Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 273~283
e-ISSN: 2548-6861

The Impact of the L1/L2 Ratio on Selection Stability and Solution Sparsity along the Elastic Net Regularization Path in High-Dimensional Genomic Data

Fani Fahira [1]*, Kusman Sadik [2]*, Cici Suhaeni [3]*, Agus M Soleh [4]*
*School of Data Science, Mathematics and Informatics, IPB University
fanifahira@apps.ipb.ac.id [1], kusmans@apps.ipb.ac.id [2], cici_suhaeni@apps.ipb.ac.id [3], agusms@apps.ipb.ac.id [4]

Article Info

Article history:

Received 2025-12-18
Revised 2025-12-29
Accepted 2026-01-13

Keywords:

Elastic Net,
LASSO,
High-Dimensional Data,
Feature Selection Stability,
Sparsity,
Regularization Path.

ABSTRACT

High-dimensional genomic datasets ($p > n$) pose persistent challenges for predictive modeling and biomarker-oriented feature selection due to multicollinearity and instability of selected feature sets under resampling. Although Elastic Net is widely used to address correlated predictors via combined L1/L2 regularization, the practical role of the L1/L2 mixing ratio (α) is often treated as a secondary tuning choice driven primarily by predictive accuracy. This study investigates how varying α shapes the trade-off among selection stability, solution sparsity, and predictive performance along the Elastic Net regularization path. Experiments were conducted using the publicly available METABRIC breast cancer cohort (n = 1,964) with 21,113 gene expression features and a binary overall survival status outcome. Logistic regression with Elastic Net penalty was fitted across a grid of α values, with the regularization strength (λ) selected by cross-validation. Feature selection stability was evaluated under repeated resampling using the Jaccard index, Dice coefficient, and Adjusted Rand Index (ARI), while sparsity was summarized by the average number of non-zero coefficients; predictive performance was assessed using AUC, accuracy, and F1-score. Results show a monotonic decline in stability as α increases: α = 0.2 yields the highest stability (Jaccard 0.324, Dice 0.487, ARI 0.434), whereas LASSO (α = 1.0) produces the lowest stability (Jaccard 0.278, Dice 0.431, ARI 0.400). In contrast, predictive performance varies only marginally across α (AUC 0.696–0.704; accuracy 0.666–0.671; F1-score 0.738–0.742), while sparsity changes substantially (average selected features 110–204). Coefficient path analyses further illustrate abrupt shrinkage under LASSO versus smoother, group-preserving shrinkage under Elastic Net, consistent with improved reproducibility under lower-to-moderate α. Frequency-of-selection analysis highlights genes repeatedly selected across resampling, supporting interpretability of stable configurations without claiming causal biomarker validity. Overall, the findings demonstrate that α is a substantive modeling choice that materially affects stability and sparsity even when accuracy is similar, motivating stability-aware tuning for high-dimensional genomic prediction and reproducible feature discovery.

I. INTRODUCTION

The rapid advancement of high-throughput technologies in biomedical and genomic research has led to the routine generation of high-dimensional datasets, in which the number of measured variables far exceeds the number of observations. Such data structures pose fundamental challenges for statistical modeling and machine learning, as conventional estimation methods often become unreliable due to overfitting, noise accumulation, and spurious correlations. These difficulties are widely associated with the curse of dimensionality, which motivates the need for regularization and feature selection strategies that impose parsimony and

enable the recovery of meaningful low-dimensional signals from complex data spaces [1]. These challenges are particularly pronounced in biomedical applications, where simulation studies have shown that sample size and correlation structure strongly influence variable screening performance, even for widely used regularization methods [2].

In genomic applications, high dimensionality is frequently accompanied by strong correlations among predictors. These correlations arise from genuine biological mechanisms, including gene co-regulation, shared molecular pathways, and interaction networks. While biologically informative, such correlation structures complicate variable selection by amplifying instability, whereby small perturbations in training data, sampling schemes, or parameter choices result in markedly different subsets of selected features. Both theoretical and empirical studies identify strong correlation as a principal driver of inconsistency in high-dimensional variable selection, emphasizing its detrimental impact on model reliability and interpretability [3], [4]. Complementary theoretical insights highlight how tuning parameters influence convergence rates, bias, and variable selection consistency in ultra-high dimensional settings [5].

Regularization-based approaches have become essential tools for addressing these challenges. Among them, the Least Absolute Shrinkage and Selection Operator (LASSO) has been widely adopted due to its ability to induce sparsity and perform variable selection simultaneously. However, LASSO exhibits well-documented limitations in the presence of multicollinearity. When predictors are correlated, LASSO tends to select a single variable from a correlated group, ignoring others that may be equally informative. This behavior can lead to unstable selection outcomes and reduce the reproducibility of results [3]. Despite these drawbacks, LASSO continues to be extensively used in biomedical research, including clinical and epidemiological studies, underscoring the importance of critically evaluating its behavior in high-dimensional settings [6].

Elastic Net regularization was introduced to address these limitations by combining L1 and L2 penalties within a unified framework. The inclusion of an L2 component allows Elastic Net to better handle correlated predictors by promoting a grouping effect, in which correlated variables tend to be selected together. This property enhances model robustness and stability relative to purely L1-based approaches. Consequently, Elastic Net remains a relevant and widely applied method in contemporary high-dimensional applications. For example, in geological and geochemical prospectivity mapping, Elastic Net demonstrated superior prediction accuracy and robustness compared to LASSO, particularly under correlated predictor structures [7]. Similar findings have been reported in comparative studies across datasets of varying complexity, where feature selection stability was shown to degrade under data perturbations even when classification accuracy remained relatively unchanged [8].

Although predictive accuracy is often emphasized in model evaluation, a growing body of literature argues that accuracy alone is insufficient for assessing the quality of feature selection methods. Feature selection stability, defined as the consistency of selected feature sets under perturbations of the data or modeling process, has emerged as a critical dimension of model reliability. Empirical evidence shows that models can maintain stable classification accuracy while exhibiting substantial variability in selected features, revealing a disconnect between predictive performance and selection consistency [9]. In genomic research, where selected features are frequently interpreted as candidate biomarkers, such instability undermines reproducibility and scientific validity.

To quantify feature selection stability, various similarity-based and agreement-based metrics have been proposed, including the Jaccard index, Dice coefficient, and Adjusted Rand Index (ARI). These measures assess the overlap or agreement between feature subsets obtained from different perturbations. However, empirical studies demonstrate that stability metrics can behave differently depending on dataset characteristics and feature selection methods. Investigations across multiple gene expression datasets reveal high variability in stability values, highlighting the complexity of stability assessment [10]. Moreover, critical analyses indicate that no single stability metric satisfies all desirable theoretical properties, motivating ongoing methodological refinement [11]. The ARI, originally developed for clustering validation, has also been shown to be applicable for evaluating agreement in supervised classification and feature selection contexts [12].

Beyond metric development, methodological research has focused on improving stability through algorithmic innovations. Stability selection is a prominent framework that enhances feature selection by repeated subsampling and aggregation, while providing theoretical control over the expected number of false positives. However, conventional stability selection methods are often conservative, resulting in the selection of only a small number of features. Recent advances propose integrating stability paths rather than maximizing over them, yielding substantially stronger bounds on false positives and improving true positive recovery without increasing computational cost. These methods have been demonstrated on simulated data and real cancer datasets, highlighting their practical relevance [13].

Stability-oriented approaches have also been extended to other modeling frameworks. For instance, stability selection has been applied to regularized structural equation modeling to mitigate high false positive rates and inconsistent selection results associated with LASSO-based regularization. Simulation studies and empirical applications indicate that stability selection can substantially reduce false positives and improve the correctness of selected model components [14]. Although these studies focus on different model classes, they underscore a common theme: stability is a fundamental concern across regularized modeling approaches.

Adaptive regularization methods provide another avenue for improving stability and performance. Adaptive LASSO and adaptive Elastic Net introduce data-driven weights into the penalty term, allowing the degree of regularization to vary across predictors. Studies on high-dimensional sparse data with multicollinearity show that the choice of adaptive weights and their power order strongly influences predictive accuracy and selection behavior. Simulation and real-data analyses demonstrate that certain adaptive Elastic Net configurations outperform standard regularization methods in correlated settings [15]. Complementary findings in adaptive LASSO-based sparse logistic regression further emphasize the importance of tuning choices, including the selection of initial estimators, in determining model performance under multicollinearity [16].

Ensemble feature selection methods have also gained attention as a means of improving stability by aggregating results across multiple selectors or perturbations. Empirical studies on high-dimensional microarray datasets indicate that ensemble approaches, particularly voting-based methods, yield higher stability and competitive or improved accuracy compared to traditional feature selection algorithms [17]. Additional evidence from high-dimensional genetic datasets confirms that ensemble-based feature selection consistently outperforms single-method approaches in terms of stability, particularly when evaluated using similarity-based metrics such as the Jaccard index [18]. In biomedical contexts such as Alzheimer's disease biomarker discovery, ensemble feature selection combined with data-driven thresholding has been shown to produce more stable and reproducible feature sets without sacrificing predictive performance [19]. These results suggest that stability improvements can often be achieved without compromising accuracy.

The stability of feature selection is also influenced by data characteristics and pipeline design. Simulation studies examining filter-based feature selection methods reveal that stability is negatively affected by factors such as measurement error, limited sample size, and class imbalance. Moreover, different stability metrics respond differently to these data properties, with some measures exhibiting greater variability than others [20]. These findings highlight the importance of evaluating stability within the context of the entire modeling pipeline rather than attributing it solely to the selection algorithm.

The growing emphasis on stability aligns with broader concerns about reproducibility and replicability in computational research. Surveys on reproducible research identify complex computational workflows and overreliance on performance metrics as contributing factors to reproducibility challenges across scientific disciplines [21]. In this context, stability analysis provides an essential complement to accuracy-based evaluation, particularly in high-dimensional biomedical applications where reproducibility is critical.

Despite extensive methodological development, an important research gap remains. Although Elastic Net is a well-established regularization method, the specific impact of the L1/L2 mixing ratio on the trade-off between solution sparsity and feature selection stability has not been systematically examined, particularly along the regularization path. Many applied studies prioritize predictive performance, potentially overlooking how tuning the mixing parameter influences stability and reproducibility. Given evidence that correlation structure and tuning choices strongly affect selection behavior [3], [4], [5], a systematic investigation of the L1/L2 ratio is both timely and necessary.

Accordingly, this study aims to evaluate the impact of the L1/L2 ratio on selection stability and solution sparsity along the Elastic Net regularization path in high-dimensional genomic data. By jointly assessing predictive performance, stability metrics, and sparsity, this work aligns with recent stability-aware methodological perspectives and contributes to the development of genomic models that are not only predictive but also reproducible, interpretable, and reliable for downstream biomarker discovery.

II. METHOD

A. Data

The data used in this study were obtained from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), a publicly available genomic dataset that provides gene expression profiles and curated clinical information for breast cancer research. The dataset comprises 1,964 breast cancer patients and includes mRNA expression data measured using Illumina microarray platforms, along with associated clinical annotations.

The dataset is distributed under the Open Database License (ODbL) version 1.0, which permits use, sharing, and modification of the data with appropriate attribution. All patient records are fully anonymized; therefore, no additional ethical approval was required. All analyses were conducted using Python 3.12, with the NumPy, pandas, scikit-learn, and matplotlib libraries. Fixed random seeds were used throughout the analysis to ensure reproducibility. Table I summarizes the structure of the variables used in this study. The predictor space consists exclusively of high-dimensional gene expression features, while the clinical survival status serves as the response variable.

TABLE I
DATA STRUCTURE AND VARIABLES OF THE METABRIC DATASET

Data Domain	Feature Category	Indicator
Gene Expression Data (G1)	mRNA Expression Profiles	Continuous expression values for 21,113 genes measured using Illumina microarray platforms
	Correlated Gene Sets	Groups of genes exhibiting strong correlations due to co-regulation
	High-Dimensional Feature Space	Predictor space characterized by $p \gg n$
Clinical Outcome (O1)	Overall Survival Status	Binary outcome (*Deceased*, *Living*)

B. Elastic Net Model

Elastic Net regularization was employed within a logistic regression framework to simultaneously perform classification and feature selection. The regression coefficients $\beta \in \mathbb{R}^p$ were estimated by minimizing the following penalized negative log-likelihood function [3], [22]:

$$\mathcal{L}(\beta) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{p}_i) + (1-y_i)\log(1-\hat{p}_i)] + \lambda\left(\alpha\|\beta\|_1 + \frac{1-\alpha}{2}\|\beta\|_2^2\right) \quad (1)$$

where

$$\hat{p}_i = \frac{1}{1+\exp(-X_i\beta)},$$

$\lambda > 0$ is the regularization parameter, and
$\alpha \in [0,1]$controls the L1/L2 mixing ratio.

The L1 penalty promotes sparsity, while the L2 penalty stabilizes coefficient estimation under multicollinearity, making Elastic Net particularly suitable for correlated high-dimensional genomic data [4], [22].

C. Regularization Path and L1/L2 Ratio

To investigate the impact of the L1/L2 ratio, the Elastic Net mixing parameter was varied over a predefined grid $\alpha \in \{0.1, 0.2, \ldots, 1.0\}$. For each value of α, the optimal regularization strength λ was selected using cross-validation based on the minimum mean cross-validated loss ($\lambda_{\min}$), following standard practice in high-dimensional regularized regression [5]. This procedure yields a regularization path along which changes in coefficient sparsity, feature selection stability, and predictive performance can be systematically analyzed. By analyzing the entire regularization path, this framework enables systematic examination of how varying the balance between L1 and L2 penalties influences solution sparsity, feature selection stability, and predictive performance

D. Feature Selection Stability

For each resampling iteration r, the selected feature set was defined as:

$$S^{(r)} = \left\{j : \beta_j^{(r)} \neq 0\right\}$$

Feature selection stability was evaluated using three complementary metrics commonly used in genomic studies[10], [20]

1) Jaccard Index

$$J\left(S^{(r)}, S^{(s)}\right) = \frac{\left|S^{(r)} \cap S^{(s)}\right|}{\left|S^{(r)} \cup S^{(s)}\right|} \quad (2)$$

2) Dice Coefficient

$$D\left(S^{(r)}, S^{(s)}\right) = \frac{2\left|S^{(r)} \cap S^{(s)}\right|}{\left|S^{(r)}\right| + \left|S^{(s)}\right|} \quad (3)$$

3) Adjusted Rand Index (ARI)

The ARI measures agreement between two feature selection partitions while correcting for chance agreement [12]. For each value of α, stability scores were obtained by averaging pairwise comparisons across all resampling runs.

E. Solution Sparsity

Solution sparsity for a given α was defined as the number of non-zero coefficients:

$$Sparsity(\alpha) = \frac{1}{R}\sum_{r=1}^{R}\left|S^{(r)}\right| \quad (4)$$

where R denotes the total number of resampling iterations. Averaging across resampling runs ensures that sparsity reflects the typical model complexity rather than a single realization [6].

F. Predictive Performance Evaluation

Predictive performance was assessed on the held-out test sets using standard classification metrics:

- Area Under the ROC Curve (AUC),
- Accuracy,
- F1-score.

Performance metrics were averaged across resampling runs to obtain robust estimates [15].

G. Trade-off Analysis

For each value of α, the following quantities were jointly evaluated:

1. Predictive performance,
2. Feature selection stability,
3. Solution sparsity.

This multi-criteria analysis enables identification of L1/L2 ratios that balance accuracy, stability, and interpretability, consistent with recent stability-aware methodological frameworks [13]

H. Analysis Procedure

The overall analysis procedure consists of the following steps:

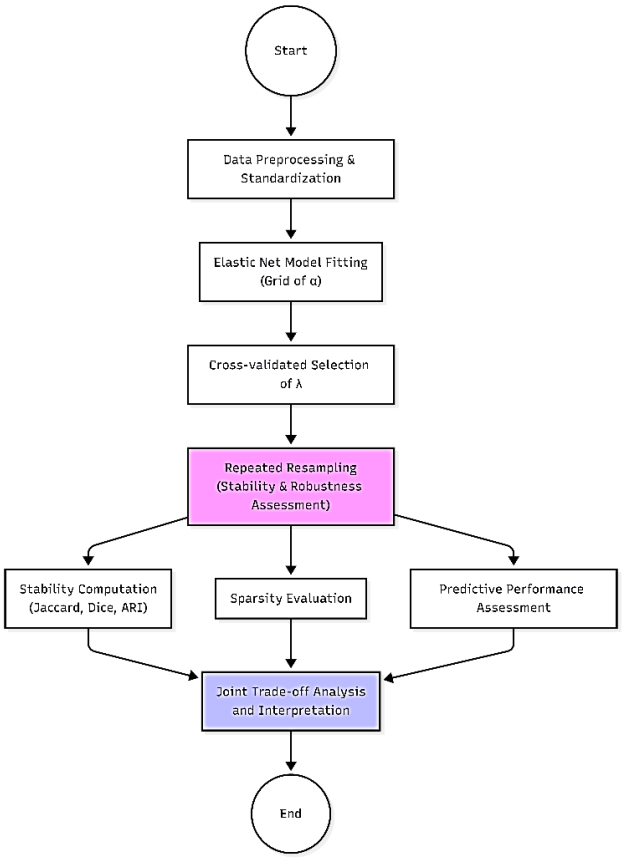

Figure 1. Research Flowchart

Figure 1 presents the overall research workflow for evaluating the impact of the L1/L2 ratio along the Elastic Net regularization path. The analysis starts with data preprocessing and standardization, followed by Elastic Net model fitting across a grid of mixing parameters (α) and cross-validated selection of the regularization parameter (λ). The procedure is embedded within repeated resampling to assess robustness and feature selection stability. Stability metrics, solution sparsity, and predictive performance are then evaluated in parallel and integrated through a joint trade-off analysis to support interpretation of reliable and interpretable models.

III. Results and Discussion

This section presents and interprets the empirical findings regarding the effect of varying the L1/L2 mixing ratio (α) in Elastic Net regularization on feature selection stability, predictive performance, and biological interpretability. The analysis was conducted on the METABRIC dataset, consisting of high-dimensional gene expression profiles from 1,964 breast cancer patients. Four representative values of α were examined—0.2, 0.5, 0.8, and 1.0—corresponding to increasing dominance of the L1 penalty, with $\alpha = 1.0$ representing the LASSO model.

A. Feature Selection Stability

Feature selection stability was evaluated using three complementary similarity metrics: Jaccard Index, Dice Coefficient, and Adjusted Rand Index (ARI). Table II summarizes the average stability scores obtained across repeated 5-fold cross-validation for each value of α.

TABLE II
Feature Selection Stability Across α Values

Model	alpha (α)	Jaccard	Dice	ARI
Elastic Net	α = 0.2	0.324	0.487	0.434
Elastic Net	α = 0.5	0.291	0.447	0.411
Elastic Net	α = 0.8	0.285	0.440	0.408
LASSO	α = 1.0	0.278	0.431	0.400

A clear monotonic decline in stability is observed as α increases. Models with stronger L2 components ($\alpha = 0.2$) consistently yield higher agreement across resampling iterations, while LASSO ($\alpha = 1.0$) exhibits the lowest stability. This pattern empirically confirms the grouping effect of Elastic Net, whereby correlated genes tend to be selected together, reducing sensitivity to small perturbations in the training data. Importantly, these results demonstrate that sparser solutions do not necessarily imply more reliable feature selection. Although LASSO enforces aggressive sparsity, it does so at the cost of reproducibility, a critical limitation in genomic biomarker discovery where consistency across studies is essential.

Although the stability metrics show a consistent decreasing trend as α increases, the absolute values provide additional insight into the inherent difficulty of stable biomarker selection in high-dimensional gene expression settings. For example, the Jaccard scores (approximately 0.28–0.32) indicate only moderate overlap between selected gene sets across resampling iterations, even in the most stable configuration. This suggests that while Elastic Net improves reproducibility relative to LASSO, the feature selection problem remains intrinsically sensitive to perturbations due to the $p \gg n$ structure and correlated predictor blocks. Importantly, the consistent ordering across Jaccard, Dice, and ARI indicates that the stability conclusion is not metric-dependent, strengthening the robustness of the inference. At the same time, the differences in scale across metrics highlight

that stability should be interpreted comparatively (across α) rather than as an absolute "good/bad" threshold.

B. Predictive Performance and Sparsity Trade-off

A notable finding is that predictive performance (AUC, accuracy, F1-score) varies only slightly across α, whereas stability and sparsity change more substantially. This pattern is consistent with the fact that, in high-dimensional settings with correlated predictors, multiple alternative feature subsets can yield similar classification performance because correlated genes may encode overlapping predictive signal. Consequently, accuracy-driven selection may identify models that are "predictively equivalent" but biologically inconsistent in terms of selected biomarkers. This reinforces the argument that stability should be included as a complementary criterion whenever interpretability and reproducible feature identification are desired. Predictive performance was evaluated using repeated train–test splits, and average metrics are reported in Table III alongside the average number of selected features.

TABLE III
PREDICTIVE PERFORMANCE AND SPARSITY ACROSS α VALUES

Model	α	Mean AUC	Accuracy	F1-score	Avg. # Features
Elastic Net	0.2	0.704	0.670	0.741	204
Elastic Net	0.5	0.699	0.671	0.742	139
Elastic Net	0.8	0.697	0.666	0.738	110
LASSO	1.0	0.696	0.668	0.739	119

The results reveal a non-trivial trade-off between model complexity and predictive behavior. The model with $\alpha = 0.2$ achieves the highest AUC, suggesting superior global class separation across all thresholds. This is consistent with its broader feature inclusion, which may capture complex multigenic patterns relevant to survival outcomes.

In contrast, the model with $\alpha = 0.5$ yields the highest F1-score while using substantially fewer features. This indicates a more balanced trade-off between precision and recall, particularly relevant in clinical risk stratification where correct identification of high-risk patients (DECEASED) is critical. The reduced feature set at $\alpha = 0.5$ may also mitigate overfitting, leading to more stable decision boundaries.

Notably, predictive performance remains relatively stable across different α values, whereas sparsity and stability vary considerably. This decoupling underscores the limitation of relying solely on accuracy-based metrics when model interpretability and reproducibility are key objectives. To complement the summary results in Table III, the predictive behavior along the regularization path was further examined using a heatmap visualization of AUC values across different values of α and λ. The heatmap (Figure 2) shows broad regions where AUC values remain comparable, indicating that similar predictive performance can be achieved across a wide range of parameter configurations. For each α, AUC was evaluated over a common grid of λ values along the regularization path, and the resulting AUC surface is summarized as a heatmap. This visualization is descriptive and complements the CV-selected $\lambda_{\min}$ by showing that near-optimal AUC is achieved across a broad λ range.

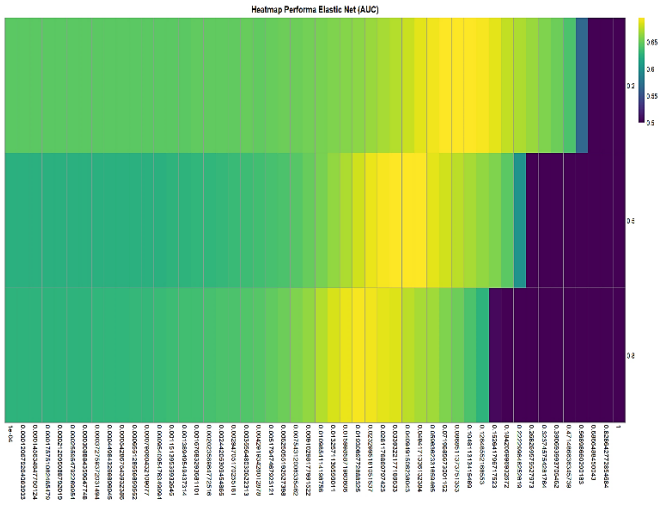

Figure 2. Heatmap of AUC values along the Elastic Net regularization path for different values of the L1/L2 mixing ratio α

Importantly, no sharply isolated maximum is observed, particularly for lower to moderate values of α, where extended performance plateaus appear along the regularization path. This pattern suggests that predictive accuracy is relatively insensitive to the exact choice of α and λ within these regions. As a result, accuracy-based tuning alone provides limited guidance for selecting the L1/L2 mixing ratio, since many configurations yield nearly identical AUC values while differing substantially in sparsity and feature selection stability.

C. Coefficient Paths and Regularization Dynamics

While increasing α generally shifts the model toward stronger sparsity, the observed number of selected features is not strictly monotonic across the reported configurations (e.g., the average number of features under LASSO is not smaller than under all Elastic Net settings). This can occur because the final sparsity is jointly determined by the mixing ratio α and the cross-validated regularization strength λ. In practice, different α values can lead cross-validation to favor different regions of the regularization path, resulting in feature counts that do not follow a simple monotonic pattern. This observation further supports the importance of analyzing regularization dynamics rather than assuming that "higher α" necessarily implies fewer selected features in the tuned model. To further examine the effect of regularization on coefficient shrinkage and sparsity, coefficient paths were analyzed separately for the LASSO and Elastic Net models as functions of the regularization parameter λ. Therefore, sparsity comparisons across α should be interpreted at their

respective CV-selected operating points rather than at matched λ values.

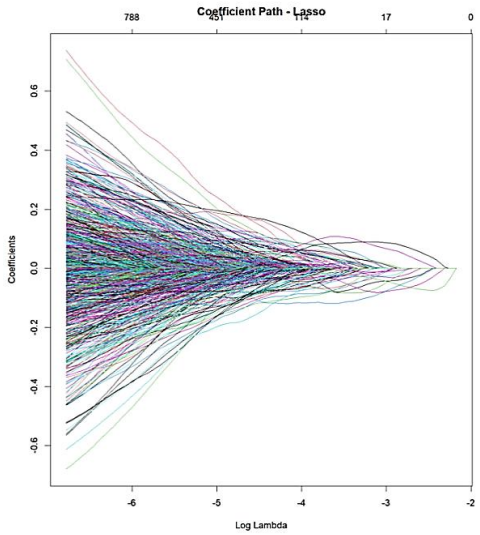

Figure 3. Coefficient paths of LASSO

Figure 3 presents the coefficient paths for the LASSO model. As λ increases, a large proportion of coefficients rapidly shrink to zero, reflecting the strong sparsity-inducing nature of the L1 penalty. This abrupt elimination of predictors indicates that LASSO enforces variable selection aggressively, often retaining only a small subset of features. While such sparsity can enhance interpretability, it also increases sensitivity to data perturbations, particularly in the presence of correlated predictors.

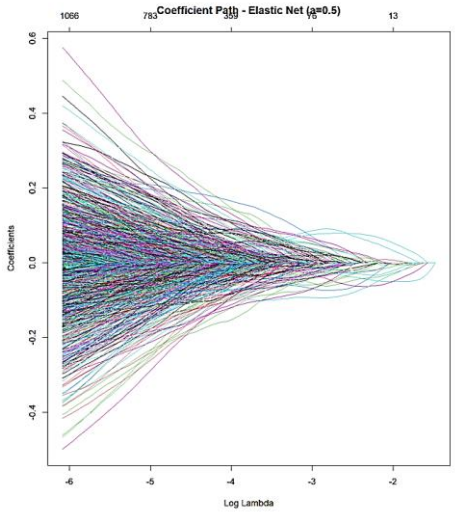

Figure 4. Coefficient paths of the Elastic Net model with α = 0.5

In contrast, Figure 4 illustrates the coefficient paths for the Elastic Net model with $\alpha = 0.5$. Compared to LASSO, coefficient trajectories exhibit smoother and more gradual shrinkage as λ increases. Many coefficients remain non-zero over a wider range of regularization strengths, highlighting the stabilizing influence of the L2 component. This behavior is consistent with the grouping effect of Elastic Net, which allows correlated genes to enter or leave the model together rather than being eliminated individually.

The visual contrast between Figures 3 and 4 underscores a fundamental difference in regularization dynamics. While LASSO prioritizes sparsity through abrupt coefficient suppression, Elastic Net balances sparsity and stability by moderating shrinkage across correlated features. These dynamics help explain the empirical findings observed in earlier sections, where Elastic Net models achieved higher feature selection stability without substantial loss in predictive performance.

D. Frequency of Feature Selection and Biological Interpretation

The frequency-of-selection analysis provides an interpretable bridge between stability metrics and biological plausibility. While pairwise similarity measures summarize agreement at the feature-set level, selection frequency highlights genes that persistently appear across resampling iterations. In biomarker discovery contexts, this perspective is particularly valuable because it prioritizes candidates that are not only predictive but also robust to data perturbations. The higher recurrence rates observed under Elastic Net are consistent with its grouped selection behavior, suggesting that Elastic Net preferentially retains coherent sets of correlated genes rather than isolated single-gene effects.

To further examine feature-level robustness, the frequency with which individual genes were selected across repeated resampling iterations was analyzed. Figure 5 displays the selection frequencies of the top 30 most frequently selected genes under the LASSO model and the Elastic Net model (α = 0.5). This visualization provides a direct comparison of how often each gene is retained across resampling runs, complementing the set-level stability metrics reported earlier.

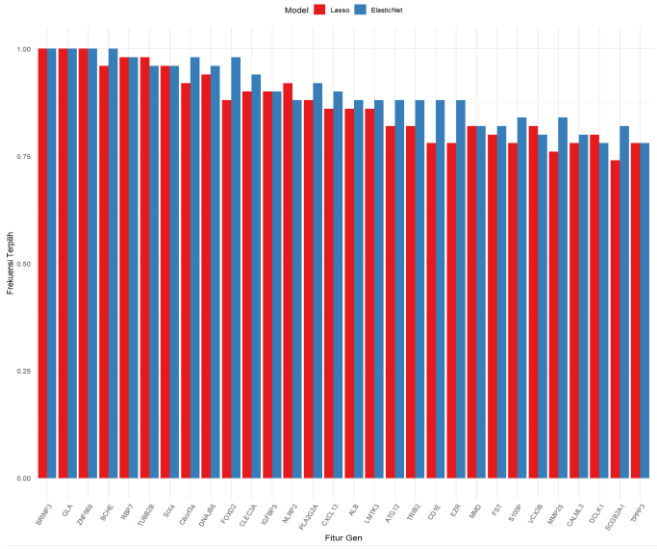

Figure 5. Selection frequency of the top 30 most frequently selected genes under LASSO and Elastic Net (α = 0.5) across repeated resampling iterations.

As shown in Figure 5, genes selected under the Elastic Net model exhibit consistently higher selection frequencies compared to those selected under LASSO. This pattern indicates improved robustness of individual gene selection under Elastic Net, which aligns with its grouping effect in the presence of correlated predictors. In contrast, LASSO demonstrates greater variability in gene inclusion, reflecting its tendency to arbitrarily select single representatives from correlated gene groups.

Several biologically well-established breast cancer–related genes emerge among the most frequently selected features under the Elastic Net model, including FOXC1, MKI67, CCNB1, and PLK1. FOXC1 (Forkhead Box C1) is strongly associated with epithelial-to-mesenchymal transition, cellular motility, and aggressive tumor phenotypes, particularly in basal-like and triple-negative breast cancer subtypes, which are characterized by poor prognosis [23]. MKI67, encoding the Ki-67 proliferation antigen, is widely used in clinical practice as a marker of tumor proliferation and grade, reinforcing the clinical plausibility of its recurrent selection. CCNB1 (Cyclin B1) and PLK1 (Polo-like kinase 1) play central roles in G2/M cell-cycle regulation and mitotic progression, and their coordinated activity reflects dysregulation of proliferative programs commonly observed in aggressive cancers [24]. Collectively, these genes participate in interconnected regulatory pathways rather than acting independently, supporting the relevance of grouped feature selection in capturing biologically meaningful gene modules. Such organization is consistent with the reported role of Forkhead Box transcription factors, particularly FOXM1, in coordinating proliferative and cell-cycle programs in cancer [25], [26].

Importantly, the recurrent selection of these genes does not imply causal biomarker validity. Rather, their consistent appearance supports the biological plausibility of the proposed stability-aware modeling framework. The inclusion of an L2 component in Elastic Net promotes grouped selection of co-regulated genes, enabling the model to capture biologically meaningful gene modules rather than arbitrarily selecting isolated predictors. Such behavior is particularly advantageous in genomic studies, where disease mechanisms are driven by interacting gene networks rather than single-gene effects.

Overall, this frequency-based analysis complements the stability metrics and coefficient path results by demonstrating that Elastic Net yields not only more stable feature sets but also more biologically interpretable and reproducible candidate biomarkers. These findings reinforce the value of incorporating stability-aware regularization strategies in high-dimensional genomic survival prediction and biomarker-oriented analyses.

Practical guidance for choosing α under different objectives.

The results provide explicit practical guidance for selecting the Elastic Net mixing ratio according to different modeling objectives. Rather than treating α as a purely technical tuning parameter, its selection should be aligned with downstream goals related to prediction, interpretability, and reproducibility.

a. Stability-oriented biomarker discovery

When reproducibility and consistent feature identification are primary objectives, lower to moderate values of α (approximately 0.2–0.5) are recommended. In this range, the L2 component remains sufficiently strong to preserve grouped selection of correlated genes, resulting in higher feature selection stability across resampling iterations.

b. Compact predictive models with balanced performance

For applications that require a reduced feature set while maintaining balanced classification performance (e.g., F1-score), intermediate values of α (around 0.5) offer a principled compromise between sparsity and stability.

c. Extreme sparsity and minimal feature sets

If interpretability is narrowly defined as selecting the smallest possible number of features, values of α close to 1.0 (LASSO) may be attractive. However, the results indicate that such configurations come at the cost of reduced selection stability and lower reproducibility of selected features.

Overall, these findings underscore that the choice of α should be explicitly aligned with the intended use case—prediction-focused versus biomarker-focused—rather than driven solely by predictive accuracy.

E. Methodological Implications

The findings of this study carry several important methodological implications for high-dimensional genomic modeling and feature selection using regularized regression. First, the results demonstrate that the L1/L2 mixing ratio in Elastic Net should not be treated as a secondary tuning parameter optimized solely for predictive performance. Prior methodological work has established that regularization choices directly influence sparsity and selection behavior, particularly under multicollinearity [3], [22]. The present study extends this understanding by empirically showing that the mixing parameter fundamentally shapes the trade-off between sparsity, stability, and interpretability along the regularization path.

Second, the observed decoupling between predictive performance and feature selection stability highlights a critical limitation of accuracy-driven model selection. Models with comparable AUC, accuracy, and F1-score can yield substantially different feature sets across resampling iterations. This finding reinforces earlier evidence that predictive accuracy alone is insufficient for evaluating feature selection methods when reproducibility is a primary objective [9], [10]. In genomic biomarker discovery, where selected features are often interpreted as biologically meaningful candidates, instability across resampling or studies can severely undermine scientific validity.

Third, the coefficient path analysis provides practical insight into the regularization dynamics underlying these trade-offs. The abrupt coefficient suppression observed under LASSO contrasts sharply with the smoother shrinkage and grouped selection behavior of Elastic Net. Such behavior is consistent with previous findings that LASSO exhibits instability in the presence of correlated predictors, often selecting arbitrary representatives from correlated groups [4]. In contrast, the grouping effect induced by the L2 component of Elastic Net promotes more reproducible selection patterns without eliminating sparsity entirely, thereby offering a principled compromise between interpretability and robustness.

Finally, the integration of stability analysis, sparsity evaluation, and predictive assessment along the regularization path represents a generalizable modeling framework that extends beyond the specific dataset analyzed in this study. Similar challenges related to high dimensionality, correlation structure, and reproducibility arise in other omics domains, including transcriptomics, proteomics, and metabolomics. Emphasizing stability-aware and path-wise analysis aligns with broader concerns regarding reproducibility and transparency in computational research [21]. By shifting emphasis from single-model optimization to systematic evaluation across the regularization path, the proposed approach contributes to more robust, interpretable, and reproducible regularized modeling practices.

Statistical interpretation of stability and regularization behavior.

From a statistical perspective, the observed stability patterns reflect the variability of the estimated support under perturbations of the data-generating process. In high-dimensional settings, the selected feature set should be regarded as a random object, and stability metrics such as Jaccard, Dice, and ARI provide empirical estimates of the variability of this support estimator. The decoupling between predictive performance and stability observed in this study highlights a fundamental distinction between risk minimization and support recovery: cross-validation optimizes predictive risk but does not control the variance of the selected support.

The L1/L2 mixing ratio plays a central role by modifying the geometry of the penalized likelihood. While LASSO induces sharp corners that favor aggressive sparsity, this geometry amplifies sensitivity to perturbations under correlated predictors, particularly when theoretical conditions for selection consistency are violated. In contrast, the inclusion of an L2 component smooths the penalization landscape, reducing abrupt coefficient thresholding and stabilizing support recovery. Importantly, the realized sparsity is jointly determined by the mixing ratio and the cross-validated regularization strength, underscoring that α should be interpreted as a structural parameter governing stability rather than a mere sparsity control.

These findings reinforce the need for stability-aware regularization strategies in high-dimensional inference, where reproducible feature selection is often as important as predictive accuracy.

F. Limitations

This study focuses on a single benchmark cohort (METABRIC) and a single outcome definition (overall survival status), which may limit direct generalizability to other genomic cohorts or clinical endpoints. In addition, feature selection was defined based on non-zero coefficients, which can be sensitive to numerical thresholds and optimization tolerance in high-dimensional settings. Although frequently selected genes were discussed to support biological plausibility, systematic pathway enrichment analysis and external validation were beyond the scope of this study and represent important directions for future work.

While the empirical analysis is restricted to a linear modeling framework and a single dataset, the observed stability patterns align with well-established theoretical properties of Elastic Net regularization. In particular, the monotonic decline in feature selection stability with increasing α reflects the reduced influence of the L2 penalty, which is known to promote grouped selection under correlated predictor structures. As such, the qualitative relationship between the L1/L2 mixing ratio and selection stability is expected to extend to other high-dimensional genomic settings characterized by strong correlations among features.

From a computational perspective, the proposed framework involves repeated resampling, cross-validated tuning of the regularization parameter, and evaluation across multiple values of the L1/L2 mixing ratio. In high-dimensional settings with tens of thousands of features, this design increases runtime and memory requirements, which may limit scalability in large-scale genomic studies or resource-constrained research environments. Although the current analysis is computationally feasible for the METABRIC cohort, applying the framework to larger cohorts or denser resampling schemes may require parallelization, warm-start strategies along the regularization path, or preliminary feature screening to improve efficiency. Future work will explicitly report computational time and resource usage to better characterize the practical cost of stability-aware regularization.

Nevertheless, the magnitude of stability gains and the optimal choice of α may vary depending on dataset-specific factors, including correlation structure, signal-to-noise ratio, outcome definition, and sample size. Future work will extend this framework to additional genomic cohorts, alternative endpoints, and non-linear modeling approaches to further assess the external consistency of the observed stability–sparsity trade-offs.

V. CONCLUSION

This study examined the impact of the L1/L2 mixing ratio along the Elastic Net regularization path on feature selection stability, sparsity, and predictive performance in high-dimensional genomic data. Using the METABRIC breast cancer dataset, the analysis demonstrated that variation in the mixing parameter α leads to systematic and interpretable differences in model behavior that are not captured by predictive accuracy alone.

The results show that lower values of α, corresponding to a stronger L2 component, consistently improve feature selection stability, while LASSO-dominated models produce more aggressive sparsity at the cost of reproducibility. Importantly, predictive performance metrics such as AUC, accuracy, and F1-score remain relatively stable across a wide range of α values. This decoupling between predictive performance and selection stability highlights the limitation of accuracy-driven model selection when the primary objective includes reliable and interpretable feature identification.

Analysis of coefficient paths further revealed fundamental differences in regularization dynamics between LASSO and Elastic Net. LASSO induces abrupt coefficient suppression, whereas Elastic Net exhibits smoother shrinkage and grouped selection of correlated genes. These dynamics explain why intermediate L1/L2 ratios, particularly moderate values of α, provide a favorable balance between sparsity, stability, and predictive performance.

From a biological perspective, the frequent selection of well-established breast cancer–related genes such as FOXC1, MKI67, CCNB1, and PLK1 supports the interpretability and translational relevance of the proposed framework. The ability of Elastic Net to retain co-regulated gene sets enhances confidence in the robustness of identified biomarkers and aligns with the network-driven nature of genomic regulation.

Overall, this study emphasizes that the L1/L2 mixing ratio should be regarded as a substantive modeling choice rather than a secondary tuning parameter. By integrating stability analysis, sparsity evaluation, and predictive assessment along the regularization path, the proposed approach advances stability-aware modeling practices and supports the development of more reliable, interpretable, and reproducible genomic prediction models.

ACKNOWLEDGEMENT

We extend our sincere gratitude to the lecturers of the Statistics and Data Science Study Program for their invaluable guidance, direction, and meaningful insights throughout this study. Their expertise and support served as a fundamental basis for this research. We also acknowledge the School of Data Science, Mathematics, and Informatics at IPB University for providing the necessary facilities and a supportive academic environment. Finally, we appreciate all those who offered constructive criticism and suggestions, which enabled the successful completion of this work.

REFERENCES

[1] G. C. Reinsel, R. P. Velu, and K. Chen, "High-Dimensional Reduced-Rank Regression," in *Lecture Notes in Statistics*, vol. 225, 2022, pp. 279–309. doi: 10.1007/978-1-0716-2793-8_10.

[2] J. Wang, P. Zhao, Y. Li, Y. Yang, and F. Chen, "Simulation study on variable selection method for high-dimensional biomedical data," *Journal of Xi'an Jiaotong University (Medical Sciences)*, vol. 42, no. 4, pp. 628–632, 2021, doi: 10.7652/jdyxb202104024.

[3] Z. T. Ke and L. Wang, "A Comparison Of Hamming Errors Of Representative Variable Selection Methods," in *ICLR 2022 - 10th International Conference on Learning Representations*, 2022. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85149719095&partnerID=40&md5=ec73609d963cf681a17ca598a85a075e

[4] R. Sarkar, S. Manage, and X. Gao, "Stable Variable Selection for High-Dimensional Genomic Data with Strong Correlations," *Annals of Data Science*, vol. 11, no. 4, pp. 1139–1164, 2024, doi: 10.1007/s40745-023-00481-5.

[5] Y. Yang and H. Yang, "Rates of convergence of the adaptive elastic net and the post-selection procedure in ultra-high dimensional sparse models," *Commun Stat Theory Methods*, vol. 50, no. 1, pp. 73–94, 2021, doi: 10.1080/03610926.2019.1628991.

[6] H. Ali, M. Shahzad, S. Sarfraz, K. B. Sewell, S. Alqalyoobi, and B. P. Mohan, "Application and impact of Lasso regression in gastroenterology: A systematic review," *Indian Journal of Gastroenterology*, vol. 42, no. 6, pp. 780–790, 2023, doi: 10.1007/s12664-023-01426-9.

[7] M. Hajihosseinlou, A. Maghsoudi, and R. Ghezelbash, "Regularization in machine learning models for MVT Pb-Zn prospectivity mapping: applying lasso and elastic-net algorithms," *Earth Sci Inform*, vol. 17, no. 5, pp. 4859–4873, 2024, doi: 10.1007/s12145-024-01404-5.

[8] O. Al Hosni and A. Starkey, "Stability and accuracy of feature selection methods on datasets of varying data complexity," in *2021 22nd International Arab Conference on Information Technology, ACIT 2021*, 2021. doi: 10.1109/ACIT53391.2021.9677329.

[9] T. Łukaszuk and J. Krawczuk, "Importance of feature selection stability in the classifier evaluation on high-dimensional genetic data," *PeerJ*, vol. 12, no. 11, 2024, doi: 10.7717/peerj.18405.

[10] Z. Mungloo-Dilmohamud, Y. Jaufeerally-Fakim, and C. Peña-Reyes, "Stability of Feature Selection Methods: A Study of Metrics Across Different Gene Expression Datasets," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, pp. 659–669. doi: 10.1007/978-3-030-45385-5_59.

[11] R. Sen, A. K. Mandal, and B. Chakraborty, "A Critical Study on Stability Measures of Feature Selection with a Novel Extension of Lustgarten Index," *Mach Learn Knowl Extr*, vol. 3, no. 4, pp. 771–787, Dec. 2021, doi: 10.3390/make3040038.

[12] J. M. Santos and M. Embrechts, "On the use of the adjusted rand index as a metric for evaluating supervised classification," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, pp. 175–184. doi: 10.1007/978-3-642-04277-5_18.

[13] O. Melikechi and J. W. Miller, "Integrated Path Stability Selection," *J Am Stat Assoc*, 2025, doi: 10.1080/01621459.2025.2525589.

[14] X. Li and R. Jacobucci, "Regularized Structural Equation Modeling With Stability Selection," *Psychol Methods*, vol. 27, no. 4, pp. 497–518, 2021, doi: 10.1037/met0000389.

[15] N. Sudjai, M. Duangsaphon, and C. Chandhanayingyong, "Adaptive Elastic Net on High-Dimensional Sparse Data with Multicollinearity: Application to Lipomatous Tumor Classification," 2024.

[16] N. Sudjai, M. Duangsaphon, and C. Chandhanayingyong, "Adaptive Lasso sparse logistic regression on high-dimensional data with multicollinearity," *Science, Engineering and Health Studies*, vol. 19, Feb. 2025, doi: 10.69598/sehs.19.25020002.

[17] D. Theng, K. K. Bhoyar, and P. Pawade, "Feature Selection and Stability Analysis using Ensemble Techniques," *Journal of Intelligent Systems and Internet of Things*, vol. 16, no. 1, pp. 41–48, 2025, doi: 10.54216/JISIoT.160104.

[18] A. S. Sumant and D. Patil, "Stability Investigation of Ensemble Feature Selection for High Dimensional Data Analytics," in *Lecture Notes in Networks and Systems*, 2022, pp. 801–815. doi: 10.1007/978-3-031-12413-6_63.

[19] A. Spooner, G. Mohammadi, P. S. Sachdev, H. Brodaty, and A. Sowmya, "Ensemble feature selection with data-driven thresholding for Alzheimer's disease biomarker discovery," *BMC Bioinformatics*, vol. 24, no. 1, Dec. 2023, doi: 10.1186/s12859-022-05132-9.

[20] R. Bertolini and S. J. Finch, "Stability of filter feature selection methods in data pipelines: a simulation study," *Int J Data Sci Anal*, vol. 17, no. 2, pp. 225–248, 2024, doi: 10.1007/s41060-022-00373-6.

[21] B. Antunes and D. R. C. Hill, "Reproducibility, Replicability and Repeatability: A survey of reproducible research with a focus on high performance computing," *Comput Sci Rev*, vol. 53, 2024, doi: 10.1016/j.cosrev.2024.100655.

[22] H. Zou and T. Hastie, "Regularization and Variable Selection Via the Elastic Net," *J R Stat Soc Series B Stat Methodol*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/J.1467-9868.2005.00503.X.

[23] S. Biyani, A. S. Patil, and V. Swami, "The Influence of FOXC1 Gene on Development, Organogenesis, and Functions," *Clinical and Translational Metabolism*, vol. 22, no. 1, 2024, doi: 10.1007/s12018-024-09297-0.

[24] X. T. R. Moore, L. Gheghiani, and Z. Fu, "The Role of Polo-Like Kinase 1 in Regulating the Forkhead Box Family Transcription Factors," May 01, 2023, *MDPI*. doi: 10.3390/cells12091344.

[25] D. Kalathil, S. John, and A. S. Nair, "FOXM1 and Cancer: Faulty Cellular Signaling Derails Homeostasis," Feb. 15, 2021, *Frontiers Media S.A.* doi: 10.3389/fonc.2020.626836.

[26] M. Li *et al.*, "FOXM1 transcriptional regulation," *Biol Cell*, vol. 116, no. 9, 2024, doi: 10.1111/boc.202400012.