**Project Development Phase**

**Model Performance Test**

|  |  |
| --- | --- |
| Date | 10 February 2025 |
| Team ID | LTVIP2025TMID38625 |
| Project Name | revolutionizing liver care |
| Maximum Marks | 10 Marks |

## 🚧 **Project Development Phase: Model Performance Testing in Liver Care**

### 🎯 **Purpose**

This phase ensures that the **machine learning (ML) or AI model** developed during earlier stages performs **accurately, reliably, and ethically** on liver-related clinical data before being deployed in real-world healthcare settings.

## 🔍 **Why It’s Critical in Liver Care**

In liver care, the consequences of a model's error could be life-threatening. Testing ensures:

* Early and accurate diagnosis (e.g., fibrosis staging)
* Reliable outcome predictions (e.g., risk of liver failure)
* Safe personalization of diet/medication
* Fairness across populations (age, gender, ethnicity)

## 🧩 **Outputs of This Phase**

* **Performance report** (metrics with visualizations)
* **Validation summary** (internal + external)
* **Model audit log** (versioning, bias findings)
* **Go/No-Go decision** for next phase (deployment or re-training)

## ✅ **Best Practices**

* Involve clinicians during evaluation to validate model outputs
* Use real-world, diverse datasets to avoid bias

**Model Performance Testing:**

| **Category** | **Description** |
| --- | --- |
| **Phase Objective** | To evaluate the AI/ML model's accuracy, reliability, fairness, and clinical utility in liver care |
| **Model Purpose** | - Disease detection (e.g., NAFLD, Hepatitis B/C, Cirrhosis) - Risk prediction - Personalized treatment support |
| **Input Data** | - Liver function tests (ALT, AST, Bilirubin) - Imaging (Ultrasound, MRI) - Patient demographics - Lifestyle/diet logs |
| **Key Performance Metrics** | - Accuracy - Precision - Recall (Sensitivity) - F1 Score - AUROC - Mean Absolute Error (MAE) / MSE (if regression) |
| **Validation Techniques** | - Train/Test/Validation split - K-fold Cross-Validation - External Dataset Testing |
| **Bias & Fairness Checks** | - Gender, age, and ethnicity-based performance analysis |
| **Robustness Testing** | - Model tested with missing, noisy, or outlier values |
| **Interpretability Tools** | - SHAP (SHapley Additive exPlanations) - LIME (Local Interpretable Model-agnostic Explanations) |
| **Testing Tools Used** | - Scikit-learn, TensorFlow, PyTorch - Jupyter, Google Colab - Matplotlib/Seaborn for visualization |
| **Output Deliverables** | - Model Performance Report - Confusion Matrix - Bias & Interpretability Report |
| **Review Team** | - Data Scientists - Hepatologists - Clinical Researchers - Regulatory Experts |
| **Decision Point** | - Proceed to deployment - Re-train model - Collect more data for better generalization |