UTM UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING
UTM Johor Bahru

SECB3203-01(PROGRAMMING FOR BIOINFORMATIC)

Semester 01 2025/2026
Section 01

Progress 1

Faculty of Computing

Dataset:
https://www.kaggle.com/datasets/miadul/tuberculosis-x-ray-dataset-synthetic

Project Title:
Tuberculosis Disease Classification Using Synthetic Chest X-Ray Images and Machine Learning Techniques

---

Student Name	Matric Number
NATIJAH BINTI HUDA	A23CS0142
NUR IMAN BINTI MOHAMAD ZAHARI	A23CS0158
LIANA DARWISYAH BINTI AZMAN	A23CS0102

LECTURER'S NAME : DR. SEAH CHOON SEN

SUBMISSION DATE :

Table of contents

1. Software and Hardware Requirements

1.1 Software Requirements

The following software will be used to develop and implement the tuberculosis classification system:

- Operating System:
  Windows 10 / Windows 11
- Programming Language:
  Python 3.x
- Development Environment:
  - Jupyter Notebook
  - Anaconda Navigator
- Python Libraries:
  - NumPy – numerical computations
  - Pandas – data manipulation and preprocessing
  - Matplotlib – data visualization
  - Scikit-learn – machine learning models and evaluation
  - OpenCV / PIL – image processing
  - TensorFlow / Keras (if deep learning is applied)
- Version Control:
  - GitHub for code management and documentation
- Dataset Source:
  - Synthetic Tuberculosis Chest X-Ray Dataset from Kaggle

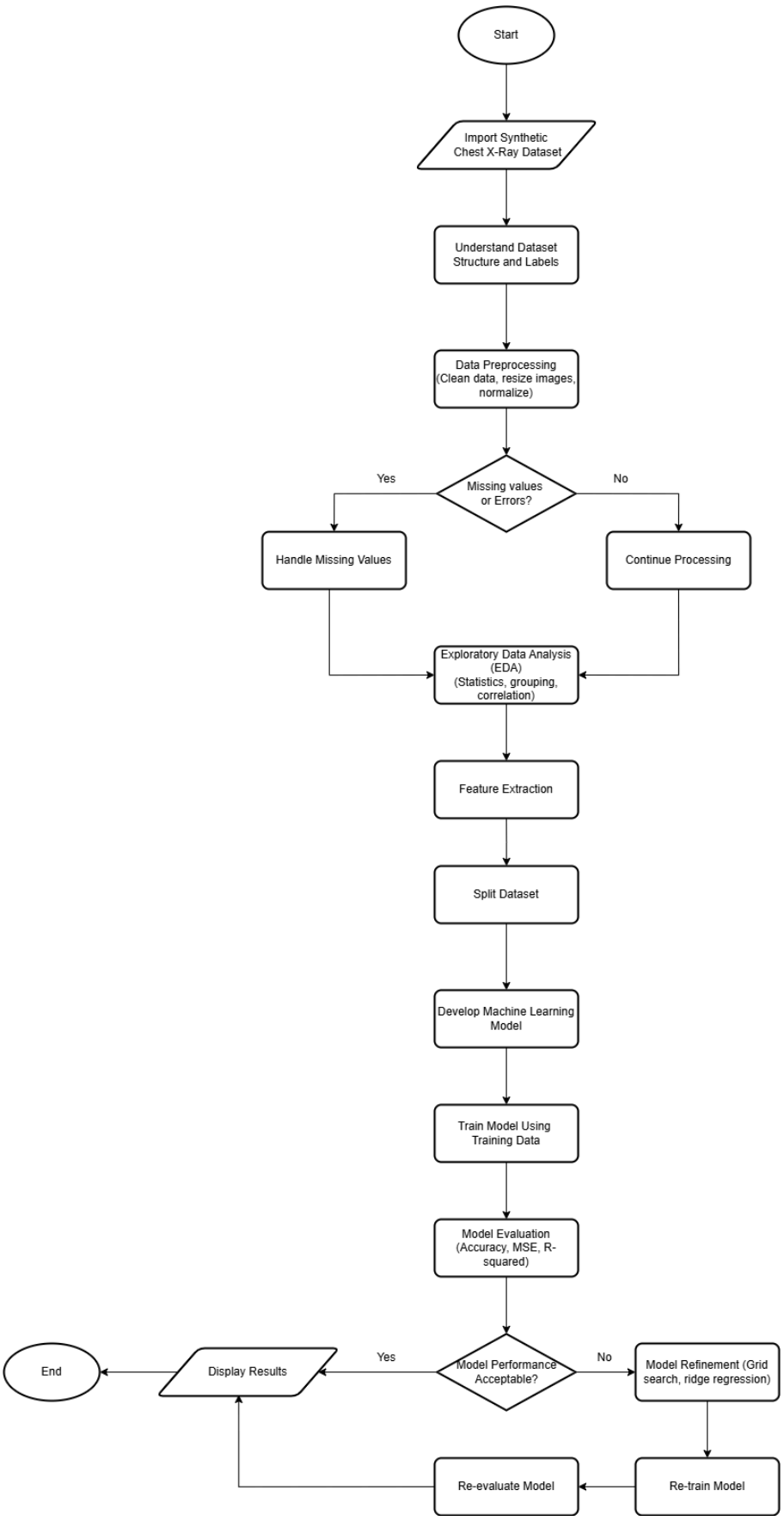
1.2 Hardware Requirements

The project will be implemented using student-level computing resources:

- Processor:
  Intel Core i5 or equivalent
- RAM:
  Minimum 8 GB
- Storage:
  At least 20 GB free disk space
- Display:
  Standard monitor with minimum 1366 × 768 resolution
- Internet Connection:
  Required for dataset download, library installation, and GitHub access

2. Flowchart of the Proposed Approach

The flowchart below shows the overall workflow of the project from data collection to model evaluation.


Start
Import Synthetic Chest X-Ray Dataset
Understand Dataset Structure and Labels
Data Preprocessing (Clean data, resize images, normalize)
Missing values or Errors?
Yes
No
Handle Missing Values
Continue Processing
Exploratory Data Analysis (EDA) (Statistics, grouping, correlation)
Feature Extraction
Split Dataset
Develop Machine Learning Model
Train Model Using Training Data
Model Evaluation (Accuracy, MSE, R-squared)
Model Performance Acceptable?
Yes
No
End
Display Results
Model Refinement (Grid search, ridge regression)
Re-train Model
Re-evaluate Model