



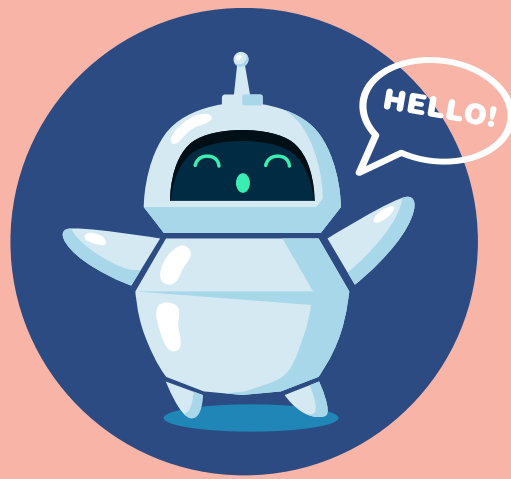
SDAIA X STANFORD HACKATHON

# Predicting Progression Free Survival In NSCLC using Multi-Modal Data

Team #2



# The Team



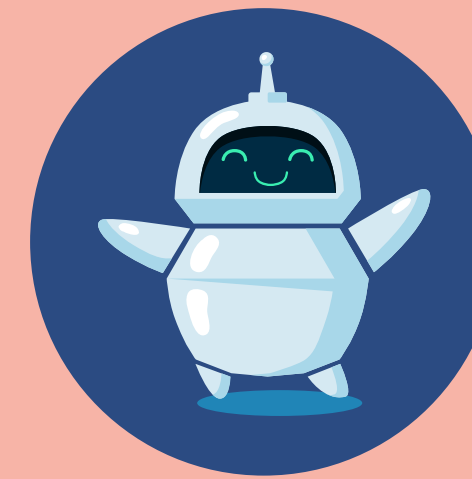
**Ahmed Abul  
Hasanaath**  
Computer Science



**Amjad  
Almubarak**  
Information System

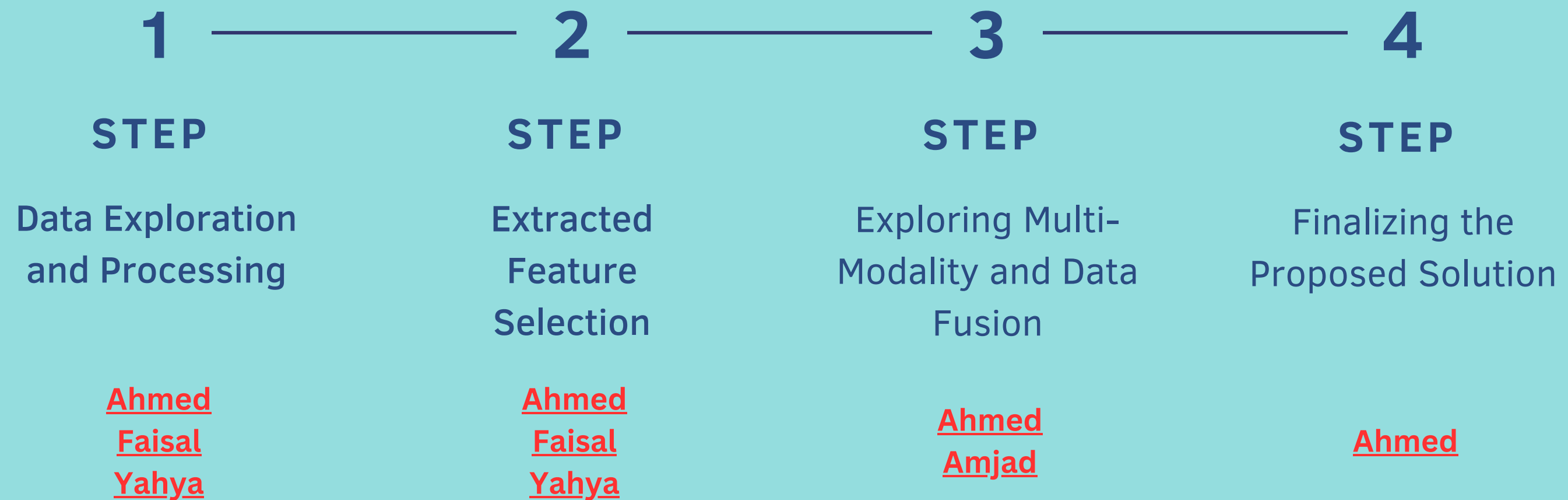


**Faisal  
Alarjani**  
Health Informatics



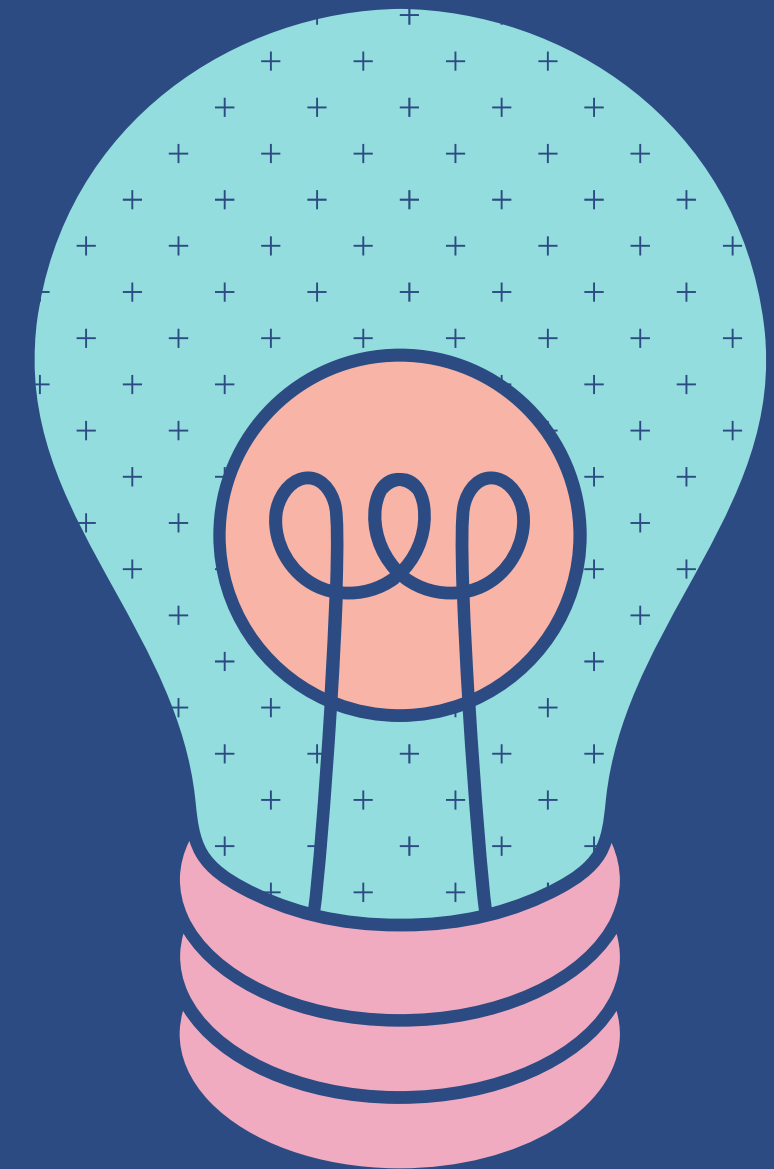
**Yahya  
Laghbi**  
Pharmacist

# Project Timeline



# Step 1: Exploration and Processing

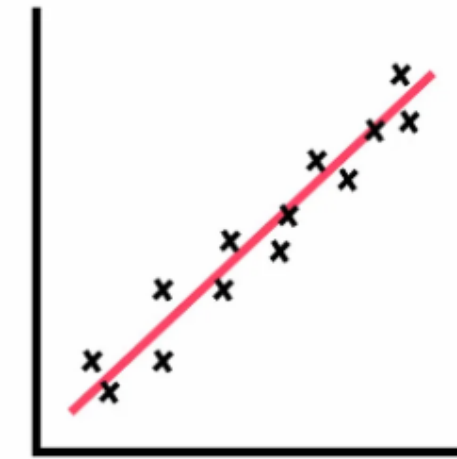
- **EXPLORED DATA; PARTICULARLY:**
  - **CLINICAL DATA**
  - **GENOMIC DATA**
  - **EXTRACTED PATHOLOGY FEATURES**
  - **RADIOMICS DATA**
- **CLEANED THE DATA**
  - **REMOVE UNNECESSARY COLUMNS**
  - **DELETED ROWS WITH EMPTY CELLS**
  - **ETC**



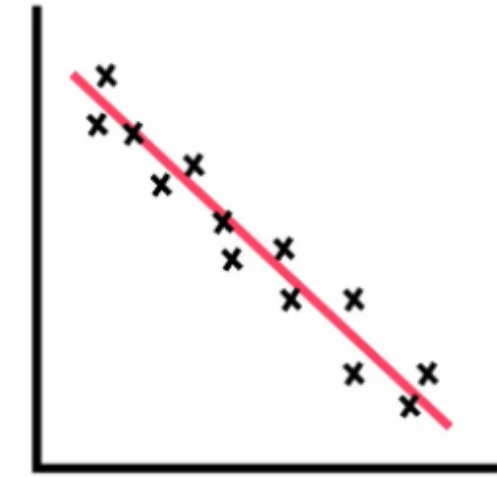
# Step 2: Extracted Feature Selection

## Target and Feature Correlation Based Selection

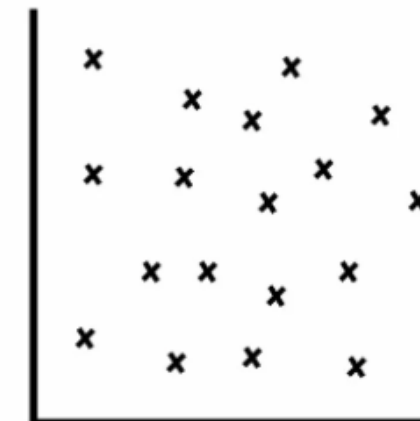
- Trained ML models on features with high correlation
  - Trained ML models on individual datasets:
    - radiomic
    - genomic
    - clinical
    - pathology
  - Trained ML models on combinations of aforementioned datasets



Positive  
Correlation



Negative  
Correlation

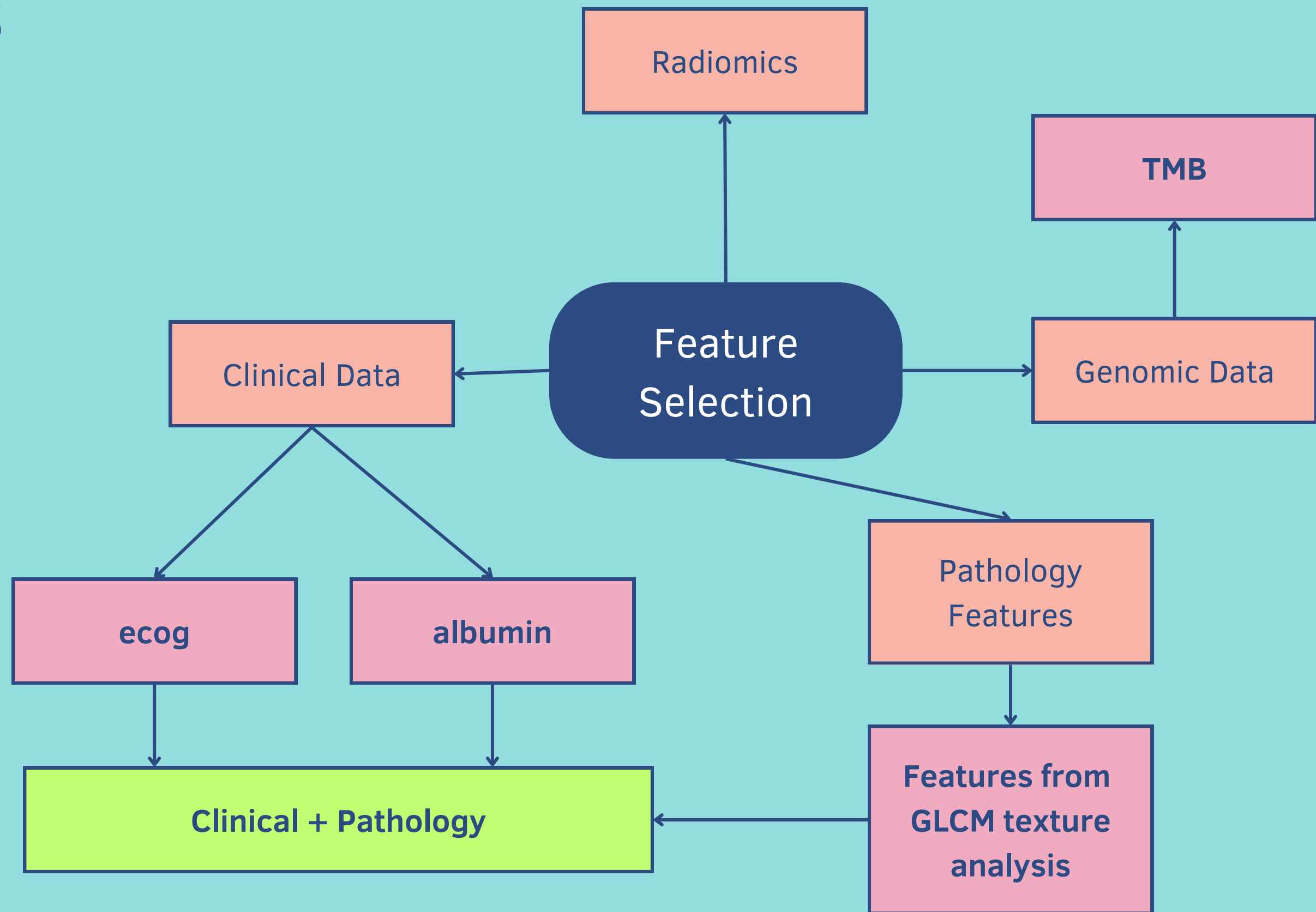


No  
Correlation

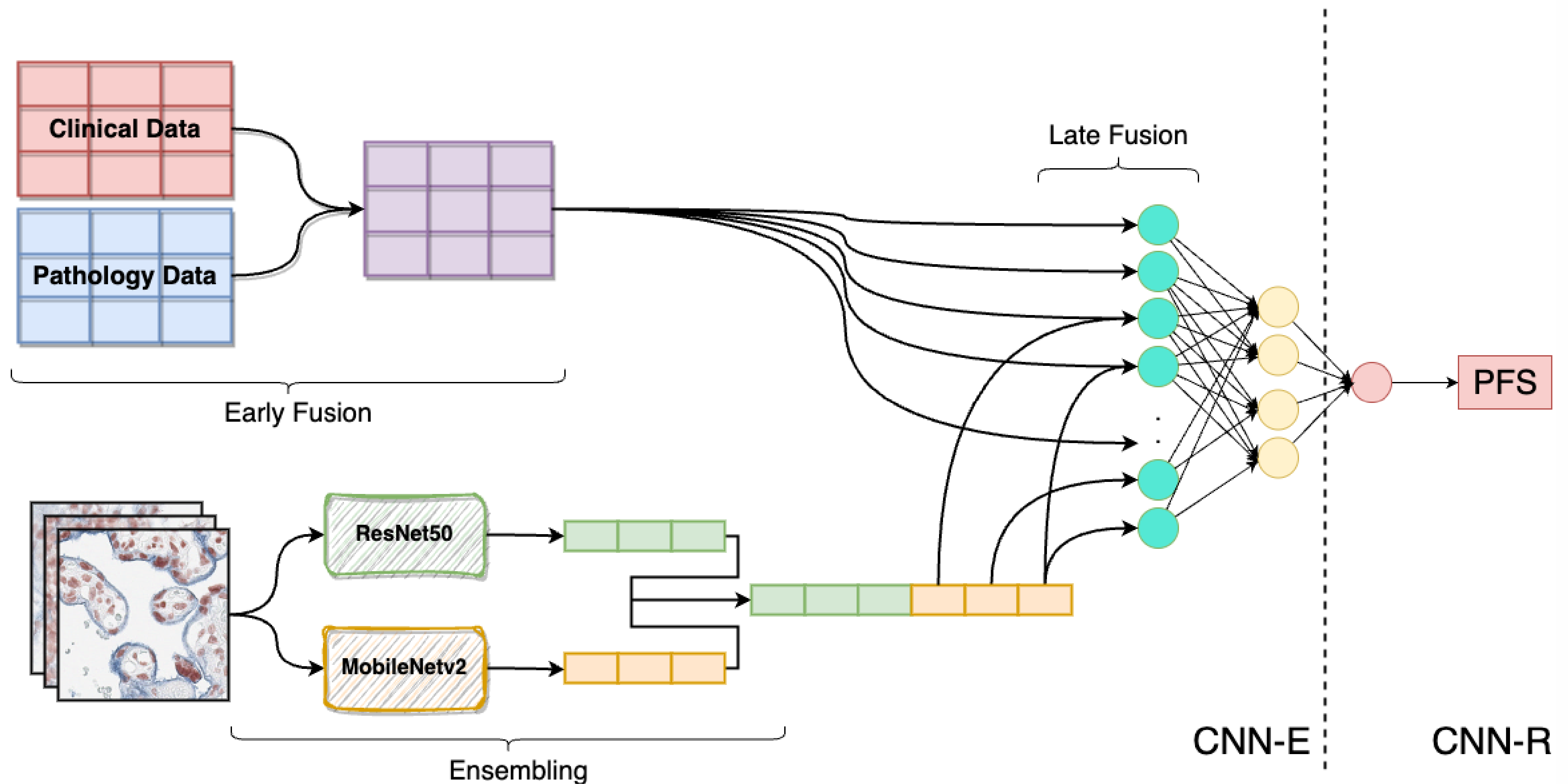
# Selected Features

## Criteria

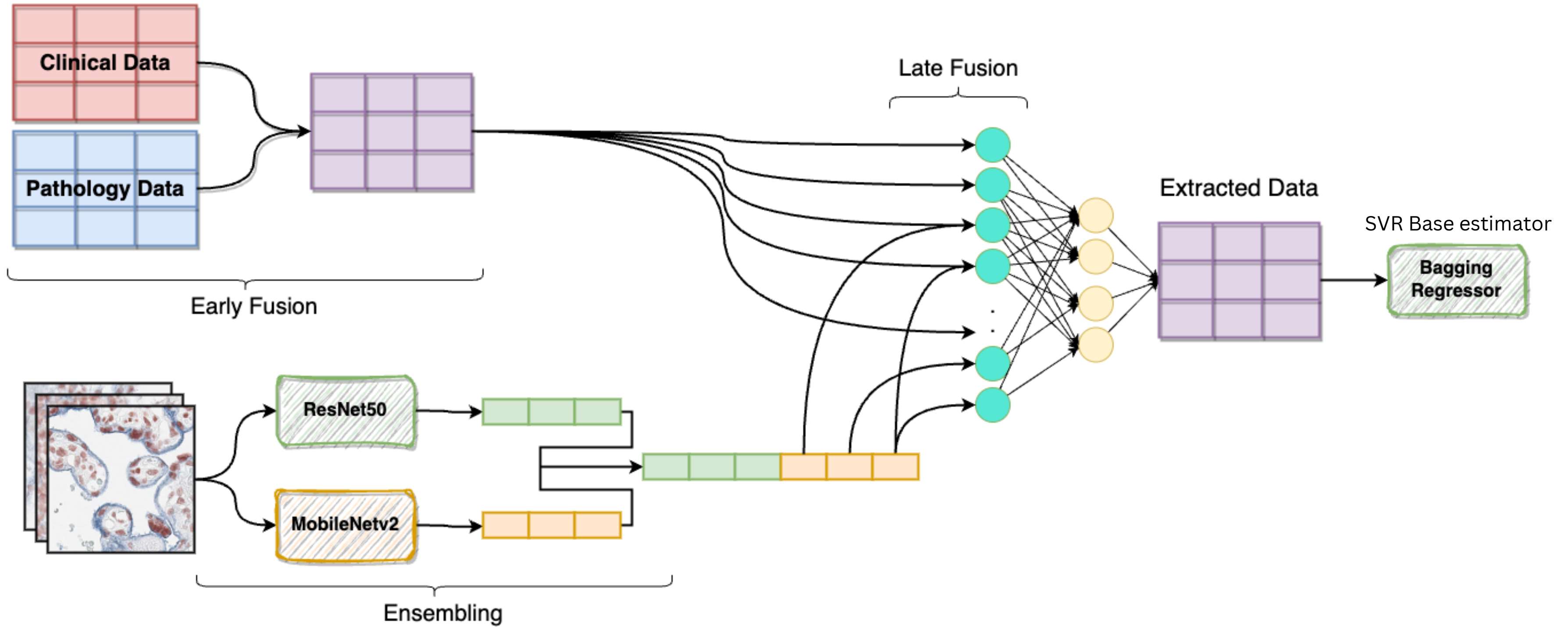
- Significant Correlation with PFS
- Decent Performance on Preliminary Training of ML Models



# Step 3: Multi-Modality and Data Fusion



# Step 4: Final Solution





# Results

Clinical Data + Extracted Pathology Features on Bagging Regressor

MSE: 36.91 | PCC: 0.4991

Clinical Data + Pathology Features + Pathology Patches on CNN-R

MSE: 41.02 | PCC: 0.3922

Clinical Data + Path. Features + Path. Patches on CNN-E + BaggingRegressor

MSE: 35.43 | PCC: 0.4009

NOTE: These scores were generated on the validation splits as instructed

*Thank  
you!*

