UTM UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING
UTM Johor Bahru

SECB3203-01(PROGRAMMING FOR BIOINFORMATIC)

Semester 01 2025/2026
Section 01

Progress 5

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 :

5.0 Model Evaluation

In this phase, the trained machine learning classification models are evaluated to determine their performance in classifying tuberculosis cases. Model evaluation is an important step to assess how well each model generalizes to unseen data. The evaluation process is conducted using the scikit-learn library.

5.1 Evaluation Metrics

Several evaluation metrics are used to measure the performance of the classification models:

- Accuracy: Measures the overall correctness of the model predictions.
- Precision: Indicates how many predicted positive cases are actually positive.
- Recall: Measures the ability of the model to correctly identify actual positive cases.
- F1-score: Represents the harmonic mean of precision and recall.

Since the dataset is imbalanced, accuracy alone is not sufficient. Therefore, F1-score is emphasized as the primary metric for model comparison.

5.2 Model Performance Comparison

The performance of multiple classification models is compared using the selected evaluation metrics. The models evaluated include:

- Logistic Regression
- Decision Tree
- Random Forest
- Naive Bayes

The comparison results show that several models achieve high accuracy but fail to detect the minority class, resulting in low recall and F1-score values.

5.3 Overfitting and Underfitting Analysis

An analysis of overfitting and underfitting is conducted based on the performance results:

- Models with high accuracy but low recall are identified as underfitting the minority class.
- Models with balanced precision and recall demonstrate better generalization ability.

This analysis helps in understanding the strengths and limitations of each model.

5.4 Model Selection

Based on the evaluation metrics, particularly the F1-score, the Decision Tree model is selected as the best-performing model. Although its accuracy is lower compared to

other models, it provides a better balance between precision and recall, making it more suitable for the imbalanced dataset.

5.5 Model Refinement and Discussion

The selected model is further analyzed to understand its performance and behavior. Model refinement is discussed in terms of potential improvements such as feature engineering, handling class imbalance, and applying advanced tuning techniques. These refinements can further improve model performance in future work.

5.6 Summary of Evaluation Results

In summary, the evaluation phase highlights the importance of using appropriate metrics for imbalanced datasets. The Decision Tree model demonstrates the most reliable performance among the evaluated models, making it the preferred choice for tuberculosis classification in this study.