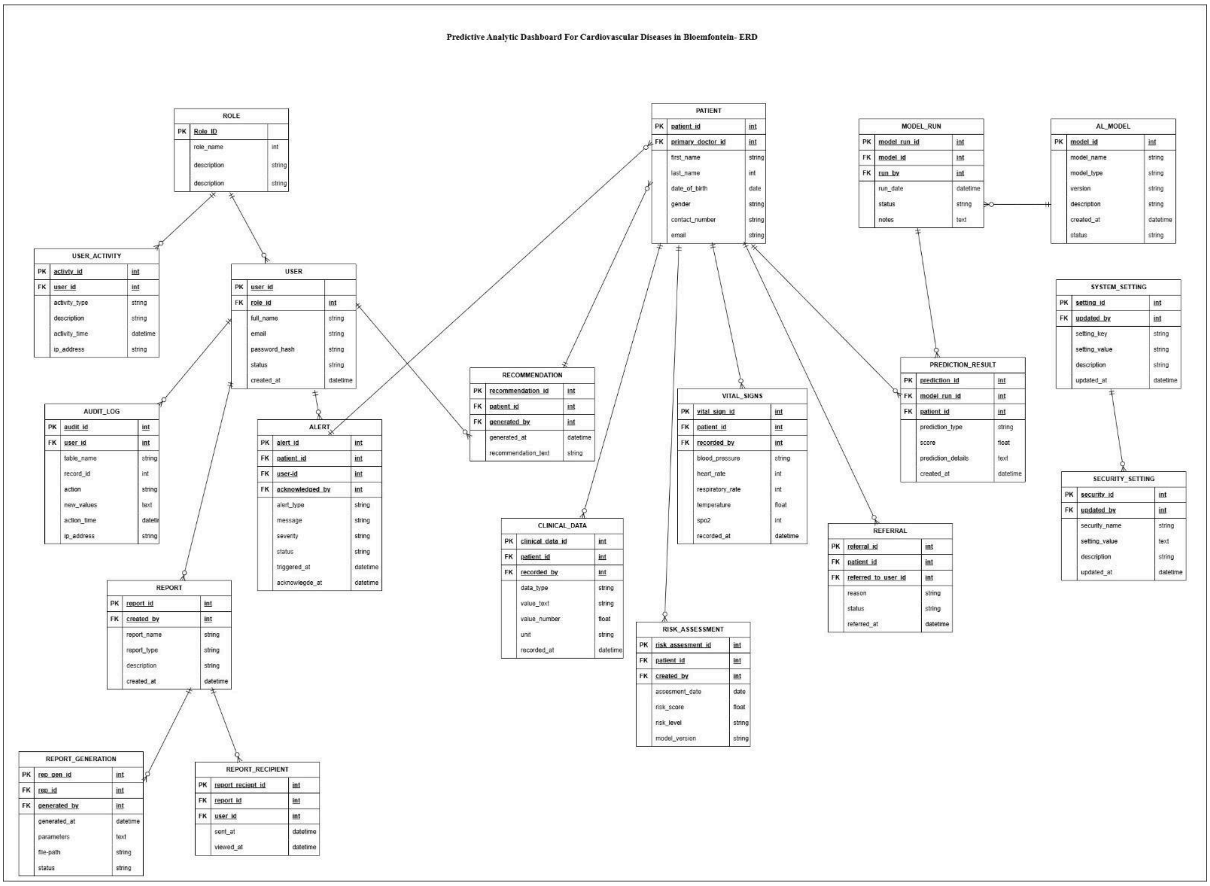

Figure 2 ERD of CVD

Explanation of the Entity-Relationship Diagram (ERD) for the Predictive Analytics Dashboard

The Entity Relationship Diagram (ERD) is a visual blueprint of the database designed for the Predictive Analytics Dashboard for Cardiovascular Diseases in Bloemfontein. It shows how different pieces of information relate to one another inside the system, it is like a detailed filing system where each drawer has a specific label and the drawers are connected where necessary.

---

1. Users and Their Roles

The diagram recognizes that different people use the dashboard doctors, cardiologists, nurses and administrators. These users are assigned specific roles (e.g., "Doctor" or "System Admin") which determine what they can see and do within the system. For example, a nurse may record vital signs, while a cardiologist views predictive reports.

---

2. Patient Data and Clinical Information

Each patient has their own record containing personal details, clinical measurements (such as blood pressure and cholesterol levels) and vital signs (such as heart rate and temperature). This information is not stored randomly; rather every measurement is linked to the correct patient and the correct healthcare worker who recorded it. This ensures that data remains accurate and traceable.

---

3. Risk Assessment and Predictions

One of the main goals of the dashboard is to predict a patient's risk of developing cardiovascular disease. The ERD shows how risk assessments are linked to patients and how AI models are used to generate those predictions. Every time a prediction is made the system records which AI model was used and when it was run. This makes the prediction process transparent and repeatable.

4. Alerts, Referrals and Recommendations

When a patient is identified as high risk, the system can generate an alert for the doctor, create a referral to a specialist (such as a cardiologist) or produce a recommendation for treatment. The ERD clearly shows how alerts, referrals and recommendations are each connected to both the patient and the responsible healthcare professional. This helps ensure that no critical information is lost or ignored.

5. Reports and Accountability

The dashboard allows users to generate reports for review, research or patient management. Every report generation is logged and the system records who received the report and whether it was viewed. In addition the system keeps an audit log of significant actions (such as when a record is changed) and a user activity log of general system use. These features support accountability, data security and regulatory compliance.

6. System and Security Settings

The ERD includes tables that store system settings and security settings. These control how the dashboard behaves (for example, which AI model is active) and who has access to sensitive information. Only authorized system administrators can modify these settings.

CONCLUSION- First part

This research proposal has presented the design and preliminary development of a predictive analytics dashboard for early cardiovascular disease risk detection in South African public hospitals with specific focus on the Bloemfontein national district hospital . The study addresses a critical healthcare challenge where cardiovascular disease accounts for approximately 17.3% of deaths in South Africa yet nearly 20% of high risk patients remain undetected until complications arise. The reliance on descriptive analytics in existing healthcare information systems creates a reactive environment that delays clinical interventions and compromises patient outcomes.

The proposed solution integrates four machine learning algorithms Logistic Regression, Random Forest, XGBoost and Support Vector Machine to transform raw clinical data into actionable cardiovascular risk profiles. The concurrent mixed methods research design combines quantitative model validation with qualitative feedback from healthcare professionals ensuring that the dashboard meets real world clinical requirements. Key system features include real time risk scoring, colour coded visual indicators, automated alerts for high risk patients, explainable AI components to build clinical trust and offline functionality to accommodate resource constrained environments with limited connectivity.

The Work Breakdown Structure and Gantt Chart demonstrate a realistic project timeline from February to November 2026 aligning technical development phases with academic milestones. The Use Case Diagram illustrates how doctors, nurses, system administrators and hospital managers interact with the system across nine core functionalities. The system architecture employs a three layer design presentation, application and data supported by security protocols compliant with the Protection of Personal Information Act.

The prototype screenshots demonstrate a user friendly interface that addresses questionnaire feedback regarding the need for speed, clarity and integration with existing workflows. The dashboard generates risk assessments within seconds, displays top contributing factors through intuitive visualisations and provides AI generated clinical recommendations with transparent explanations. The offline mode ensures continuity of care in environments with intermittent connectivity, a feature identified as critical by 70.6% of questionnaire respondents.

This research contributes to addressing three identified gaps in existing literature: the contextual gap of South African population specific models, the implementation gap between predictive analytics research and clinical practice, and the system gap of integrated, user friendly dashboards. By grounding the design in empirical data from healthcare professionals and focusing on resource appropriate technology this study offers a practical pathway toward proactive cardiovascular care in public hospital settings. Future work will focus on external validation with larger more diverse datasets, integration with live hospital EHR systems and longitudinal assessment of clinical impact on patient outcomes.

Display of prototype

Figma Prototype Link -

https://www.figma.com/design/nw5sYa2BmTmZogPhqVYtXP/Untitled?node-id=0-1&t=tDMYi82YhBdnySC4-1

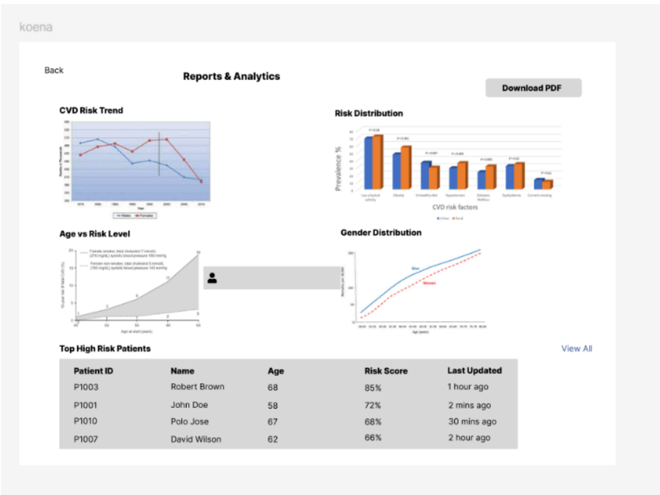

Figure 3 Report and Analytics

These diagram shows patient's clinical data which is used to predict the analytics

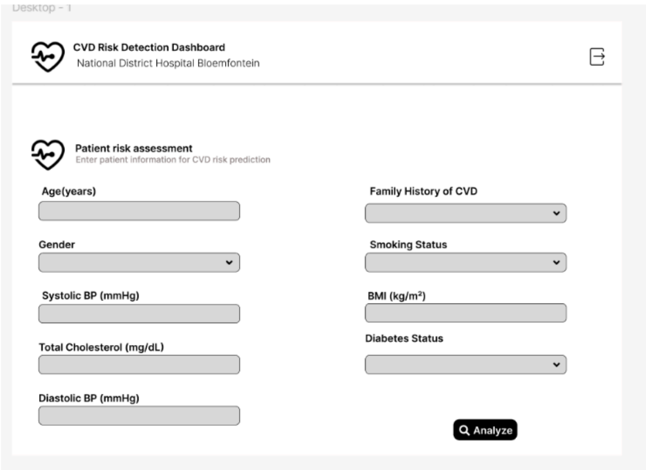

Figure 4 CVD Risk Detection Dashboard

This diagram shows what patient data is captured and needed for risk detection assessment

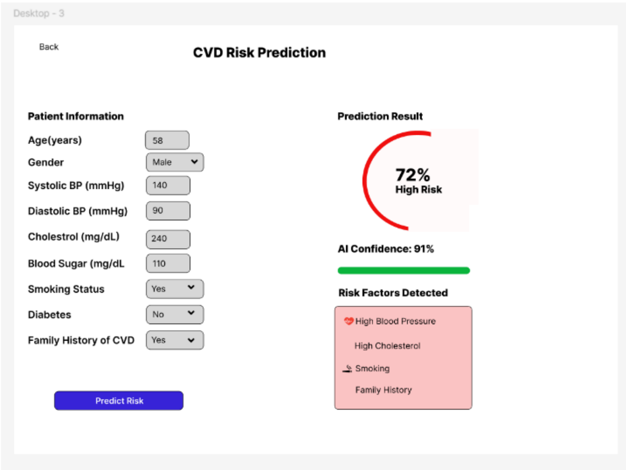

Figure 5 CVD Risk Prediction

This diagram shows the results of CVD detected using patient clinical data

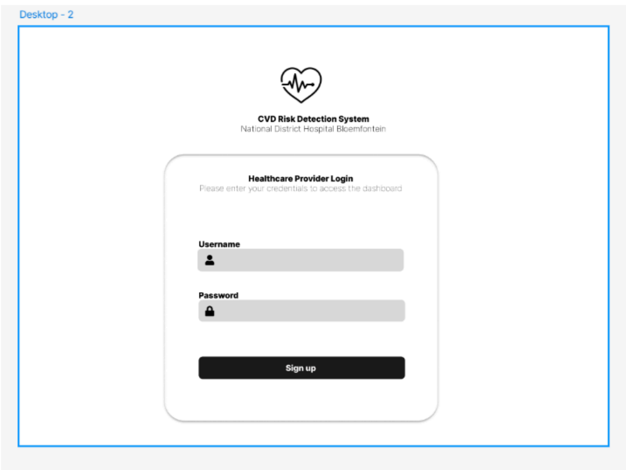

Figure 6 CVD Risk Detection System Login screen

This is for user authentication and authorization to identify who should access the system

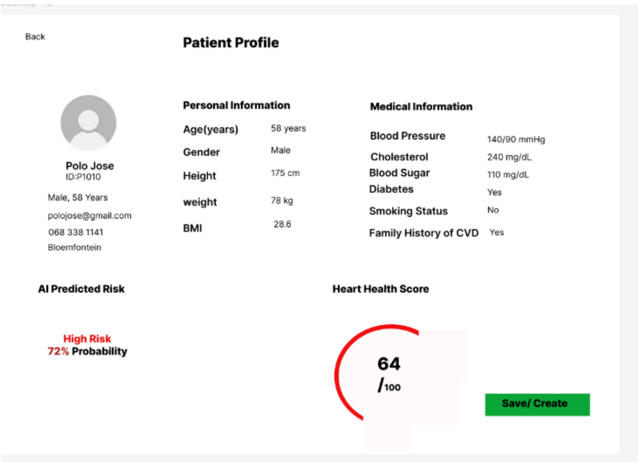

Figure 7 Patient Profile

This is where the patients' information can be accessed and modified.

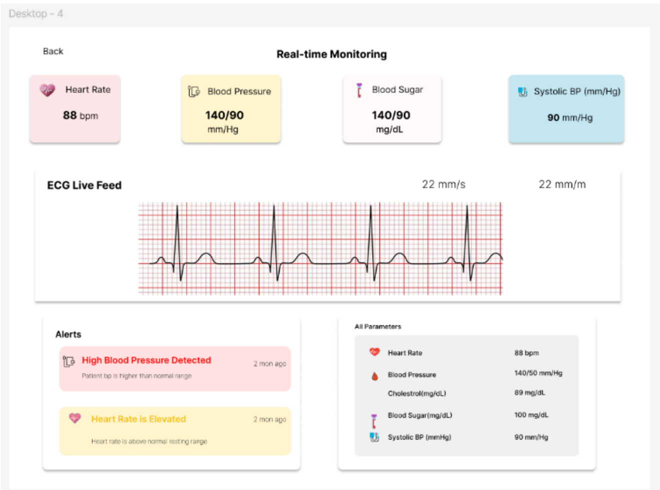

Figure 8 Real-time Monitoring

This shows the real-time monitoring of patients' risk assessment records and alert status