A Novel Machine Learning Framework for Identifying Predictive Biomarkers of FGFR Targeted Therapy in Breast Cancer

Rishiraj Sinharay - 32435851
Supervisor: A/Prof. Lan K. Nguyen Co-Supervisor: Dr. Sungyoung Shin
MONASH University Information Technology

1. INTRODUCTION

Breast cancer, a complex and heterogeneous illness, presents therapeutic obstacles. Traditional therapies target rapidly dividing cells, including cancer cells, which can have systemic consequences. Targeted therapy is promising because it identifies and attacks cancer cells while causing minimal damage to normal cells.

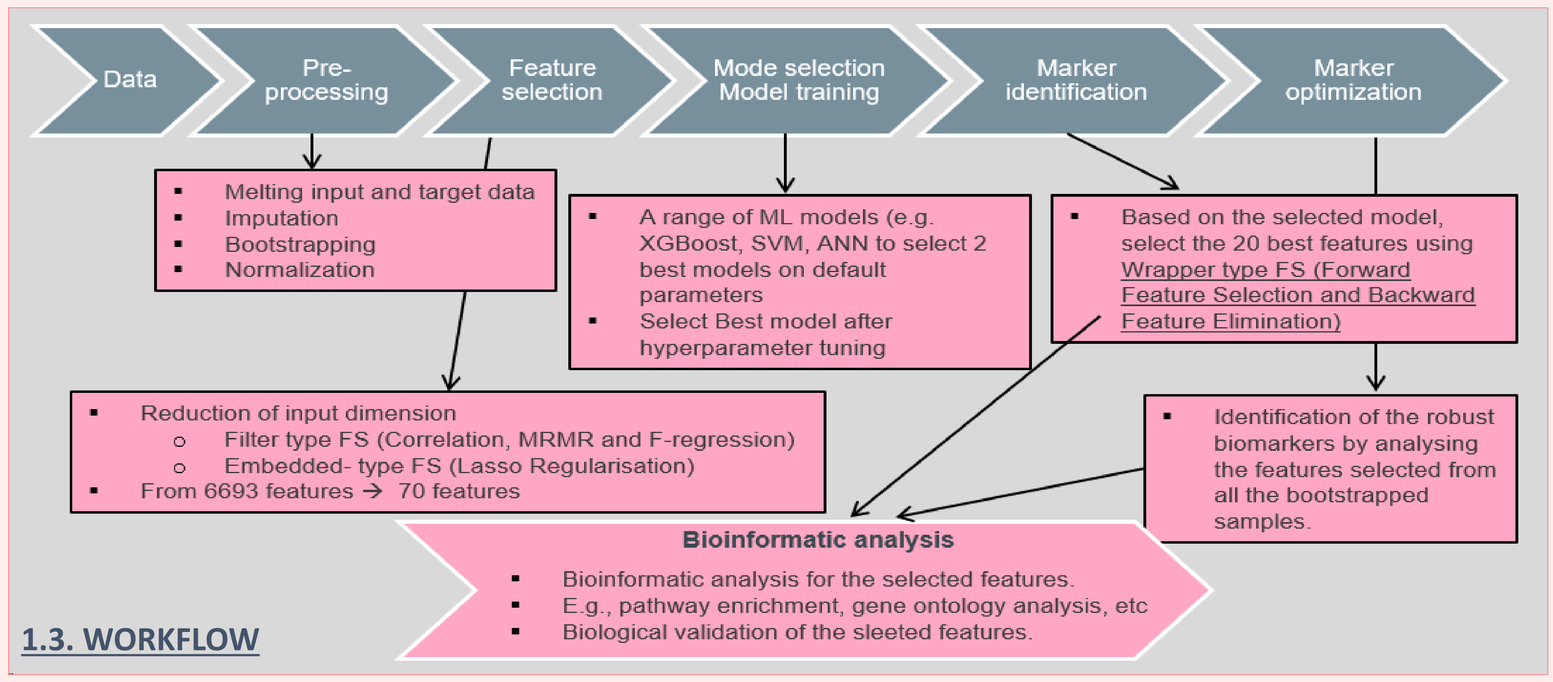
1.1. Why FGFR (Fibroblast Growth Factor Receptor)?

The FGFR pathway is essential in cell processes and its alterations are linked to breast cancer subtypes. FGFR inhibitors, in clinical trials, show promise for targeted treatment. However, the absence of reliable biomarkers to predict drug response is a challenge. The complexity of breast cancer makes single biomarkers insufficient, highlighting the need for comprehensive solutions to enhance treatment precision and efficacy.

1.2. What are we doing?

We employ machine learning and enrichment analysis to identify potential biomarkers and pathways for AZD4547 FGFR-targeted therapy. The research bridges computational analysis and biological insight, and helps understand gene enrichment in cancer-cell lines.

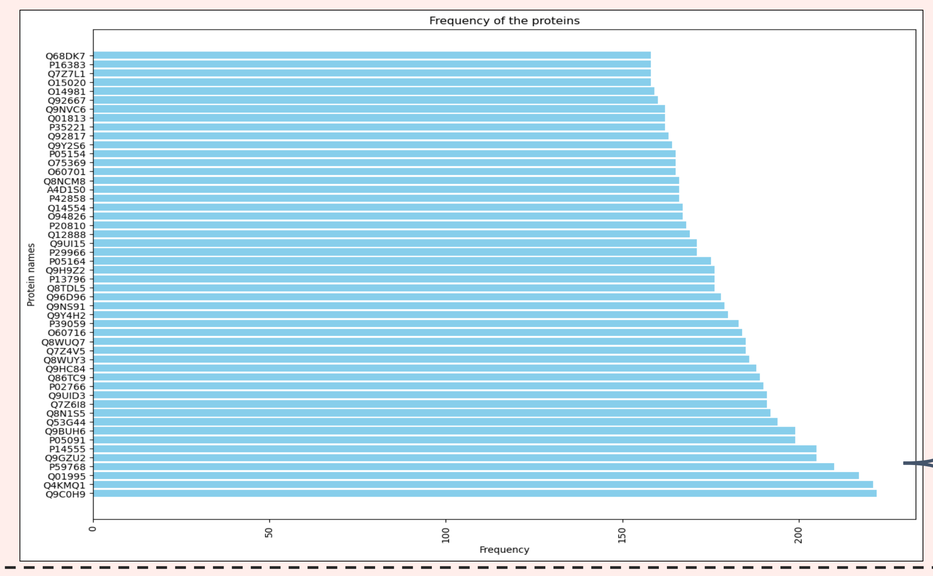
1.3. WORKFLOW



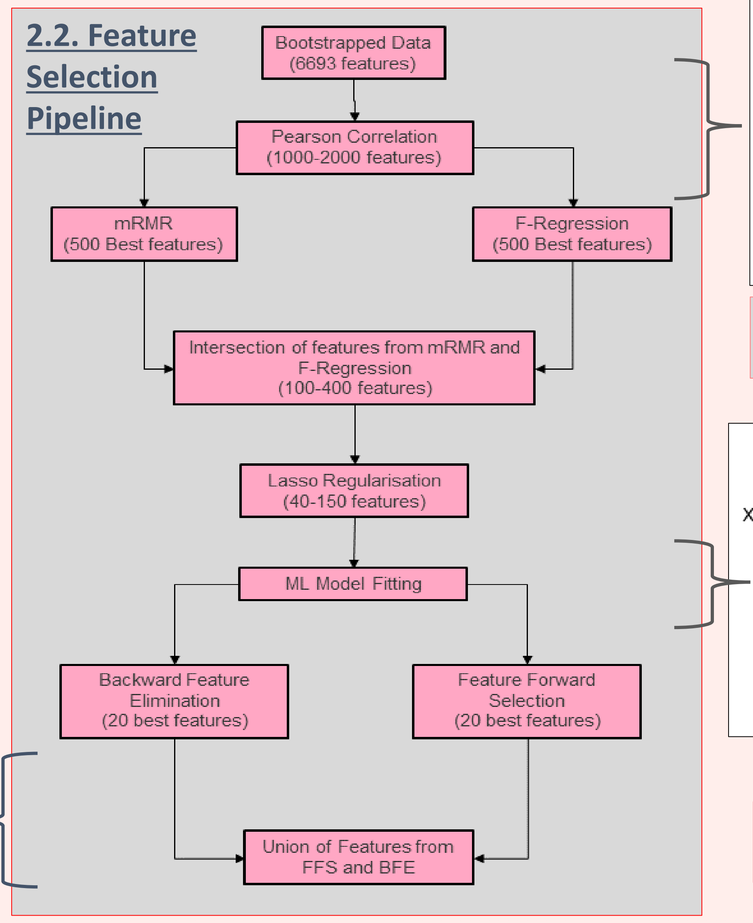
2. IMPLEMENTATION

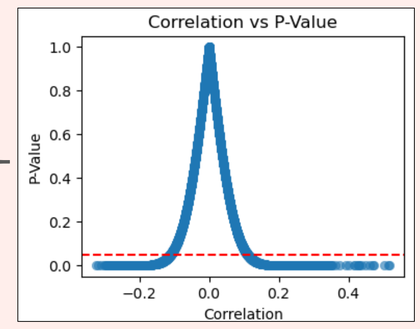
2.1. Data and Pre-Processing

- Pan-cancer dataset created by integrating proteomics data from Cancer Cell Line Encyclopaedia (CCLE) and AZD4547 drug response data from DepMap portal.
- Gibbs-sampler based imputation for missing values.
- The dataset is bootstrapped over 1000 times for feature selection, ensuring that the selection was not influenced by random variance in any single dataset iteration.



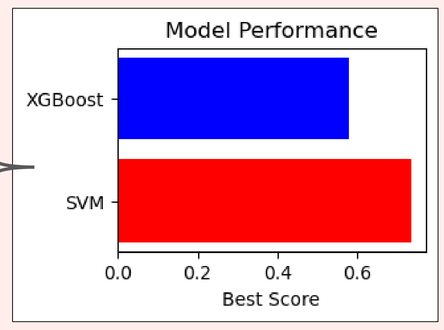
2.2. Feature Selection Pipeline





Using Pearson correlation, we identify proteins with a statistically significant linear relationship (p-value < 0.05) to drug response.

Utilizing multiple FS methods ensured a nuanced identification of robust biomarkers.



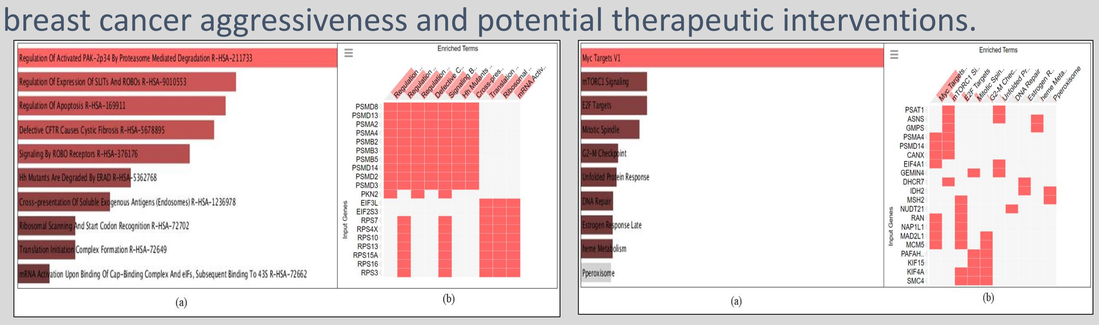
We choose XGBoost, SVM and MLPRegressor models for model fitting. Based on the data in the bootstrap iteration, the top 2 models which performs the best on default parameters is selected for model fitting.

The best model after hyperparameter tuning is selected for FFS and BFE (SVM for this particular iteration)

3. RESULTS

Enrichment analysis of genes derived from different stages of the feature selection pipeline for a particular bootstrap iteration

- Filter feature selection methods yielded 247 genes, with notable pathways like the **Regulation of Expression of SLITs and ROBOs**, and **Myc Targets V1** associated with breast cancer aggressiveness and potential therapeutic interventions.

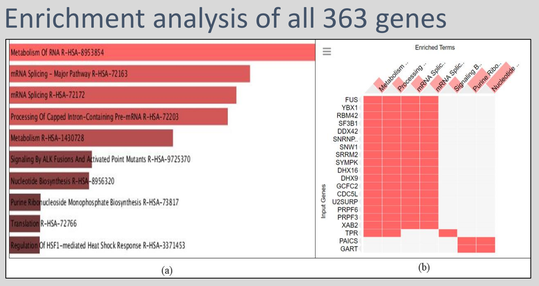


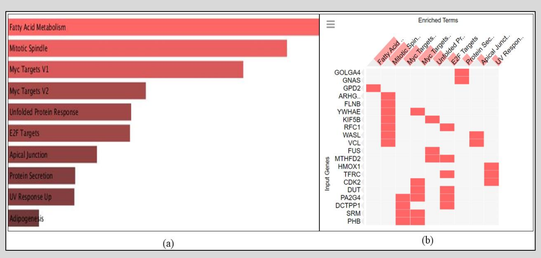
- Embedded feature selection methods resulted in 89 genes. Highly enriched pathways included **Signalling By Membrane-Tethered Fusions Of PDGFRA Or PDGFRB**, and **Oestrogen Response Early**, both implicated in breast cancer pathologies.

Enrichment analysis of genes amassed from bootstrap iterations. Genes were selected based on their frequency of occurrence in all the iterations.

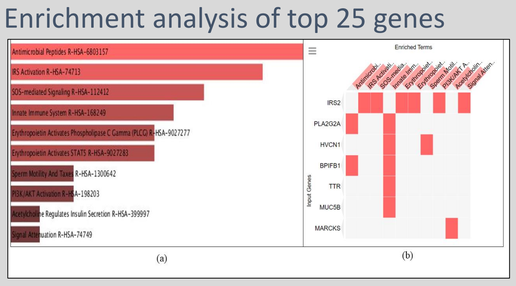
Frequency of Occurrence	Number of Genes	Enriched Pathways which van be Potential Target in Breast Cancer Therapy
No filter	363	RNA metabolism, mRNA splicing, and fatty acid metabolism
>100	168	Membrane trafficking pathways and E2F pathway
>150	50	IRS proteins, Unfolded Protein Response (UPR), and Estrogen Response Late
No filter	25, 20,10	Antimicrobial Peptides, TNF-alpha signaling, PI3K/AKT activation pathway, and IL-6/JAK/STAT3 signaling system

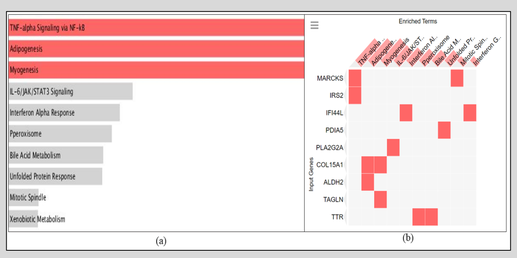
Enrichment analysis of all 363 genes



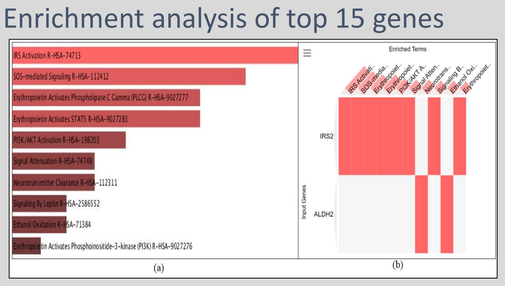


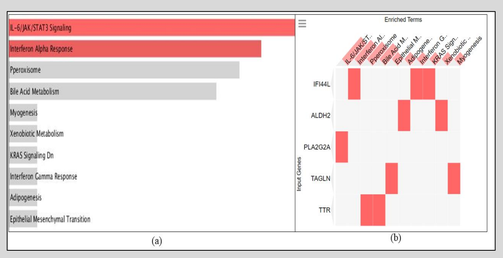
Enrichment analysis of top 25 genes





Enrichment analysis of top 15 genes





4. CONCLUSION

The examination of gene sets via Enrichr revealed multiple enriched pathways with potential implications in breast cancer target therapy. The recurrent appearance of genes **PLA2G2A**, **BPIFB1**, and **IRS2** across enriched pathways underscores their potential significance in targeted breast cancer therapies with AZD4547 FGFR inhibitor. The different enriched pathways found across different versions of the gene set show how biological processes interact in breast cancer in a complicated way. The results give us a strong base for more research into targeted therapeutic methods, which could lead to more successful, individualised treatments for breast cancer.