

## **Assignment 3 – Detailed Marking Rubric (Total = 30 marks)**

*(Tasks 1–6, each 5 marks)*

---

### **Task 1 – Data Preparation & Pre-processing (5 marks)**

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Excellent</b>	Correctly imports all required libraries; correctly loads image dataset; clearly separates input vs target; correct train-test split; applies appropriate scaling/pre-processing (e.g., normalization); code is clean, organised, and reproducible.	<b>5</b>
<b>Proficient</b>	Minor issues in loading or splitting but overall correct; scaling/pre-processing applied but not fully justified; code runs with minimal adjustment.	<b>4</b>
<b>Satisfactory</b>	Basic loading and splitting done; some missing steps OR inconsistent scaling; code requires fixes but structure is understandable.	<b>3</b>
<b>Weak</b>	Incorrect or incomplete dataset loading; train-test split flawed; missing scaling when required; unclear or disorganised code.	<b>1-2</b>
<b>Poor / No submission</b>	Minimal or no attempt.	<b>0</b>

---

### **Task 2 – Model Training, Comparison & Evaluation (5 marks)**

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Excellent</b>	Trains <b>three ML/DL models</b> correctly; clearly compares them; uses F1-score appropriately; provides evaluation tables/plots; explanation is technically sound.	<b>5</b>
<b>Proficient</b>	Three models trained but comparison lacks depth OR minor mistakes in F1-score use; evaluation is present but not well explained.	<b>4</b>
<b>Satisfactory</b>	At least two models trained properly; F1-score computed but misinterpreted; limited comparison.	<b>3</b>

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Weak</b>	One model only OR major errors in training or metric computation; no meaningful comparison.	<b>1–2</b>
<b>Poor / No submission</b>	No valid model training.	<b>0</b>

---

### **Task 3 – XAI Method Implementation (GradCAM / RISE / LIME) (5 marks)**

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Excellent</b>	Correctly applies a <b>suitable</b> XAI method; provides clear, high-quality visual explanations (e.g., GradCAM heatmaps); implementation is correct and linked to the chosen model.	<b>5</b>
<b>Proficient</b>	Method used correctly but visuals or code explanation could be clearer; minor technical inaccuracies.	<b>4</b>
<b>Satisfactory</b>	XAI method implemented but visuals weak OR incomplete; interpretation limited.	<b>3</b>
<b>Weak</b>	Incorrect XAI method use OR output not meaningful.	<b>1–2</b>
<b>Poor / No submission</b>	No XAI attempt.	<b>0</b>

---

### **Task 4 – Written XAI Explanation for Task 3 (5 marks)**

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Excellent</b>	Insightful explanation of how the model made decisions; links heatmaps/features to pneumonia pathology; critical reflection on model reliability; excellent clarity.	<b>5</b>
<b>Proficient</b>	Good explanation but lacks depth OR minor clinical or technical misunderstandings.	<b>4</b>
<b>Satisfactory</b>	Describes results but not the underlying reasoning; limited critical evaluation.	<b>3</b>
<b>Weak</b>	Very general explanation; misinterprets the visuals.	<b>1–2</b>

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Poor / No submission</b>	No explanation.	<b>0</b>

---

#### **Task 5 – Second XAI Method Implementation (Different from Task 3) (5 marks)**

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Excellent</b>	Chooses and implements a <b>different XAI method</b> correctly; produces clear visual or textual outputs; compares suitability relative to Task 3.	<b>5</b>
<b>Proficient</b>	Correct method applied but comparison or visuals could be improved.	<b>4</b>
<b>Satisfactory</b>	Method implemented but limited quality or weak explanation.	<b>3</b>
<b>Weak</b>	Incorrect method, poor implementation, or low-quality outputs.	<b>1-2</b>
<b>Poor / No submission</b>	No second XAI implementation.	<b>0</b>

---

#### **Task 6 – Written Explanation for Second XAI Method (5 marks)**

<b>Level</b>	<b>Description</b>	<b>Marks</b>
<b>Excellent</b>	Clear interpretation of second XAI output; meaningful comparison with Task 3 results; identifies strengths/weaknesses; shows critical thinking about model trustworthiness.	<b>5</b>
<b>Proficient</b>	Good explanation but missing deeper comparison or critique.	<b>4</b>
<b>Satisfactory</b>	Basic interpretation; lacks depth or analytical insight.	<b>3</b>
<b>Weak</b>	Misinterprets results OR gives minimal explanation.	<b>1-2</b>
<b>Poor / No submission</b>	No explanation.	<b>0</b>

---

**Total Marks = 30 (Convert to 20%)**

You can scale:

**Final Score (%) = (Total Marks / 30) × 20**