SYSTEM ANALYSIS & DESIGN DOCUMENTATION

Breast Cancer Molecular Subtype and Survival Prediction System

This document presents the Unified Modeling Language (UML) design for the Breast Cancer Molecular Subtype and Survival Prediction System.

The system is designed as a single-user clinical decision-support prototype that predicts:

- Breast cancer molecular subtype
- Binary survival outcome
- Multi-class vital status

Using structured clinical and pathological data.

Unlike full hospital information systems, this prototype does not include authentication or multi-user access control. The system assumes a single authorized clinical user. This design decision aligns with the primary objective of the project, which is to develop and evaluate the predictive modeling framework rather than implement a complete hospital management system.

The system aims to provide a cost-effective computational alternative to genomic testing methods such as the PAM50 assay, particularly in resource-limited environments.

Scope of the UML Design

The UML documentation models:

- System functionality
- Internal structure
- Behavioral workflow
- Model training process
- Deployment structure

Use Case Modeling

Use Case Overview

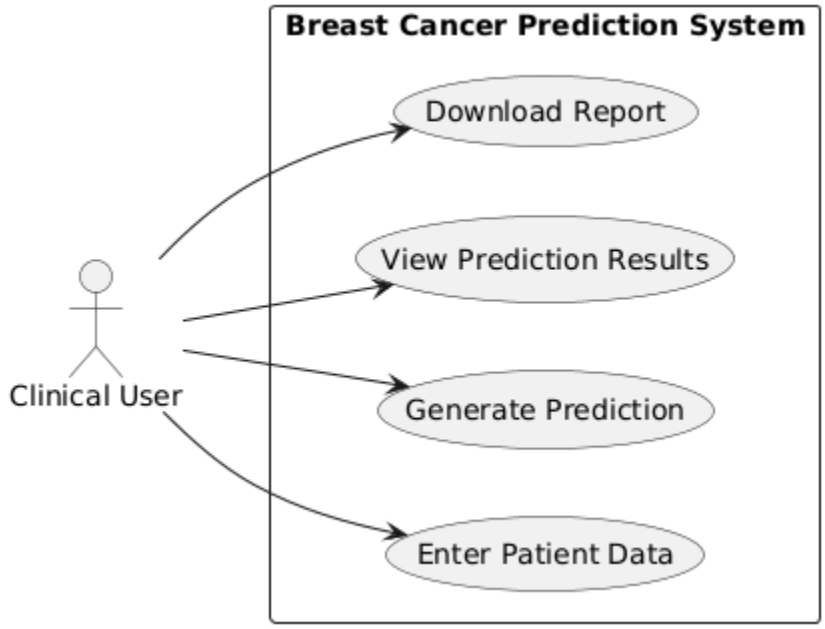
Since the system assumes a single user type, the only external actor is:

- Clinical User

The Clinical User can:

- Enter patient clinical data
- Generate predictions
- View results
- Download prediction reports

Use Case Diagram (PlantUML Representation)



Use Case Explanation

The Clinical User inputs structured patient data such as:

- Age
- Tumor size
- Tumor stage
- Lymph node involvement
- ER/PR/HER2 receptor status
- Treatment information

The system processes the data through a trained machine learning model and returns probabilistic predictions for:

- Molecular subtype
- Survival status
- Vital status classification

Structural Modeling

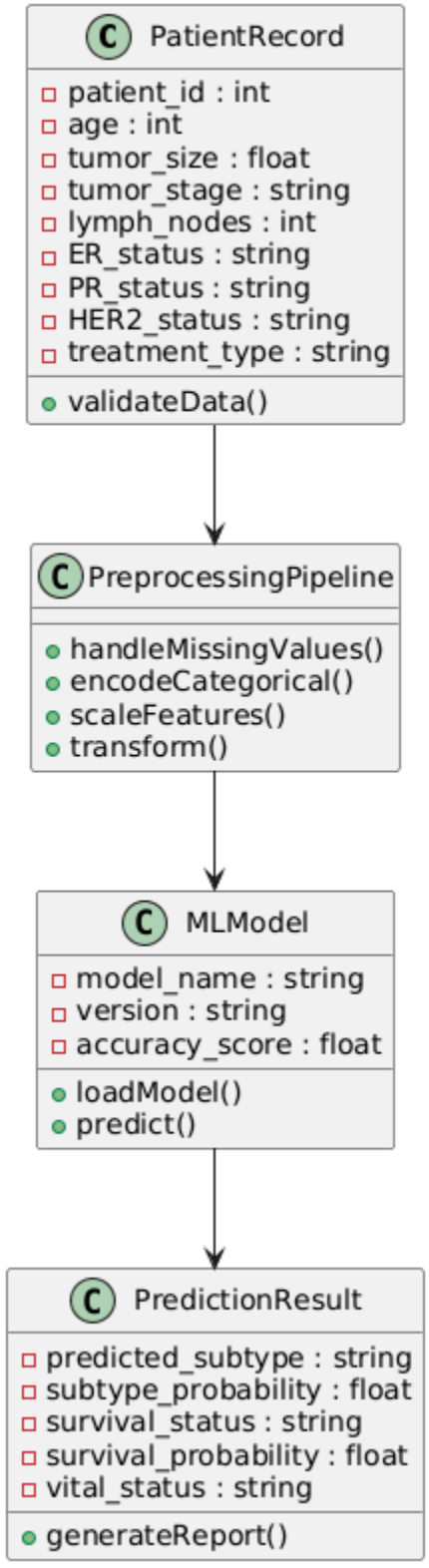
Class Diagram

The Class Diagram illustrates the static structure of the system.

Main Classes:

- PatientRecord
- PredictionResult
- MLModel
- PreprocessingPipeline

Class Diagram (PlantUML Representation)



Structural Interpretation

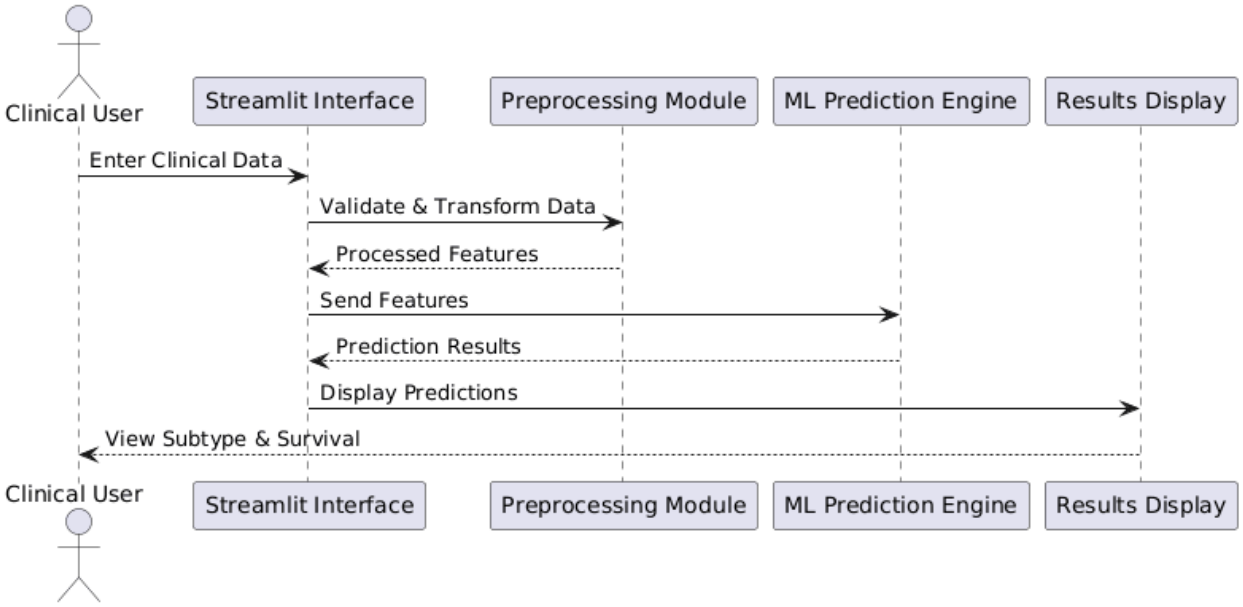
A PatientRecord contains clinical attributes.

- The PreprocessingPipeline transforms raw data into model-ready features.
- The MLModel generates predictions.
- The PredictionResult stores and formats outputs for display.
- This modular structure improves maintainability and reproducibility.

Behavioral Modeling

Sequence Diagram – Prediction Generation

This diagram illustrates the interaction flow when generating a prediction.

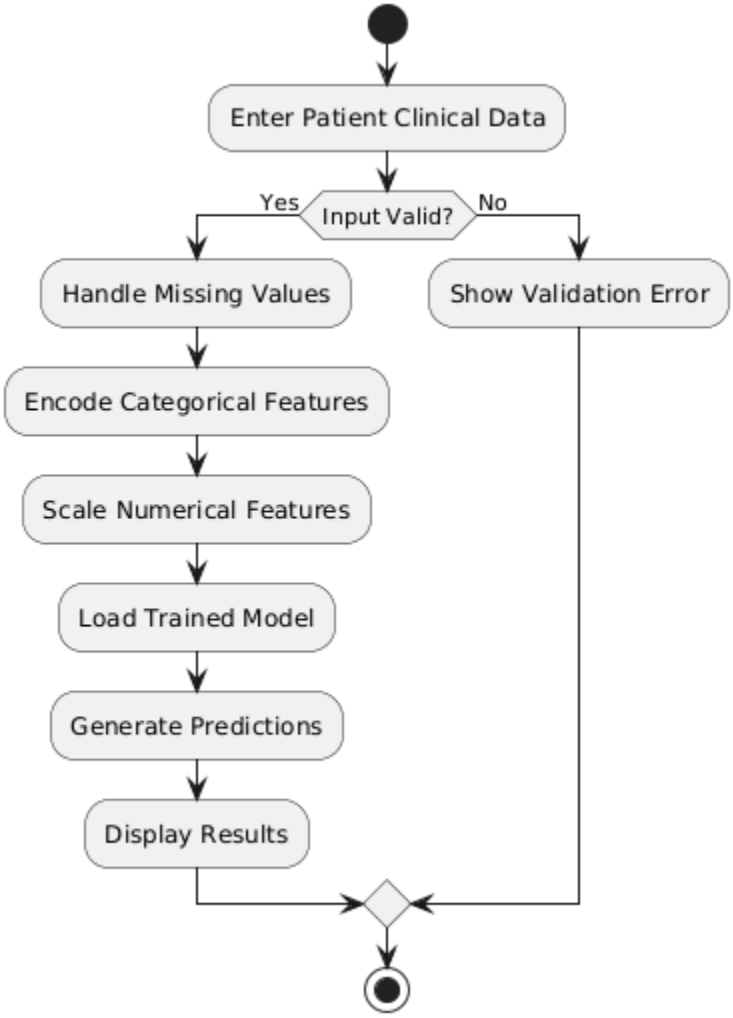


Explanation;

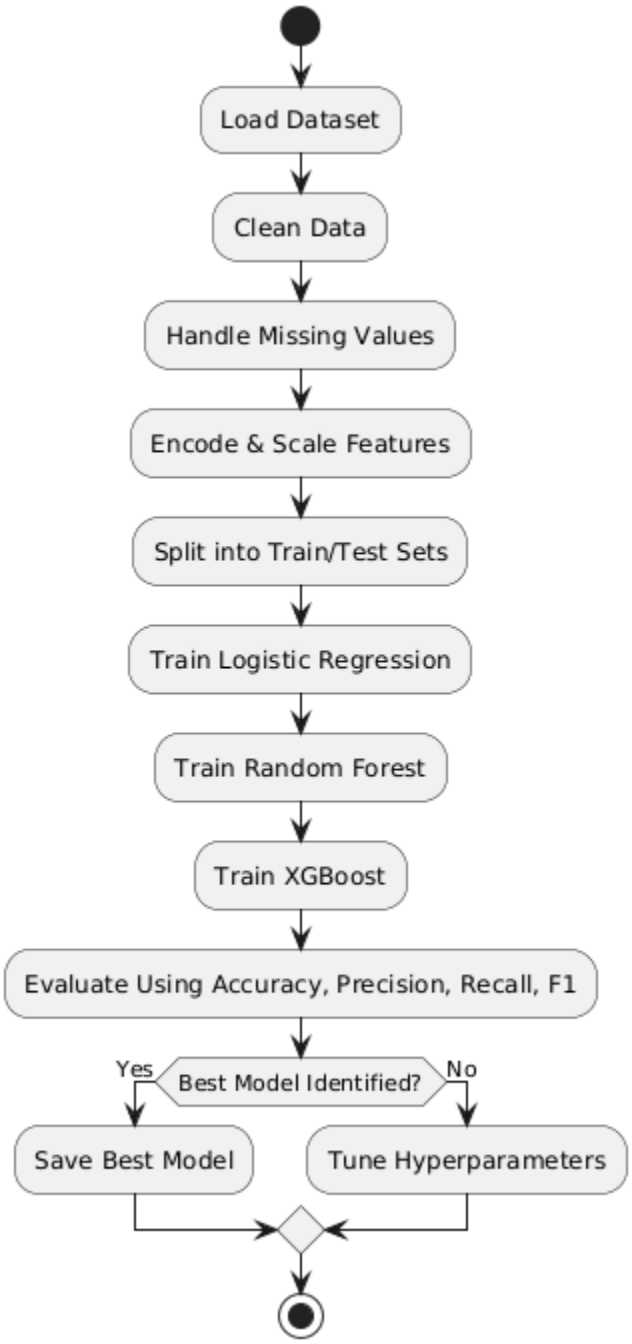
a) The user inputs clinical data.
b) The system preprocesses the data.
c) The trained ML model generates predictions.
d) Results are displayed immediately.

Activity Modeling

Activity Diagram – Prediction Workflow

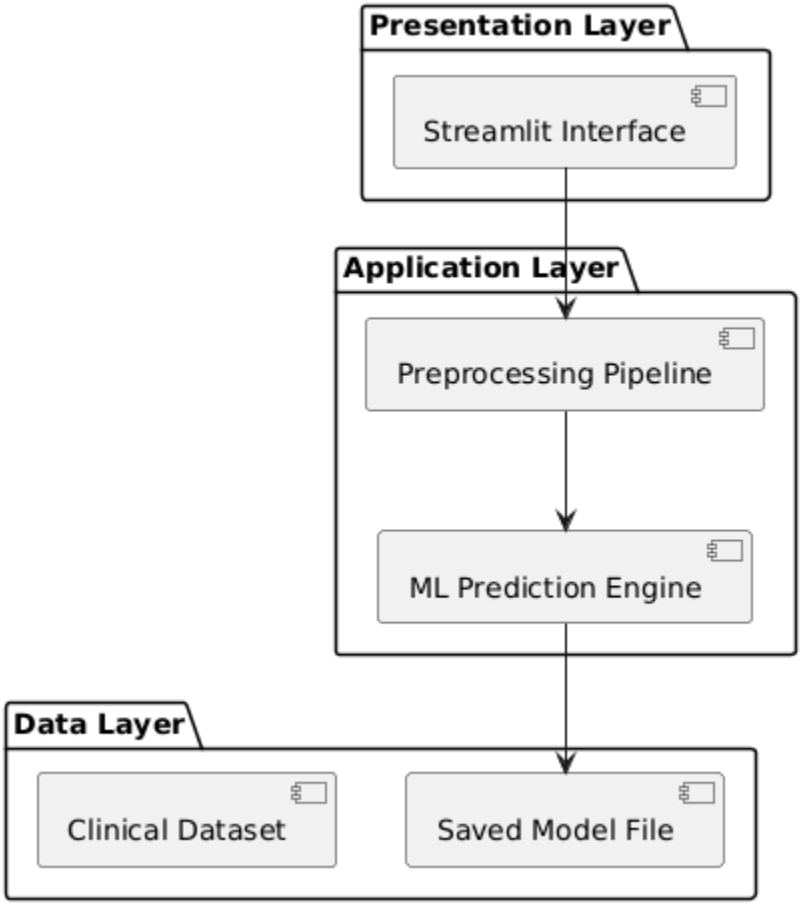


Activity Diagram – Model Training Workflow



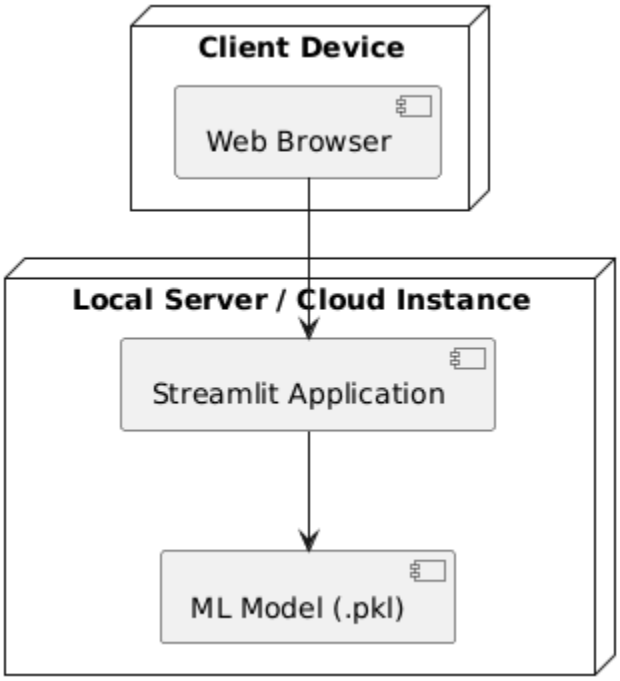
Component Diagram

The system follows a modular architecture.



Deployment Diagram

The Deployment Diagram shows the physical configuration of the prototype system.



Design Justification

The simplified single-user architecture:

- Eliminates authentication complexity
- Reduces system overhead
- Focuses on predictive modeling accuracy
- Enhances clarity for academic evaluation

The modular design ensures:

- Clear separation between preprocessing and prediction
- Ease of model updates
- Scalability for future integration into hospital systems