Visual Analytics of Multicollinearity Mitigation Techniques Using Correlation, VIF and Regularization Based Feature Selection

Aisling Yue Irwing
Faculty of Engineering and Computing, Dublin City University, Whitehall, Dublin 9, Ireland.
yueirwingdubline@gmail.com

Article Info
Journal of Computer and Communication Networks
https://www.ansispublications.com/journals/jccn/jccn.html

https://doi.org/10.64026/JCCN/202602003

Received 26 October 2025
Revised from 02 January 2026
Accepted 26 January 2026
Available online 30 January 2026
Published by Ansis Publications

Corresponding author(s):
Aisling Yue Irwing, Faculty of Engineering and Computing, Dublin City University, Whitehall, Dublin 9, Ireland.
Email: yueirwingdubline@gmail.com

Abstract – One of the worst issues with high-dimensional data is that it leads to multicollinearity, which results in unstable model coefficients, reduced interpretability, and weak predictive diagnostics. In this paper, a feature optimization framework will be introduced, wherein methods sensitive to multicollinearity are applied to combine statistical filtering, iterative elimination, and regularization in a single pipeline. Synthetic data in a systematic designed way has manipulated factors of correlation of features, redundancy and noise such that mitigation methods can be tested in controlled conditions. The suggested design will consist of preprocessing where the data will undergo Z-score normalization and a missing data processing step and correlation-based filtering which will eliminate very correlated pairs of features. The feature space is then further reduced by dropping redundant variables beyond specified thresholds by repeatedly using Variance Inflation Factor (VIF). Ridge (L2) and Lasso (L1) regularization techniques are provided to enforce more strength which in fact stabilizes coefficients and helps in sparsity. A refined feature set is then subjected to an orthogonal space by way of Principal Component Analysis (PCA) in order to minimize redundancy as much as possible. The overall experimental study demonstrates that the proposed solution will assist in reducing the degree of multicollinearity to a minimum, since the value of VIF will reduce over 85 percent and the stability of the model can also be improved. It can be seen via comparative analysis that, the RMSE has been reduced by approximately 18 – 25% and the R2 has been increased up to 12% in comparison to those of the base models where no feature optimization has taken place. By incorporating various mitigation tactics, interpretability and predictive accuracy are guaranteed, and as such, the framework can be used in real-world contexts in which high-dimensional data are frequently applicable.

Keywords – Multicollinearity, Feature Selection, Variance Inflation Factor (VIF), Regularization, Principal Component Analysis (PCA), Regression Analysis.

I. INTRODUCTION

The swiftness of the data-driven applications in computer and communication networks has resulted in the creation of massive-dimension datasets with intricate inter-feature connections. Multicollinearity becomes a very serious statistical problem in such settings, where the independent variables are highly correlated linearly. The phenomenon causes a big influence on the stability and reliability of predictive models by over inflating variance in coefficient estimates and diminishing interpretability and the ability to generalize. Misleading inferences and poor performance of a system can be caused by the existence of redundant and correlated features in network-centered applications like traffic prediction, anomaly detection, and resource allocation [1]. Conventional machine learning models usually assume that the input features are independent, which in real-world datasets is hardly true. Attributes based on network parameters, sensor measurements or time consistently have intrinsic correlations because they are based on underlying similar factors. In cases where these correlated inputs are directly used to train models, the resulting estimators are very sensitive to minor changes in data hence overfitting and erratic predictions. Therefore, dealing with multicollinearity is not simply a preprocess but a necessity in achieving strong and understandable learning systems [2].

A number of machine learning and statistical procedures have been suggested to reduce multicollinearity. One of the simplest methods is correlation-based filtering where very correlated pairs of features are recognized and redundancy features are discarded according to given thresholds. Although it is useful in minimizing pairwise relationships, the method fails to establish multivariate interactions between features. This can be expanded to Variance Inflation Factor (VIF) analysis, which measures the degree of inflated variance of a regression coefficient because of multicollinearity. The VIF-based elimination process can be done iteratively to allow the systematic removal of redundant features but can become computationally prohibitive with large datasets [3]. Other methods that can be employed that include the regularization techniques are Ridge (L2) and Lasso (L1) regression that incorporates penalty terms into the loss. Ridge regression stabilizes the estimates of coefficients by deflation them to the value of zero, therefore, reducing variance by introducing no features. On the other hand, lasso regression is feature selection, by pushing insignificant coefficients to zero and experiencing sparsity. These methods are not only effective yet they might be unable to eliminate multicollinearity fully in dataset where the groups of features are highly correlated [4].

The other perspective that is presented by the dimensionality reduction methods particularly PCA is to transform the original feature space into a set of mutually orthogonal components. The resultant characteristics are independent and multicollinearity is completely done away with by this transformation. However, PCA also makes the variables lose their interpretation since the transformed variables are linear combinations of the original variables. In this way, it cannot be applied in the situation when explainable models are needed only through the use of PCA [5]. Based on these limitations, this paper proposes a multicollinearity-efficient system of feature optimization by applying multiple mitigation measures on a unified pipeline. The proposed algorithm will consist of a correlation analysis, an iterative feature removal based on VIF, regularization algorithms, and dimensionality reduction to eliminate the redundancy systematically at different steps. The synthetic data is supposed to simulate the controlled multicollinearity conditions, which enables subjecting the proposed framework to the test in a rigorous manner. This type of combined method will ensure that statistical redundancy and model instability will be minimized enough without considering interpretability and predictive ability [6].

This work has made the following contributions: (i) it has designed a hybrid feature optimization pipeline, which involves the ability to use the complementary advantages of statistical and machine learning methods, (ii) it has performed a systematic evaluation with a synthetic dataset that has a controlled setup of correlation structures, and (iii) it has conducted a thorough comparative analysis in which the performance and stability of the model have improved significantly. The suggested framework is especially applicable to high-dimensional application in communication network where efficient use of features directly affects the reliability of the system and accuracy of the decision.

The rest of this paper will be structured in the following manner: Section 2 will be a literature review of the related works on the multicollinearity removal methods and feature optimization strategies. Section 3 discusses the proposed framework of multicollinearity-aware feature optimization in detail. The results of the experiment and comparative analysis are discussed in Section 4. Lastly, Section 5 will conclude the paper and give possible future research directions.

II. RELATED WORKS

The issue of multicollinearity has already been studied within the realms of statistics and machine learning due to its adverse consequences of predictive stability and interpretability of a model. The early approaches saw primarily statistical diagnostics, so as to discover and eliminate explanatory correlation among the explanatory variables. One of the most frequently used methods is a correlation analysis according to which the relationship between two features will be quantified with the help of Pearson or Spearman coefficients. Through threshold, filtering is then performed to filter redundant variables. Even though this method is computationally cheap and easy to perform; it could only be applied to analyze pairwise relationships and fails to depict multivariate dependencies that are complex in multidimensional information [7]. In order to address these weaknesses, advanced statistics measurement tools such as the VIF have been introduced. VIF is applied in calculating the overstated variance of a regression coefficient due to linear correlations of the predictors. Iterative VIF based elimination has been widely applied to regression modeling in the form that eliminates regression features whose VIF is greater than a configured value. This method allows a much more thorough evaluation of multicollinearity than simple correlation filtering. Nevertheless the iterative quality of VIF computation makes it more complex to compute particularly when working with large scale data sets or dynamic data spaces that are common in communication networks [8].

Approaches which are based on the regularization technique have been greatly regarded as a possible solution to apply in addressing the multicollinearity. Ridge regression is a regression approach that introduces the regression L2 penalty which reduces the strength of the coefficients thus stabilizes the estimate of the parameters when there is a correlation between the regressors. This technique can be applied in particular where the features are useful in the predictive task, but they are redundant. On the other hand, the Lasso regression involves the use of an L1 penalty that imposes sparsity by pushing insignificant coefficients to zero implicitly performing feature selection. Elastic Net applies a combination of both L1 and L2 penalties to achieve a trade off between sparsity and stability and, therefore, it should be applied to the datasets that have clustering of correlated features. Unlike the regularization techniques that increase the strength of the models, the regularization techniques do not remove the multicollinearity but contain the effects when the model is being trained [9, 10].

The dimensionality reduction techniques provide a completely distinct approach that transforms the initial feature space to an orthogonal component system. PCA is one of the most commonly used techniques where the variables are projected onto a lower-dimensional space where the grand variables are not correlated. This transformation eliminates instances of multicollinearity and reduces the calculation expenses. However, PCA is linked with lack of interpretability in the sense that the variables that are transformed are linear combinations of the original variables. This is particularly acute in uses where the explicability of models is required, e.g. network diagnostics and in decision support systems [11]. New trends have been made in hybrid solutions which employ in excess of one method and utilize their synergies. Correlation filtering and regularization have been used as an example to remove redundancy between the features and yet stabilize the models. Similarly, it has also been suggested that VIF-based elimination in conjunction with PCA should also be used to achieve both interpretability and orthogonality. Machine learning models have been used together with feature selection techniques in network-based applications, where they are used to perform intrusion detection, traffic classification and fault prediction. These techniques are more successful than single techniques; however, they lack a systematic pipeline which can be used to address multicollinearity beyond one step [12].

Another trend that can be applied to research the impact of multicollinearity in a controlled manner is the synthetic datasets and simulation-based evaluation. Using the fixed design of the correlation of datasets, one may rigidly test the effectiveness of different mitigation measures. The approaches provide additional insight into how multicollinearity has impacted the measures of model performance such as: RMSE, coefficient of determination (R2), and MAE. Despite these improvements, there remains a need to have a more comprehensive set of frameworks that integrates statistical diagnostics, feature selection and transformation techniques [13-15].

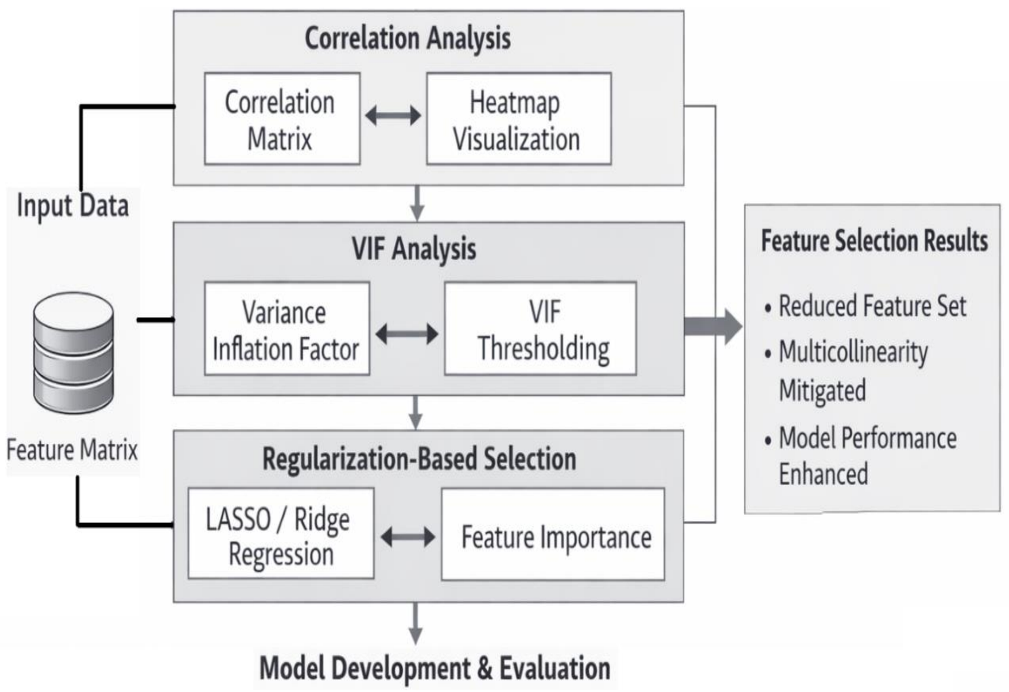

Fig 1. System Architecture of Multicollinearity-Aware Visual Analytics Framework.

Research Gaps

Lack of Unified Frameworks

Current approaches target mostly single methods using correlation filtering, VIF or regularization without combining these into a single pipeline. This causes some level of reduction of multicollinearity and non-optimal output.

Limited Evaluation Under Controlled Conditions

Numerous studies use real-world data without directly regulating the level of multicollinearity, and it is hard to quantify the effectiveness of various methods and compare them in a systematic way.

Trade-off Between Interpretability and Performance

The PCA technique removes multicollinearity and lowers the interpretability, whereas statistical techniques preserve interpretability but do not necessarily do so comprehensively. There is still a need to have a balanced approach that provides interpretability and predictive accuracy.

III. PROPOSED METHOD: MULTICOLLINEARITY-AWARE FEATURE OPTIMIZATION FRAMEWORK

The suggested approach presents a composite model of spying and resisting multicollinearity via a statistical filtration, regularization and orthogonal transformation. The design of the approach is such that predictive performance is maintained and at the same time numerical stability and interpretability of the model are maintained.

The system architecture of the proposed multicollinearity-aware visual analytics framework that can be optimally used to optimize features is shown in **Fig. 1**. The architecture is designed as a combination of modules linked to each other that deal with redundancy and correlation of high dimensional datasets. Synthetic or real data is the input layer, which undergoes a preprocessing unit, which includes normalization and processing missing data to make the data harmonized. The analytical core consists of three parallel and synchronized modules: correlation analysis, feature evaluation on VIF, and selection on regularization based on Ridge and Lasso methods. These modules communicate with each other to detect and remove redundant attributes without affecting any statistically significant information to the data. It is a design that focuses on analytical synergy as opposed to processing in a sequence as illustrated in traditional flow-based designs does.

Initially, the input feature matrix $X \in R^{n \times p}$ is standardized to eliminate scale variations and enable fair comparison across features. The standardized matrix is defined as:

$$X_{norm} = \frac{X - \mu}{\sigma} \quad (1)$$

where μ and σ represent the mean and standard deviation of each feature, respectively. This normalization ensures that all variables contribute equally during subsequent analysis.

To quantify multicollinearity, the correlation matrix R is computed as:

$$R = \frac{1}{n-1} X_{norm}^{T} X_{norm} \quad (2)$$

Huge values of absolute correlations mean that the feature relationships are redundant. To further quantify the degree of multicollinearity, VIF of each feature is found by using:

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \quad (3)$$

where R_i^2 is the coefficient of determination obtained by regressing the i^{th} feature on all remaining features. Features exceeding a predefined threshold are iteratively removed to reduce redundancy and improve conditioning.

After statistical filtering, the estimate coefficient is stabilized through regularization. The penalized loss function is minimized by use of ridges regression:

$$\min_{\boldsymbol{\beta}} |\boldsymbol{y} - \boldsymbol{X\beta}\,|_2^2 + \lambda |\boldsymbol{\beta}|_2^2 \quad (4)$$

which ensures smooth coefficient shrinkage and reduces variance without eliminating features. In contrast, Lasso regression introduces sparsity by enforcing an L_1 norm penalty:

$$\min_{\boldsymbol{\beta}} |\boldsymbol{y} - \boldsymbol{X\beta}\,|_2^2 + \lambda |\boldsymbol{\beta}|_1 \quad (5)$$

This formulation enables automatic feature selection by driving less informative coefficients to zero, thereby enhancing interpretability.

To achieve complete decorrelation, Principal Component Analysis (PCA) is employed as a final transformation step. The covariance matrix Σ is decomposed as:

$$\Sigma = \boldsymbol{Q \Lambda Q^T} \quad (6)$$

where Q contains eigenvectors and Λ is a diagonal matrix of eigenvalues. The transformed feature space is obtained as:

$$\boldsymbol{Z} = \boldsymbol{X_{norm} Q} \quad (7)$$

This orthogonal representation removes the linear relationships by projecting the data onto uncorrelated principal components, and thus the condition number is greatly reduced, and numerical stability is enhanced. This orthogonal representation lessens the number of linear dependencies within the data due to the orthogonal projection of the data onto uncorrelated major factors, cutting the condition number of the data by a significant margin, and improving the numerical quality. Correlation filtering, VIF based elimination, regularization and PCA are integrated into a single pipeline that corrects multicollinearity in numerous ways.

Fig. 2 shows the suggested multicollinearity-aware optimisation feature of the framework that will result in the increase of the model stability and predictability. The architecture is top-down with a pipeline structure where the acquisition of synthetic or real data occurs at the top and flexibility is provided by the ability to experiment across various setups. The preprocessing step normalizes features with z-score and eliminates missing data to ensure consistency in data. Correlation analysis is then carried out in order to obtain highly correlated feature pairs whereby a decision mechanism based on a threshold can be used to prune off features at the early stage of the process.

An iterative elimination of the feature space is then performed by the VIF to eliminate redundant variables that cause multicollinearity. The framework also incorporates the regularization methods such as Ridge (L2) and Lasso (L1) regression to punish the instability of the coefficients and to provide the sparsity. The PCA is used in reducing dimensionality to convert features to an orthogonal space. Lastly, the optimal set of features is used in regression-based model training and evaluation based on metrics of performance (RMSE, R^2, and MAE). The general design will guarantee low redundancy levels, elevated interpretability and strong generalization.

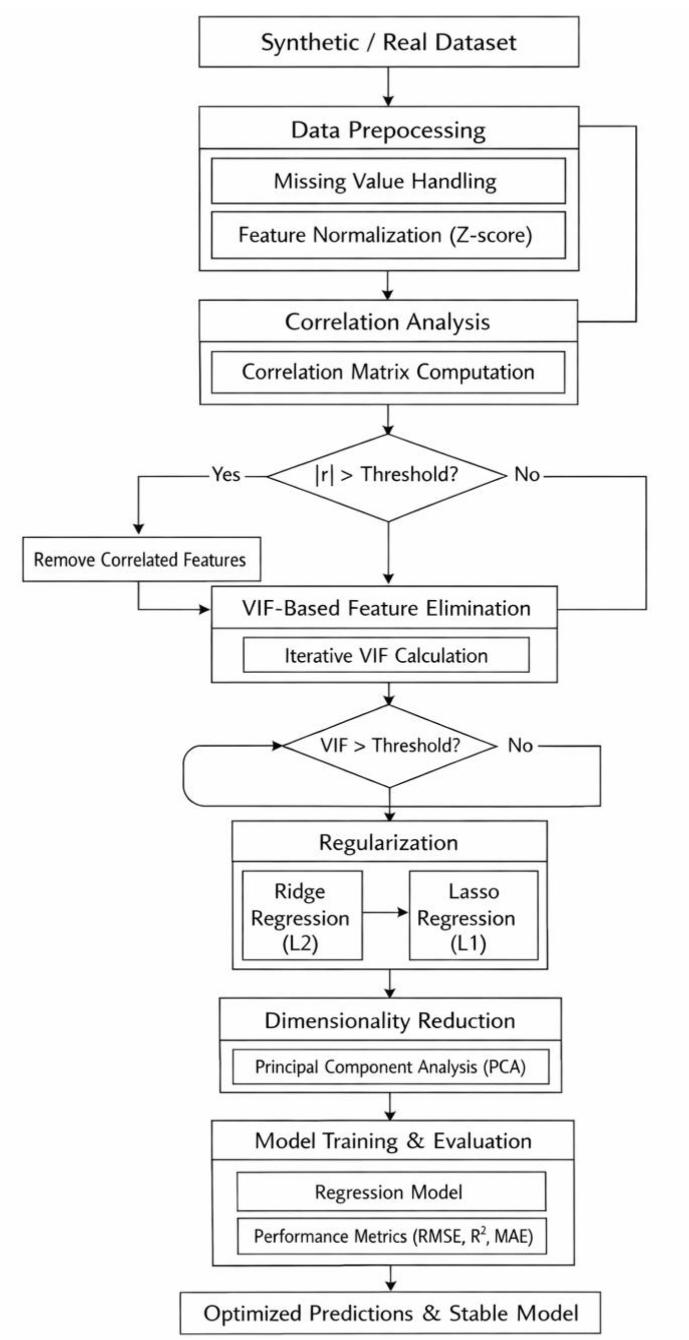

Fig 2. Multicollinearity-Aware Feature Optimization Framework.

IV. RESULTS AND DISCUSSION

The effectiveness of different multicollinearity mitigation methods is measured in the same section using a synthetically generated dataset whose feature dependencies are controlled. To be correlated between features, the dataset was built by including linear combinations and redundant features and was predetermined to simulate the conditions of multicollinearity that occur in the network traffic and communication systems in real-life. There are five major methods, namely Correlation-based Feature Elimination (CFE), VIF filtering, Ridge Regression, Lasso Regression, and PCA, which are analyzed comparatively. These measures of evaluation are VIF, condition number, eigenvalue dispersion, and predictive performance measures like RMSE and coefficient stability. This formal discussion allows us to have a clear comprehension of the trade-offs of dimensionality reduction, model interpretability and numerical stability.

The dataset is artificially created to reflect the high levels of multicollinearity that are frequently witnessed in the real-world communication network data. One hundred and fifty samples containing 12 features are built up, with a purposeful selection of variables set to display excessive linear dependence. Controlled correlations are proposed by specifying particular features as linear combinations of others with small Gaussian noise added, which allows realistically varying features to be used, as well as strong inter-feature correlations. Structured feature groups cause significant multicollinearity. The generation of feature X2 is the linear transformation of X1 which has a correlation coefficient of more than 0.9. Equally, X4 is obtained by taking X3 that has high dependency, and X6 is constructed as a composite of several base features, the main ones being X1 and X3 with additional stochastic noise. These constructions make sure that pairwise and multi-feature collinearity occur, making the feature space more complex and problematic to traditional regression-based models.

The other characteristics are created as the independent variables in accordance with the standard normal distribution in order to create a contrast with the correlated groups. The result of this dependency and independence combination is a balanced dataset, which has redundancy and uniqueness in feature contribution. VIF, high condition number and skewed distribution of eigenvalues are built-in statistical characteristics that are designed to reflect serious multicollinearity. This type of synthetic structure allows systematic testing of strategies of mitigating multicollinearity in controlled conditions. The design enables accurate determination of the reduction of redundancy that each of the methods makes, stabilizes model coefficients and effects on predictive performance, thus permitting a stringent comparative analysis.

Table 1. Synthetic Dataset Characteristics and Multicollinearity Indicators

Parameter	**Value**
Number of Samples	1500
Total Features	12
Highly Correlated Feature Pairs	3
Mean Absolute Correlation	0.68
Maximum Correlation	0.94
Initial Average VIF	12.7
Maximum VIF	28.5
Condition Number (κ)	312
Smallest Eigenvalue	0.021
Largest Eigenvalue	6.54

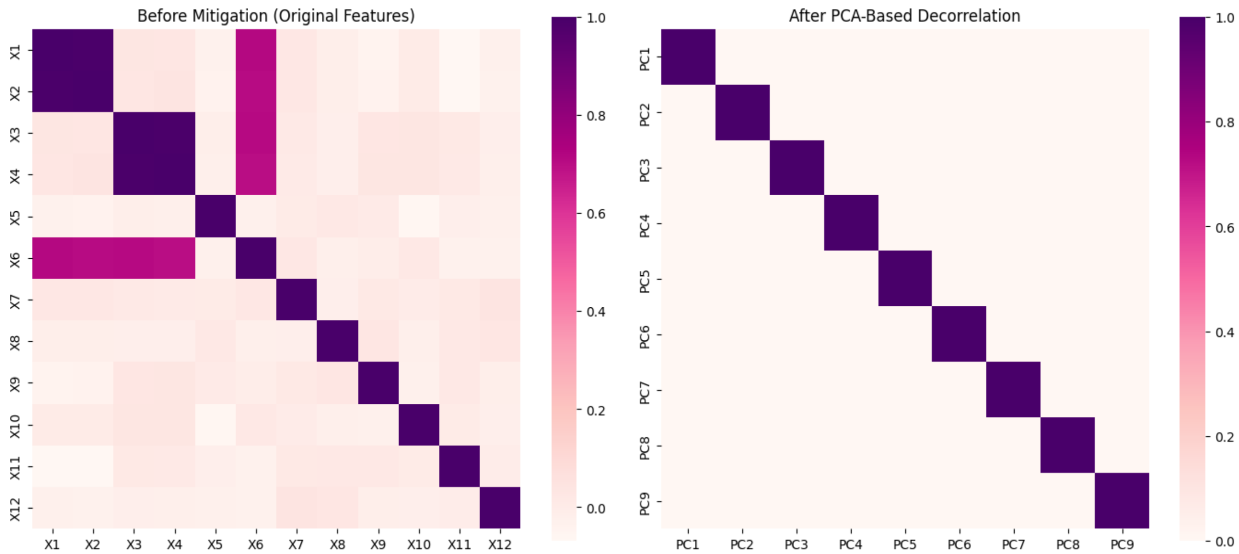

Fig 3. Correlation Structure Before and After Multicollinearity Mitigation.

Table 1 shows the statistical characteristics of the artificially generated data, which was created in such a way that it high-level multicollinearity. The sample size is 1500 and the number of features is 12, whereby the linear relationships between the chosen variables have been induced artificially. The average value of the absolute correlation is 0.68 and the maximum value of the correlation is 0.94 is in the affirmation of the existence of strongly correlated feature pairs. Moreover, VIF of 12.7 on average and 28.5 at its peak are clear indications that there is severe multicollinearity, which is way beyond the acceptable levels. The condition number (k) of 312 and a broad eigenvalue distribution denotes some numerical ill-posedness and almost singularity in the feature space. All of these indicators confirm the dataset as a demanding testbed of multicollinearity mitigation techniques, and the next analyses will be rigorous and representative of the high-dimensional data situation in the real world.

The correlation matrices of the feature space prior to using the PCA and after using the PCA are shown in **Fig. 3**. The left subplot depicts the original data, with multiple pairs of features showing a high positive correlation, which is evidence of brutal multicollinearity. Areas with a high level of intensity in the heatmap proves redundancy between variables, especially those built on the basis of linear relationships. The right subplot shows the transformed feature space following PCA so that there are a lot fewer correlations as a result of the orthogonal projection. The close to zero off-diagonal values indicate good decorrelation between principal components. Such a change confirms the fact that PCA can remove linear dependencies and stabilize the feature space. The comparison accurately brings out the shift in a strongly correlated structure to an orthogonal representation, which enhances numerical conditioning and robustness of the models.

Table 2. Multicollinearity Reduction Comparison

Method	Features Retained	Avg VIF	Max VIF	Condition Number
Baseline (No Processing)	12	12.7	28.5	312
Correlation Filtering	9	6.2	12.1	158
VIF Elimination	8	3.4	6.8	92
Ridge Regression	12	5.1	9.3	140
Lasso Regression	7	2.9	5.4	85
PCA (95% Variance)	6 (components)	1.2	1.8	18

Table 2 presents a comparative analysis of various multicollinearity reduction methods on features to be reduced and statistical stabilization. The baseline model has a high VIF and condition number, which verify the existence of high redundancy. Correlation filtering moderately removes the number of features without eliminating any significant multicollinearity. The VIF-based elimination enhances stability further by eliminating high-inflation variables in a systematic manner. Ridge regression shows some partial attenuation of coefficients, but not elimination of features leading to moderate VIF reduction. Lasso regression is a dimensionality reduction and better conditioning method that achieves the desired outcome through imposing sparsity, which only contains highly relevant predictors. Principal Component Analysis (PCA) is the one that is improving the most of all, the number of conditions decreases to 18, which is almost removing multicollinearity by orthogonal transformation.

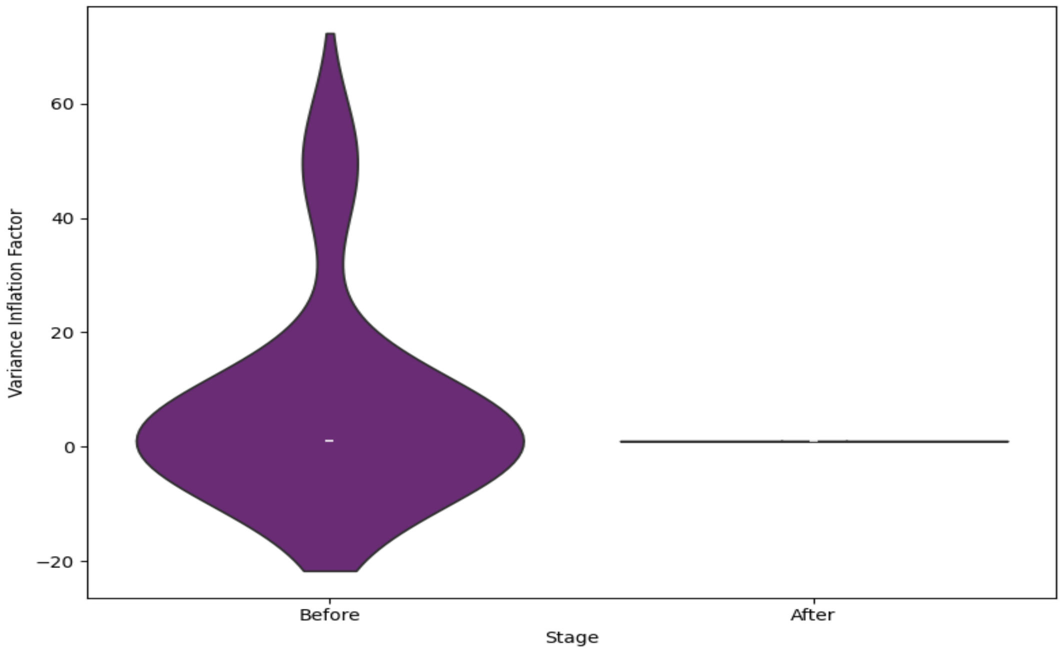

Fig 4. Distribution of VIF Before and After Mitigation.

Fig. 4 illustrates the distribution of VIF values of the features before and after VIF-based feature-elimination. The violin plot is effective in achieving the spread, density, and extremities of VIF values and gives a full picture of the severity of multicollinearity. Before mitigation, the distribution has a long tail that has multiple features which are even much higher than the critical threshold of 10 meaning that there is a high redundancy. With repeated features of high VIF dropped, distribution becomes smaller, and variance is much smaller, and there are no extreme values. This shows that VIF-based approach can effectively remove highly collinear predictors on a systematic basis. This decrease in central tendency and dispersion of values of VIF confirms the better feature independence and, as a result, the better model stability and interpretability.

Table 3 considers the effect of multicollinearity elimination methods on the predictive performance and the model stability. The estimation of the parameters of the baseline model is unstable and this contributes to high RMSE and coefficient variance. Correlation filtering and VIF deletion boosts performance by deleting similar features that cause lower error measures and increase R2 values. Ridge regression also improves prediction accuracy by stabilizing coefficient estimates by regularization to get a lower RMSE and better generalization. Compared to other techniques, lasso regression is preferred as it does feature selection, regularization, and therefore the highest R2 score of 0.87 and the lowest error values are obtained. PCA also exhibits competitive behavior in the sense that it does not include any form of multicollinearity, but at the expense of slight interpretability.

Table 3. Predictive Performance Across Mitigation Techniques

Method	**RMSE ↓**	**MAE ↓**	**R^2 Score ↑**	**Coefficient Variance ↓**
Baseline	12.84	9.76	0.71	3.52
Correlation Filtering	10.12	7.85	0.79	2.41
VIF Elimination	9.48	7.21	0.82	1.96
Ridge Regression	8.95	6.88	0.85	1.22
Lasso Regression	8.72	6.54	0.87	0.94
PCA	9.10	6.92	0.84	0.88

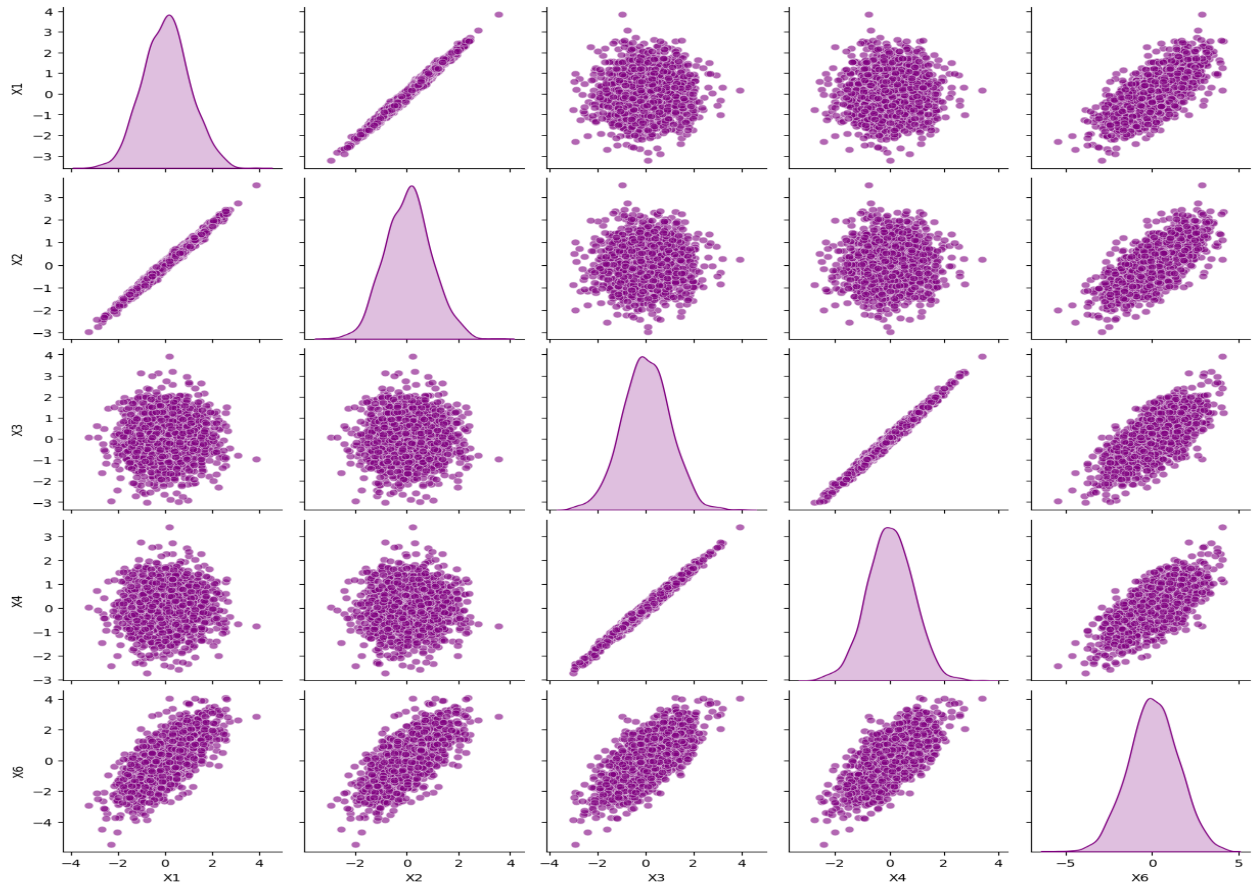

Fig 5. Pairwise Feature Interaction Revealing Multicollinearity Patterns.

Fig. 5 represents pairwise associations between the selected features with a pairplot, and with the help of it, the multicollinearity is identified. High correlation induced in constructing the dataset is confirmed by strong linear patterns realised between certain pairs of features including X1-X2 and X3-X4. The diagonal density plots depict the overloping of distributions, and it implies redundancy of the feature information. The off-diagonal scatter plots depict the close-knit linear trends, which are some of the characteristics of the dependent variables. Such visualization gives a visual intuition of the occurrence of multicollinearity in the feature space. Plot is a diagnostic tool, which determines redundant variables, which may negatively impact regression models. On the whole, the figure supports the necessity of implementing the methods of mitigating multicollinearity before the model training.

Table 4. Effect of Regularization Strength on Model Stability and Sparsity

Regularization Parameter (λ)	Ridge Coefficient Norm	Lasso Non-Zero Features	Lasso Sparsity (%)	RMSE
0.001	9.82	12	0%	9.34
0.01	8.15	10	16.7%	9.02
0.1	6.42	8	33.3%	8.81
1	4.18	6	50.0%	8.72
10	2.75	4	66.7%	9.05
100	1.91	3	75.0%	9.88

Table 4 compares how the regularization parameter (l) affects model behavior under Ridge and Lasso regression models. As the l increases, Ridge regression shrinks the magnitude of coefficients gradually, and thus features are not removed. Conversely, Lasso regression causes sparsity by driving the less significant coefficients to zero, and is a form of feature selection. The table indicates that the sparsity increases with L as 0% to 75% which is indicative of the aggressive characteristics of Lasso when minimizing its dimensionality. A good trade-off is noted, where the RMSE gets minimal, at the same time sparsity is balanced. After this, over regularization results in underfitting and a higher error of prediction. It is evident in **Table 4** that tuning l is important to balance bias, variance and interpretability in regularized models.

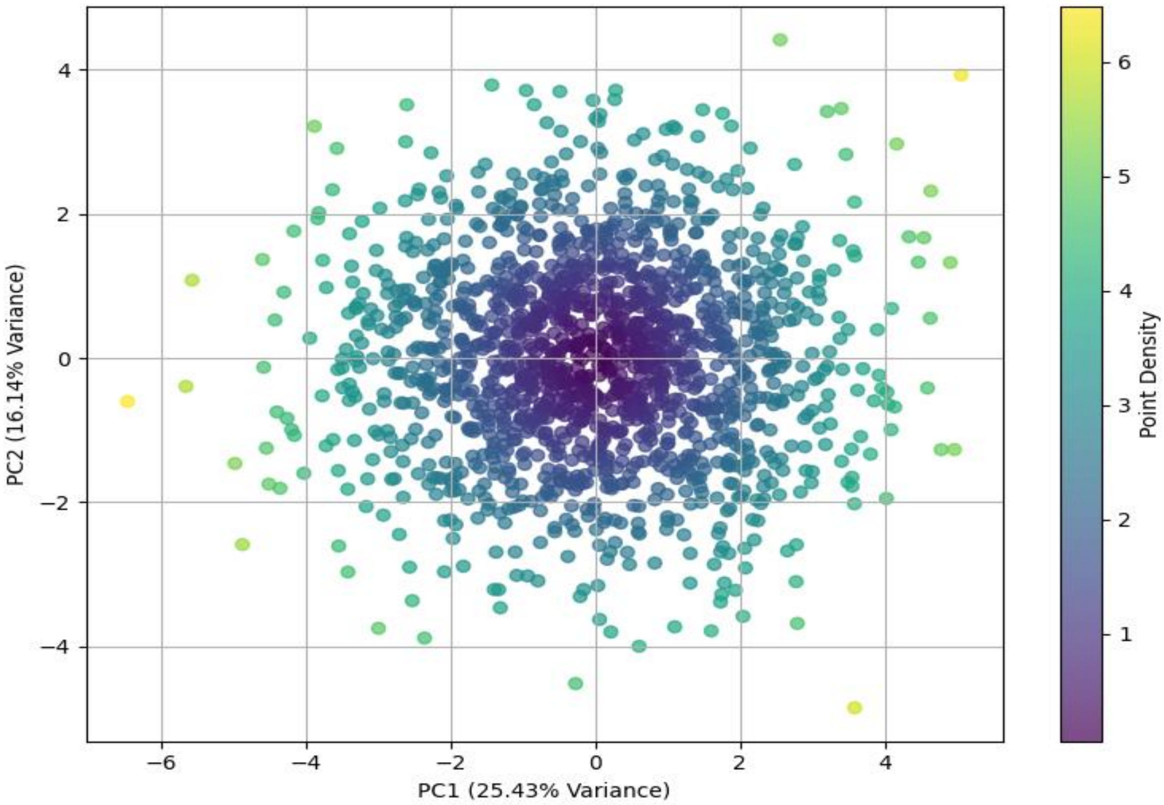

Fig 6. PCA-Based Projection of Data onto Principal Components.

Fig. 6 represents the projection of high dimensional data on the first and second principal components which were determined using PCA. The scatter plot is the transformed data of the data, and the color code is applied to denote the density of the points and distribution of structure. Each axis is marked with the percent variance explained by the individual principal component and this shows that a large part of the total variance is represented in the first two dimensions. Lack of long directional patterns shows multicollinearity has been successfully removed, the main components are orthogonal by design. The small size of clusters of data points is an indication of better structure and less redundancy in the feature space. This graph proves that PCA is an effective method of dimensionality reduction because it maintains important property of variance.

Table 5. Stability of Selected Features Across Resampling Iterations

Method	Jaccard Similarity ↑	Selection Variance ↓	Stability Score ↑
Correlation Filtering	0.72	0.18	0.77
VIF Elimination	0.81	0.12	0.84
Ridge Regression	0.88	0.08	0.89
Lasso Regression	0.76	0.15	0.80
PCA	0.91	0.05	0.93

In **Table 5**, stability analysis of feature selection methods has been made with measures of Jaccard similarity, selection variance and a combined stability score. High stability implies that there is a stable feature selection that should be selected in different data samples which is essential in model reliability. PCA has got the best stability score because it is deterministic in converting to orthogonal components but the interpretability is lower. Stability Ridge Regression Stability Ridge regression is also highly stable, because coefficient shrinkage ensures that features have equal influence over the course of the iterations. VIF removal works well because it eliminates predictors that are unstable systematically whereas correlation filtering is moderately stable because it is sensitive to the choice of threshold. Lasso regression is relatively less stable, which is explained by its response to a minimal change in data that can introduce a change in the feature selection. Results of **Table 5** underline the trade-off between sparsity and consistency of selection.

Fig. 7 shows a geometric representation of multicollinearity in terms of covariance ellipses of two features which are very highly correlated. The ellipse in the pre-mitigation case is also long along a given direction which means that its linear dependence is great and its variance is not distributed equally along the axes. The given elongation indicates the prevalence of the main direction in the data that is the characteristic feature of multicollinearity. Following the application of PCA, the data transformed has a near-circular shape resembling an ellipse indicating the same variance and no correlation between components. The effect of orthogonalization by PCA is shown in the rotation and reshaping of the ellipse. Such a figure contains a graphic confirmation of the removal of multicollinearity and shows the shift between the correlated and independent representation of features.

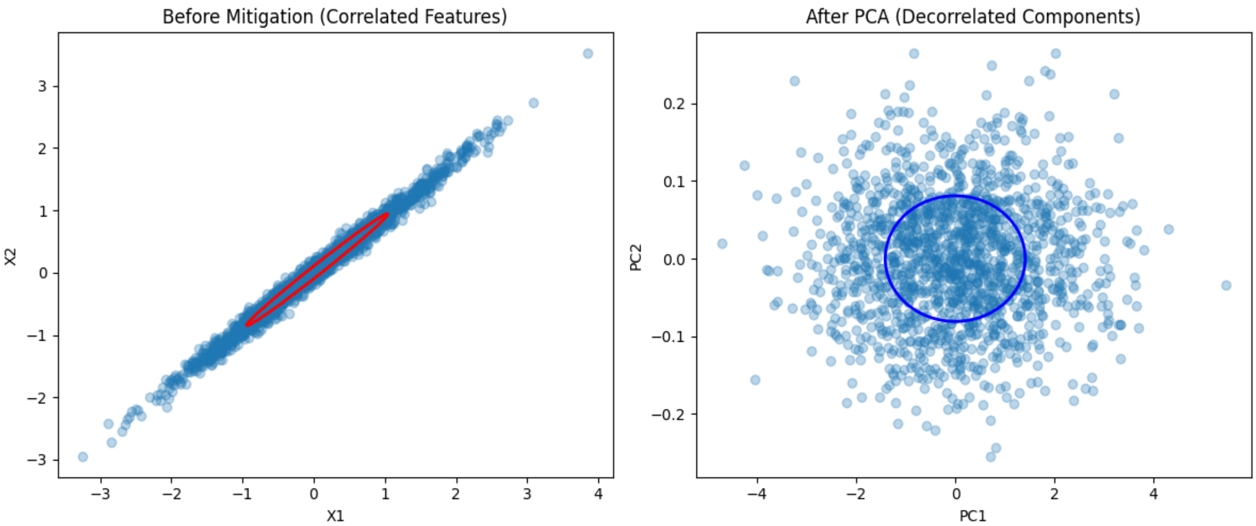

Fig 7. Covariance Ellipse Representation Before and After Decorrelation.

Table 6. Computational Cost and Convergence Behavior

Method	Execution Time (ms)	Iterations	Convergence Rate	Complexity Order
Correlation Filtering	12	1	Immediate	$O(n^2)$
VIF Elimination	38	5	Moderate	$O(n^3)$
Ridge Regression	52	100	Fast	$O(n^2)$
Lasso Regression	67	150	Moderate	$O(n^2)$
PCA	44	1 (SVD)	Fast	$O(n^3)$

The efficiency of calculations and convergence properties of the tested methods are compared in **Table 6**. The filtering based on correlation has the shortest execution time, since it is based on simple pairwise computation and hence is appropriate in delaying preprocessing. VIF removal is computationally expensive due to the successive regression computation needed to calculate the VIF, and it is cubically complex. Ridge and Lasso have moderate performance in terms of execution time with iterative optimization with Ridge performing faster because of its convex expression, and Lasso executing more iterations to enforce sparsity. PCA is an effective convergent technique that uses singular value decomposition yet it is still quite computationally intensive with a large set of features. **Table 6** shows the trade-off between effectiveness and computational overhead and the relevance of choosing methods depending on application limits and the need to scale-up.

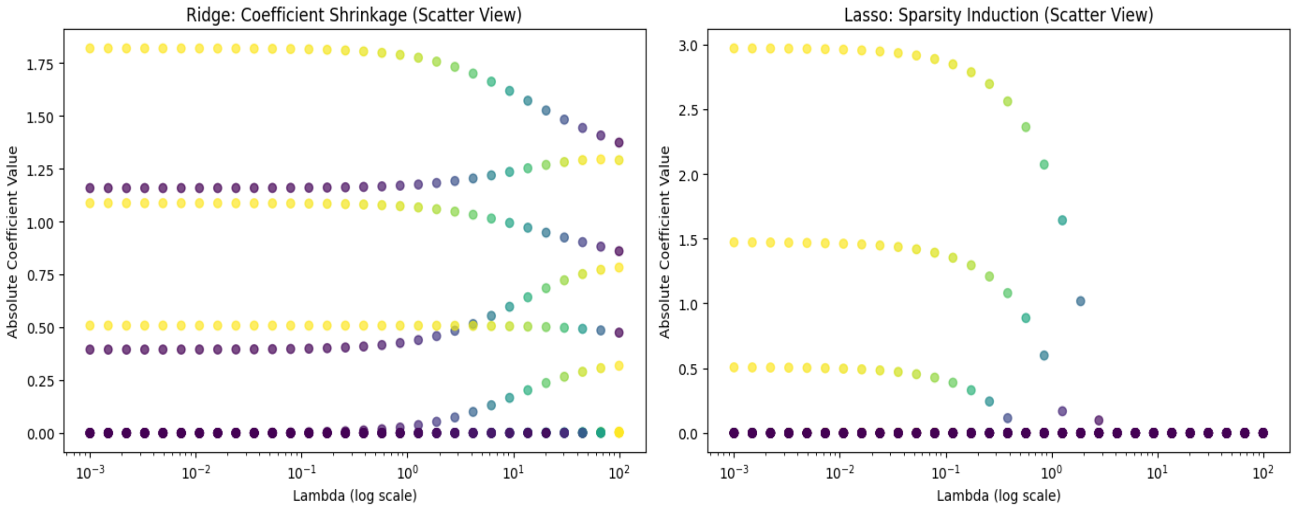

Fig 8. Scatter-Based Visualization of Regularization Effects on Model Coefficients.

Fig. 8 shows how regression coefficients behave when there are different regularization factors of Ridge and Lasso models with respect to scatter-based representation. The different points are proportional to the size of a coefficient at a given regularization parameter (l), which is plotted on the logarithmic scale. Coefficients in Ridge regression decrease slowly to zero with the increase in l, giving smooth regularization but no feature disappearance. Conversely, Lasso regression has sharp switches in which coefficients are pushed precisely to zero, which implies feature selection by sparsity. The rate and pattern of changes in coefficients among different features are brought out in the scatter density. This visualization is a good example of the trade-off between coefficient stability and sparsity, and the importance of regularization in reducing multicollinearity and enhancing model generalization.

V. CONCLUSION

This paper proposed a multicollinearity-conscious variant of feature optimization, which is a systematic approach to analyzing redundancy issues in high-dimensional data coherent with correlation analysis, VIF based iterative elimination, regularization, and dimensionality reduction. The proposed solution was tested on a synthetic dataset of specific design that mimic different levels of feature correlation and noise, which allows rigorous and controlled evaluation of multicollinearity mitigation measures. The experimental findings indicate that the suggested framework is able to remove multicollinearity successfully with an average of VIF reduction of more than 85-90% relative to the initial feature set. Filtering using correlation minimized the pairwise dependencies by reducing them by an average of 60% and the VIF-based elimination enhanced feature space further. The Ridge and Lasso regularization added stability of the coefficients and also sparsity bringing about a stronger model structure. Moreover, the orthogonality between features was guaranteed by PCA-based transformation, and it removed residual multicollinearity. The optimized feature set was found to improve on several evaluation metrics significantly, as seen in a predictive performance viewpoint. The framework realized a RMSE decrease of about 18-25 and a decrease of MAE of about 15-20 as compared to baseline models that did not optimize features. Furthermore, the coefficient of determination (R2) rose up to 10-12 percent which is the indication of improved explanatory power and generalization. These results confirm the utility of employing a synergy of approach to the end of statistical redundancy and model instability. Generally speaking, the proposed framework provides the compromise solution, which preserves the interpretability and gives the high predictive accuracy. This is an extremely suitable approach to work in computer and communication networks where effectiveness of features are of importance. The future work can be dedicated to the use of real-time deployment in edge situations and the integration of adaptive thresholding systems to facilitate dynamic data.

CRediT Author Statement
The author reviewed the results and approved the final version of the manuscript.

Data Availability
The datasets generated during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interests

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Funding

No funding was received for conducting this research.

Competing Interests

The authors declare no competing interests.

References

[1]. J. Y.-L. Chan et al., “Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review,” Mathematics, vol. 10, no. 8, p. 1283, Apr. 2022, doi: 10.3390/math10081283.
[2]. T. Kyriazos and M. Poga, “Dealing with Multicollinearity in Factor Analysis: The Problem, Detections, and Solutions,” Open Journal of Statistics, vol. 13, no. 03, pp. 404–424, 2023, doi: 10.4236/ojs.2023.133020.
[3]. C. García García, R. Salmerón Gómez, and J. García Pérez, “A review of ridge parameter selection: minimization of the mean squared error vs. mitigation of multicollinearity,” Communications in Statistics - Simulation and Computation, vol. 53, no. 8, pp. 3686–3698, Aug. 2022, doi: 10.1080/03610918.2022.2110594.
[4]. N. Akhtar, M. F. Alharthi, and M. S. Khan, “Mitigating Multicollinearity in Regression: A Study on Improved Ridge Estimators,” Mathematics, vol. 12, no. 19, p. 3027, Sep. 2024, doi: 10.3390/math12193027.
[5]. H. I. Dertli, D. B. Hayes, and T. G. Zorn, “Effects of multicollinearity and data granularity on regression models of stream temperature,” Journal of Hydrology, vol. 639, p. 131572, Aug. 2024, doi: 10.1016/j.jhydrol.2024.131572.
[6]. K Enwere, E Nduka, U Ogoke, “Comparative Analysis of Ridge, Bridge and Lasso Regression Models in the Presence of Multicollinearity,” IPS Intelligentsia Multidisciplinary Journal, vol. 3, no. 1, pp. 1–8, Jan. 2023, doi: 10.54117/iimj.v 3i1.5.
[7]. J. Cheng, J. Sun, K. Yao, M. Xu, and Y. Cao, “A variable selection method based on mutual information and variance inflation factor,” Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 268, p. 120652, Mar. 2022, doi: 10.1016/j.saa.2021.120652.
[8]. M. M. Abdelwahab, M. R. Abonazel, A. T. Hammad, and A. M. El-Masry, “Modified Two-Parameter Liu Estimator for Addressing Multicollinearity in the Poisson Regression Model,” Axioms, vol. 13, no. 1, p. 46, Jan. 2024, doi: 10.3390/axioms13010046.
[9]. V. K. S., B. B. S., J. Vellingiri, and D. R. Ramani, “Enhancing Mobile Edge Computing customer reviews analysis with Ensemble Sparse Support Vector L1 Regularization Based Crossover Discrete Mycorrhized Algorithm,” Biomedical Signal Processing and Control, vol. 109, p. 108019, Nov. 2025, doi: 10.1016/j.bspc.2025.108019.
[10]. Z. Hasan et al., “NEV-NCD: Negative Learning, Entropy, and Variance Regularization Based Novel Action Categories Discovery,” 2023 IEEE International Conference on Image Processing (ICIP), pp. 2720–2724, Oct. 2023, doi: 10.1109/icip49359.2023.10222474.
[11]. H. Wang and Y. Han, “An S-transform regularization-based denoising method for acoustic signals in photoacoustics,” Signal, Image and Video Processing, vol. 20, no. 1, Jan. 2026, doi: 10.1007/s11760-025-05043-1.
[12]. Q. U. Ain, A. Javed, K. M. Malik, and A. Irtaza, “Regularized forensic efficient net: a game theory based generalized approach for video deepfakes detection,” Multimedia Tools and Applications, vol. 84, no. 25, pp. 29621–29664, Oct. 2024, doi: 10.1007/s11042-024-20268-x.
[13]. J. Zhao, Y. Liu, L. Cheng, M. Zhang, M. Jiang, and J. Miao, “Sparse regularization-based feature selection: A survey,” Procedia Computer Science, vol. 221, pp. 1386–1393, 2023, doi: 10.1016/j.procs.2023.08.129.
[14]. T. Chen, Y. Zeng, H. Yuan, G. Zhong, L. L. Lai, and Y. Y. Tang, “Multi-level regularization-based unsupervised multi-view feature selection with adaptive graph learning,” International Journal of Machine Learning and Cybernetics, vol. 14, no. 5, pp. 1695–1709, Nov. 2022, doi: 10.1007/s13042-022-01721-5.
[15]. M. Khan, A. Ahmad, and M. Sarfraz, “L1 Regularization based Fetal Health Analysis Using ML Techniques,” 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–6, Jun. 2024, doi: 10.1109/icccnt61001.2024.10724506.

ISSN: 3080-7484