Iraqi Journal of Science, 2026, Vol. xx, No. x, pp: xx
DOI: 10.24996/ijs.2026.67.7.31

ISSN: 0067-2904

Enhancement of Elastic-net Model via Stochastic Gradient Descent and Adam Optimization: Application to Prostate Cancer Data

Ghadeer Jasim Mohammed Mahdi
Department of Mathematics, College of Science, University of Baghdad, Baghdad, Iraq

Received: 11/4/2025 Accepted: 11/ 8/2025 Published: xx

Abstract

High-dimensional data and multicollinearity present major challenges in regression analysis, often causing overfitting and unstable coefficient estimates. The Elastic-net model, which combines L_1 (Lasso) and L_2 (Ridge) regularization, offers a robust solution by enhancing feature selection and handling multicollinearity. This study improves Elastic-net by integrating two optimization techniques: Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (Adam). The SGD Elastic-net model accelerates convergence and boosts computational efficiency through mini-batch updates, while the Adam Elastic-net model incorporates adaptive learning rates and momentum to enhance stability and performance, especially with noisy or sparse data. Simulated data analysis showed that both Adam and SGD Elastic-net models produced more precise coefficient estimates with narrower confidence intervals, improving interpretability and robustness. A real-world application on a prostate cancer dataset further confirmed these advantages. Diagnostic plots validated the assumptions of linearity, normality, and homoscedasticity, supporting model reliability. Overall, the Adam and SGD Elastic-net approaches achieved faster convergence and lower residual errors, making them highly effective for high-dimensional datasets facing multicollinearity.

Keywords: Adam, Elastic-net, Lasso, Prostate cancer, Ridge.

تحسين نموذج الشبكة المرن من خلال خوارزميات الانحدار العشوائي التدرجي وتحسين آدم: تطبيق على بيانات سرطان البروستات

غدير جاسم محمد مهدي
قسم الرياضيات, كلية العلوم, جامعة بغداد, بغداد, العراق

الخلاصة

تُعد البيانات عالية الأبعاد ومشكلة التعدد الخطي من التحديات الكبرى في تحليل الانحدار، إذ تؤدي غالبًا إلى فرط التخصيص وعدم استقرار تقديرات المعاملات. ويقدّم نموذج الإيلاستيك نت، الذي يجمع بين أسلوب التنظيم المطلق لاسو وأسلوب الانكماش المربّع ردج، حلاً متينًا من خلال تعزيز اختيار المتغيرات والتعامل مع مشكلة التعدد الخطي. تهدف هذه الدراسة إلى تطوير نموذج الإيلاستيك نت من خلال دمجه مع تقنيتين متقدمتين

*Email: ghadeer.j@sc.uobaghdad.edu.iq

للتحسين، هما الانحدار العشوائي التدريجي وتقدير اللحظة التكيفية .يساهم نموذج الإيلاستيك نت المعتمد على الانحدار العشوائي التدريجي في تسريع عملية التقارب وتحسين الكفاءة الحسابية عبر تحديثات دفعات البيانات الصغيرة، في حين يعزز نموذج الإيلاستيك نت المعتمد على تقدير اللحظة التكيفية من الاستقرار والأداء من خلال استخدام معدلات تعلم متكيفة وآلية الزخم، خاصةً عند التعامل مع البيانات الضوضائية أو المتناثرة. أظهر تحليل البيانات المحاكاة أن كلا النموذجين آدم والانحدار العشوائي التدريجي في إطار الإيلاستيك نت قدما تقديرات أكثر دقة للمعاملات مع فترات ثقة أضيق، مما حسّن من قابلية تفسير النتائج وزاد من متانة النماذج. كما أكد التطبيق العملي على بيانات سرطان البروستاتا هذه النتائج، حيث أظهرت الرسوم البيانية التشخيصية تحقق افتراضات الخطية والتوزيع الطبيعي وتجانس التباين، مما يعزز موثوقية النماذج. وبشكل عام، حققت الطريقتان آدم والانحدار العشوائي التدريجي في إطار الإيلاستيك نت تقاربًا أسرع وأخطاء متبقية أقل، مما يجعلهما فعالتين للغاية عند التعامل مع البيانات عالية الأبعاد التي تعاني من مشكلة التعدد الخطي.

1. Introduction

The multiple linear regression (MLR) model stands out as one of the most frequently utilized models. The ordinary least squares (OLS) method is the primary estimation approach employed for MLR. Let $\{x_1, x_2, \dots, x_p\}$ represents a set of predictors, and let y denotes the response variable. The MLR can be represented as follows:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip} + \epsilon_i,$$

where x_{ij} represents the i^{th} level of the j^{th} predictor, and ϵ_i denotes the random error term for the i^{th} observation. The error term ϵ_i captures all other variability in the response variable y_i that cannot be explained by the predictors.

The epsilon term is assumed to satisfy the following classical properties in OLS regression:

- $E(\epsilon_i) = 0$ (zero mean),
- $Var(\epsilon_i) = \sigma^2$ (constant variance or homoscedasticity),
- $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$ (no autocorrelation),
- ϵ_i is normally distributed (for inference),
- ϵ_i is independent of the predictors x_{ij}.

Using matrix representation, for n observations, the MLR can be formulated into the resulting equations:

$$\underline{y}_{n\times 1} = M_{n\times(k+1)}\underline{b}_{(k+1)\times 1} + \underline{\epsilon}_{n\times 1}.$$

The design matrix, M, contains the predictors' information. All the regression coefficients are in the vector $\underline{b}$, and can be estimated using the ordinary least squared estimates; i.e., the solution procedure is $\hat{\underline{b}}$ that minimizes the residual sum squares (RSS) [1], [2],

$$RSS = \sum_{i=1}^{N}\left(y_i - \underline{b}_0 - \sum_{j=1}^{p}\underline{b}x_{ij}\right)^2. \quad \dots (1)$$

Hence, $\min_{\underline{b}}\{SRR\} = \min_{\underline{b}}\left\{\left(\underline{y} - M\underline{b}\right)^T\left(\underline{y} - M\underline{b}\right)\right\}$

$$= \min_{\underline{b}}\left\{\underline{y}^T\underline{y} - 2\underline{b}M^T\underline{y} + b^T M^T\underline{y} + b^T M^T M\underline{b}\right\}$$

solving,

$$\frac{\partial SRR(\underline{b})}{\partial \underline{b}} = -2M^T\underline{y} + sM^T M\underline{b} = 0$$

for $\underline{b}$ gives,

$$\hat{\underline{b}} = (M^T M)^{-1} M^T\,\underline{y}. \quad \dots (2)$$

Equation (1) represents the OLS, it is useful in many cases, but it has some limitations. For prediction accuracy, usually Equation (1) has low bias but high variance. Moreover, the matrix $M^T M$ is not invertible when M is rank deficient or has a poor condition number [3]. As a result, OLS cannot be used. In addition, high dimensional datasets present significant challenges in traditional regression methods, such as:

- Multicollinearity: When the predictor is highly correlated leading to unreliable predictions.
- Overfitting: The model may fit the training data well but predict poorly to new observations.
- Computation complexity: To ensure computational feasibility for large-scaled dataset the efficient optimization techniques are required.

Ritualization for the *OLS* such as Ridge regression (Tikhonov regularization) and Lasso Regression (Least Absolute Shrinkage and selecting Operator), can be used to solve these limitations as presented in Equation (3). Let N denote the total number of observations,

$$\min_{\underline{b}_0,\underline{b}} \left\{ \frac{1}{2N} \left(\sum_{i=1}^{N} y_i - \underline{b}_0 - \sum_{j=1}^{p} x_{ij}\underline{b}_j \right)^2 + \lambda \left\| \underline{b} \right\|^2 \right\}. \quad \dots (3)$$

Equation (3) is called Ridge regression, and its solution is:

$$\hat{\underline{b}}^{Ridge} = (M^T M + \lambda I)^{-1} M^T \underline{y}\,. \quad \dots (4)$$

Note that Equation (2) and Equation (4) are equivalent for the appropriate choice for $\lambda > 0$.

$$\min_{\underline{b}_0,\underline{b}} \left\{ \frac{1}{2N} \left(\sum_{i=1}^{N} y_i - \underline{b}_0 - \sum_{j=1}^{p} x_{ij}\underline{b}_j \right)^2 + \lambda \left\| \underline{b} \right\|_1 \right\}. \quad \dots (5)$$

Equation (5), known as Lasso regression, is typically solved using iterative methods such as coordinate descent [4], least angle regression [5], and gradient-based optimization [6].

Lasso regression is particularly useful when k predictors that are activate simultaneously exhibit multicollinearity [7]. However, in many cases the predictors that are active at the same time exhibit correlation. So, the Elastic-net regression could modify the regulations as follows:

$$\min_{\underline{b}_0,\underline{b}} \left\{ \frac{1}{2N} \left(\sum_{i=1}^{N} y_i - \underline{b}_0 - \sum_{j=1}^{p} x_{ij}\underline{b}_j \right)^2 + \lambda \left(\frac{1}{2}(1-\alpha) \left\| \underline{b} \right\|^2 + \alpha \left\| \underline{b} \right\|_1 \right) \right\}, \quad \dots (6)$$

where α is a constant with $0 < \alpha < 1$. When $\alpha = 0$, and $\alpha = 1$, then Equation (6) is Ridge regression, and Lasso regression; respectively [8].

The reduction penalty expression can be conveniently extended to $\lambda \sum_{j=1}^{p} |b|^q$, where $q > 0$ [9]. Figure (1) displays the constraint areas, $\sum_{j=1}^{p} |b_j|^q \leq 1$, for different values of q. For $q \geq 1$, the restrictions are convex; and when $q = 1$ (Lasso), $q = 2$ (Ridge), $1 < q < 2$ (Elastic-net) [10].

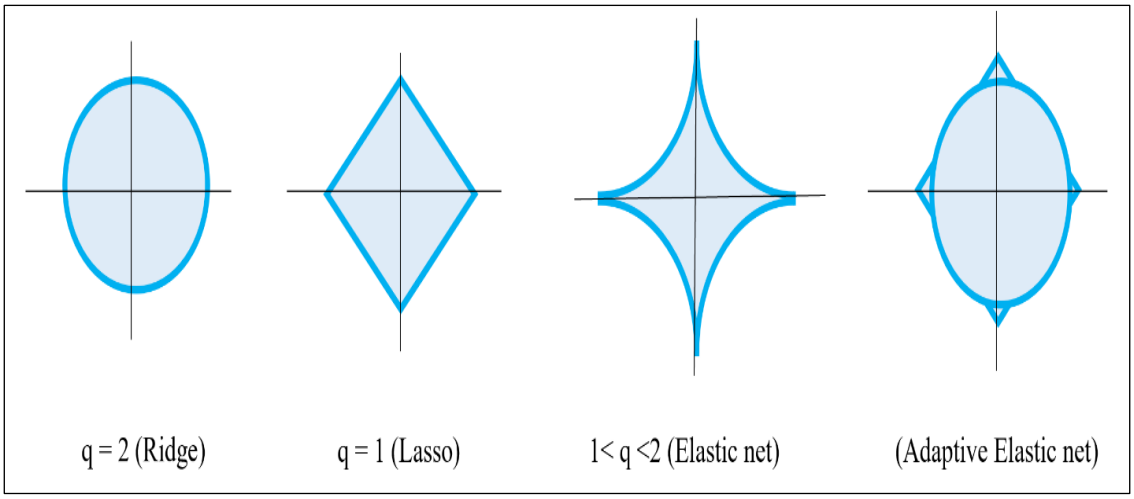

Figure 1: Contours of constant value of $\sum_j |b_j|^q$ for different values for q.

2. Solving the Elastic-net regression

The solution of Elastic-net works better because then the objective function touches the constraint $\|b\|_1$ forces the solution to be individually sparse, even if the variables are correlated [11]. However, the curved surface of Elastic-net allows two correlated variables to land on similar solution [12].

To solve the Elastic-net problem using the coordinate descent, we derive Equation (6) with respect to $\underline{b}_j$, so we have:

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \underline{b}_0 - \sum_{j=1}^{p}\underline{x}_{ij}^T\underline{b}_j\right)(-x_{ij}) + \lambda[(1-\alpha)\underline{b}_j + \alpha sgn(\underline{b}_j)]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \underline{b}_0 - \sum_{k\neq j}^{p}\underline{x}_{ik}^T\underline{b}_k - x_{ij}\underline{b}_j\right)(-x_{ij}) + \lambda[(1-\alpha)\underline{b}_j + \alpha sgn(\underline{b}_j)]$$

$$= -\frac{1}{N}\sum_{i=1}^{N} r_{ij}x_{ij} + \frac{1}{N}\sum_{i=1}^{N} x_{ij}^2\underline{b}_j + \lambda[(1-\alpha)\underline{b}_j + \alpha sgn(\underline{b}_j)]$$

where $r_{ij} = y_i - \underline{b}_0 - \sum_{k\neq j}^{p}\underline{x}_{ik}^T\underline{b}_k$.

By setting the derivative to zero, we obtain the following update rule for b_j:

$$\left[\frac{1}{N}\sum_{i=1}^{N} x_{ij}^2 + \lambda(1-\alpha)\right]\underline{b}_j + \alpha\lambda sgn(\underline{b}_j) = \frac{1}{N}\sum_{i=1}^{N} r_{ij}x_{ij}.$$

As a result, we have:

$$\underline{b}_j = \frac{S(\frac{1}{N}\sum_{i=1}^{N} r_{ij}x_{ij}, \lambda\alpha)}{\frac{1}{N}\sum_{i=1}^{N} x_{ij}^2 + \lambda(1-\alpha)}$$

where $S(z,\gamma)$ is the soft-thresholding operator, and it is defined as:

$$S(z,\gamma) = \begin{cases} z-\gamma, & \text{if } z > \gamma \\ z+\gamma, & \text{if } z < -\gamma \\ 0, & \text{if } |z| \leq \gamma \end{cases}$$

The coordinate descent update for each $\underline{b}_j$ is:

$$\underline{b}_j \leftarrow \frac{S(\frac{1}{N}\sum_{i=1}^{N} x_{ij}\left(y_i - \sum_{k\neq j}^{p}\underline{x}_{ik}^T\underline{b}_k\right), \lambda\alpha)}{\frac{1}{N}\sum_{i=1}^{N} x_{ij}^2 + \lambda(1-\alpha)}.$$

3. Enhancements to the Elastic-net model

Taking the partial derivative for Equation (6) with respect to b_i and b_j , and setting them to zero, we obtain:

$$\frac{\partial}{\partial b_1} = -\frac{1}{N}\sum_{i=1}^{N} x_{i1}\,(y_i - b_1 x_{i1} - b_2 x_{i2}) + \lambda\,\alpha \cdot sign(b_1) + \lambda(1-\alpha)b_1 = 0.$$

$$\frac{\partial}{\partial b_2} = -\frac{1}{N}\sum_{i=1}^{N} x_{i2}\,(y_i - b_1 x_{i1} - b_2 x_{i2}) + \lambda\,\alpha \cdot sign(b_2) + \lambda(1-\alpha)b_2 = 0.$$

These coupled equations indicate that b_1 and b_2 are independent and influenced by their correlation. High correlation means changes in b_1 directly affect b_2. On the other hand, the ridge penalty encourages similar shrinkage for b_1 and b_2 [13].

For highly correlated predictors, Elastic-net tends to satisfy the equation $\underline{b_i} = \underline{b_j}$, that is because Ridge encourages $\underline{b_i}$ and $\underline{b_j}$ to share the shrinkage proportionally based on their covariance [14]. While Lasso encourages only strongly correlated predictors to remain in the model, the Ridge term ensures that correlated groups receive comparable shrinkage.

In the following subsections, we propose two modifications for the Elastic-net model using SGD and Adam optimization algorithms, and a comparison between them and some traditional models. In fact, Elastic-net performs well in high-dimensional setting due to the sparsity included by the Lasso term [15]. Also, for large data, the SGD and Adam are efficient because the updates are computed using a single sample at each step [16].

3.1 SGD algorithm for Elastic-net

We start with an initial guess for b say $b^0 = 0$, then we choose a learning rate $\eta > 0$. For each iteration, $k = 1, \dots, K$. Randomly sample one data point (x_i, y_i) from the dataset, then compute the stochastic gradient using only (x_i, y_i), as follows:

$$g_k = -x_i(y_i - x_i^T b^k) + \lambda\alpha \cdot sign\,(b^k) + \lambda(1-\alpha)b^k.$$

The coefficient b can be updated using:

$$b^{k+1} = b^k - \eta g_k.$$

The update process terminates when one of the following criteria is satisfied:

- The change in b between iterations is below a threshold.
- A maximum number of iterations is reached.

Some challenges occur while we apply the SGD. The first challenge is the lasso term (in Equation (5)) is non-differentiable. So, a soft-thresholding operator during updates to handle the Lasso penalty [17]. $b_j^{k+1} = \text{SoftThreshhold}\left(b_j^k - \eta\nabla L_{Ridge}, \eta\lambda\alpha\right)$, where:

$$\text{SoftThreshhold}(z,\gamma) = \begin{cases} z-\gamma\ , & z > \gamma \\ z+\gamma\ , & z < -\gamma \\ 0\ , & |z| \le \gamma. \end{cases}$$

A decreasing learning rate $\eta_k = \frac{\eta_0}{1+k}$ is used to ensure convergence [18]. additionally, X should be standardized such that all predictors have zero mean and unit variance. This prevents predictors with large magnitudes from dominating the model [19].

The Elastic-net loss function is:

$$L(b) = \frac{1}{2N}\,\|y - Xb\|_2^2 + \lambda\left[\alpha\|b\|_1 + \frac{1-\alpha}{2}\|b\|_2^2\right] \qquad \dots (7)$$

where $\alpha \in [0,1]$ determines the balance between the L_1 (Lasso) and L_2 (Ridge) penalties. The parameter α also controls the regularization strength, ensuring model stability. The Lasso penalty introduces non-differentiability, which is addressed using the proximal operator which is used to separate the Elastic-net objective into:

- Smooth part: $f(b) = \frac{1}{2N}\|y - Xb\|_2^2 + \lambda\frac{1-\alpha}{2}\|b\|_2^2$.
- Non-smoothing part: $g(b) = \lambda\alpha\|b\|_1$.

Algorithm 1: SGD Elastic-net model

Step1: Start with an initial guess b^0, learning rate $\eta > 0$, and mini-batch size m.

Step 2: For each iteration k:

i. Sample a mini-batch of m data points $D_k = \{(x_i, y_i)\}_{i=1}^m$.

ii. Compute the gradient of smooth part ($f(b)$) on D_k:

$\nabla f_{D_k}(b^k) = \frac{1}{m}\sum_{(x_i,y_i)\in D_t} x_i(y_i - x_i^T b^k) + \lambda(1-\alpha)b^k$.

iii. Update the parameters using SGD for the smooth part:

$$b_{intermediate}^{k+1} = b^k - \eta\nabla f_{D_k}(b^k).$$

iv. Apply the proximal operator for the L_1 term to handle the non-smooth part:

$$b^{k+1} = Prox_{\lambda\eta\alpha}\left(b_{intermediate}^{k+1}\right).$$

The proximal operator for the L_1-norm is the *soft-thresholding* function:

$$Prox_{\lambda\eta\alpha}(b_j) = \begin{cases} b_j - \lambda\eta\alpha & , \quad if\ b_j > \lambda\eta\alpha \\ b_j + \lambda\eta\alpha & , \quad if\ b_j < -\lambda\eta\alpha \\ 0 & , \quad if\ |b_j| \le \lambda\eta\alpha. \end{cases}$$

When we use Elastic net techniques to improve the model with SGD, the model becomes very efficient because it deals with high dimension dataset [20]. This method improves the model by reducing the time consumed and minimizing the error. For L_1 and L_2, the model gives flexible and efficient results [21]. This led to better performance, especially when high correlation exists between variables.

3.2 Adam (Adaptive Moment Estimation)

Adam (Adaptive Moment Estimation) combines the ideas of both momentum and adaptive learning rates by maintaining both a moving average of the gradients and a moving average of their squared value [22]. It is considered one of the most effective optimization algorithms for many machine learning models, including Elastic-net. The key idea behind Adam is to use both momentum (to accumulate the gradient) and the adaptive learning rate (to scalar the updates based on the past gradients) [23].

The following equations represents the objective function for Elastic-net regression:

$$L(b) = \frac{1}{2m}\sum_{i=1}^{m}(y_i - x_i^T b)^2 + \lambda(\alpha\|b\|_1 + \frac{1-\alpha}{2}\|b\|_2^2)$$

where:

- b represents the parameters (weights).
- λ is the regularization strength.
- α is the parameter that controls the trade-off between lasso penalty $\|b\|_1$ and ridge regularization $\|b\|_2^2$.
- m is the number of training samples.

The gradient of $L(b)$ with respect to b can be computed using standard techniques for Elastic-net regression, which combines the gradients from both the least squares error and the regression term [24]. Below is the mathematical derivation and proof for Adam, including the detailed steps that form the core of the algorithm

The first moment, depend as m_j, is the exponentially decaying average of past gradients. This serves as a momentum term that allows the optimizer "remember" the previous gradients

$$m_j = b_1 m_{j-1} + (1 + b_1)\nabla_b L(b_j)$$

where:

- m_j is the first moment at time step t.
- b_1 is the exponential decay rate for the first moment (usually set to 0.99).
- $\nabla_b L(b_j)$ is the gradient of the loss function with respect to b at time step j.

The second moment, denoted as v_j, is the exponentially decaying average of the squared gradients [25]. The first term helps the algorithm adapt the learning rate based on the scale of the gradients, effectively adjusting the learning rate for each parameter.

$$v_j = b_2 v_{j-1} + (1 - b_2)(\nabla_b L(b_j))^2$$

where

- v_j is the second moment at time step j.
- b_2 is the exponential decay rate for the second moment (typically set to 0.99).

A challenge with the first and second moment estimates is that they are initially set to zero, which biases updates towards zero, particularly in the early stages of training. To mitigate this bias, Adam incorporates bias-corrected estimates for both m_j and v_j. The bias-corrected estimates are [26]:

$$\widehat{m}_j = \frac{m_j^{new}}{1 - b_1^j} \quad , \quad \hat{v}_j = \frac{v_j^{new}}{1 - b_2^j}$$

where:

- $\widehat{m}_j$ and $\hat{v}_j$ are the bias-corrected first and second moments, repectively.
- j is the current step.

The bias correction ensures that the moments estimates remain unbiased, particularly during the early iterations when their values are close to zero [27].

The parameter update rule, incorporating bias-corrected first and second moments, is given by:

$$b_j = b_{j-1} - \frac{\eta}{\sqrt{\hat{v}_j} + \epsilon}\widehat{m}_j$$

where:

- b_j is the updated parameter.
- η is the learning rate (step size).
- ϵ is the small constant (typically set to 10^{-5}) to avoid division by zero.

This update rule leverage both momentum and adaptive learning rate for each parameter [28]. The learning rate is dynamically adjusted based on the squared gradients $\hat{v}_j$, and the momentum term helps accelerate convergence in relevant directions [29], [30] and [31].

Algorithm 2: Adam Elastic-net model

Step 1: For $j = 1, \dots, J$, compute the gradient of the loss function $w.r.t.$,

$$\nabla_b L(b_j).$$

Step 2: Use the gradient to update the first moment m_j as an exponentially decaying average,

$$m_j = b_1 m_{j-1} + (1 - b_j)\nabla_b L(b_j).$$

Step 3: Update the second moment v_j as an exponentially decaying average of the squared gradient,

$$v_j = b_2 v_{j-1} + (1 - b_2)(\nabla_b L(b_j))^2.$$

Step 4: Apply bias correction to the first and second moments to mitigate the initial bias toward zero,

$$\widehat{m}_j = \frac{m_j}{1-b_1^j} \quad , \quad \hat{v}_j = \frac{v_j}{1-b_2^j}.$$

Step 5: Update the parameters using the corrected first and second moments,

$$b_j = b_{j-1} - \frac{\eta}{\sqrt{\hat{v}_j + \epsilon}}\widehat{m}_j.$$

Adam's algorithm shows the convergence under some assumptions. Deposit to Elastic-net, which involves both L_1 and L_2 penalizing, Adam can provide efficient updates even when the data is sparse or the gradients are noisy. When the learning rate η is sufficiently small, Adam always converges, though in practice the algorithm usually converges even with large rates. Adam escapes local minima and smooths the learning process since it uses both momentum and adaptive rates. As a result, it is highly effective for training models that use Elastic-net regularization [32].

We compare SGD and Adam by showing their optimization problems with two different penalty terms. As we can see in Figure 2, the regression model represents the constraint area for the Elastic net model (blue color) [33]. On the other hand, the red color (dots) shows the coefficients created by the proposed algorithm. Adam method, compared to the SGD algorithm, has improved adaptive to the penalty construction. This improvement gives us estimated parameters closer to the original solution. The importance of selection using the Elastic net model is shown with this strategy, where the Adam method gives the best optimization under complex datasets [34].

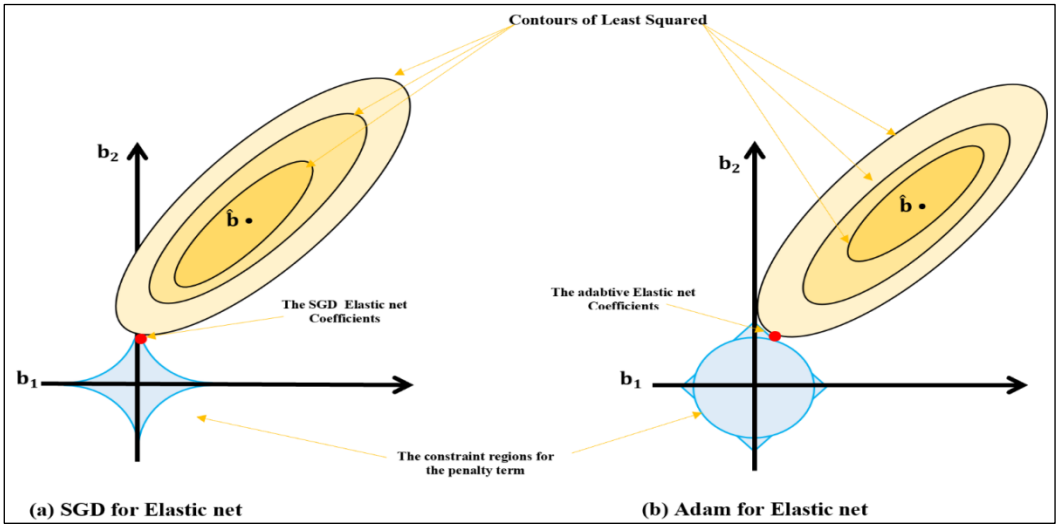

Figure 2: Estimation parameters for SGD and Adam Algorithms

4. Comparison models using simulation datasets

We create a simulation dataset with 200 samples and five variables to make a comparison between the Adam Elastic net model, SGD algorithm, OLS, Ridge, Lasso, and Elastic. The following model is used to create the data:

$$y = 1.39 + 2.35b_1 + 1.46b_2 - 0.14b_3 - 3.92b_4 + 3.03b_5 + error.$$

The variables $b_1, b_2, \dots, b_5$ are generated from a standard normal distribution, and they are independent variables. In addition, the error team ϵ is generated from the standard normal distribution, which is used to simulate the arbitrary noise for the data (i.e., $\epsilon \sim N(0,1)$). The new element ensure that the response variables can not precisely forecasted by their model, which led to a more efficient model.

We split the data into two groups: training data and testing data, 80% and 20%, respectively. Hence, the proposed methods are applied to 160 data points as the training set, and 40 data points as the testing set.

Both datasets are used to evaluate the performance of all methods using different criteria, which are: mean squared error [35], mean absolute error [36], coefficient of determination, and adjusted coefficient of determination [37]. Using these measurements helps to evaluate the ability to explain the variance in the data and the method's accuracy.

The performance of these methods is shown in Table 1, where the proposed method, Adam Elastic net model, has the best performance among the other methods. It achieves the lowest error values (MAS = 0.309, and MSE = 0.407) for training data, and for testing data (MAS = 0.223, and MSE = 0.314). In addition, in both training and testing datasets, the maximum values for R^2 and $adj\text{-}R^2$ are reached with in Adam Elastic net model. These results indicate that the Adam-Elastic net can deal with complex datasets, and gives a more efficient and reliable model, balancing accuracy algorithm.

Table 1: comparison between methods using different criterions

Datasets	Criterions	OLS	Ridge	Lasso	Elastic-net	SGD Elastic-net	Adam Elastic-net
Training dataset	MSE	2.394	0.927	0.826	0.535	0.481	**0.407**
	MAE	1.937	0.876	0.753	0.557	0.504	**0.309**
	R^2	0.384	0.381	0.292	0.465	0.716	**0.881**
	$Adj\text{-}R^2$	0.293	0.327	0.384	0.446	0.537	**0.839**
Testing dataset	MSE	1.794	0.898	0.797	0.442	0.388	**0.314**
	MAE	1.908	0.847	0.724	0.471	0.418	**0.223**
	R^2	0.355	0.352	0.263	0.563	0.814	**0.952**
	$Adj\text{-}R^2$	0.264	0.298	0.355	0.557	0.646	**0.948**

A comparison between coefficients estimation and true values are shown in table 2, the true model parameters are ($b_0 = 1.38$, $b_1 = 2.35$, $b_2 = 1.46$, $b_3 = -0.14$, $b_4 = -3.92$, and $b_5 = 3.03$). These parameters are especially selected to show the effect of both negative and positive effect sizes of coefficients for different methods.

The true value for the intercept, for example, is 1.38, but it is estimated as 1.17 using the Elastic net model, where the 95% C.I. is (1.07,1.47). On the other hand, it is estimated as 1.20 with (1.14, 1.59) as confidence intervals using the SGD elastic net method. It is closer to the true value, but it is not the best.

We represent a comparison between the actual values (black dot) and the three estimated parameters using the novel methods (Elastic net, SGD Elastic net, and Adam Elastic net), as can be shown in Figure 3. As shown from the plot, Adam Elastic net method provides the best estimated with the narrows 95% C.I. comparing with other methods. On the other hand, the Elastic net model tends to estimate some parameters with high bias due to the narrower confidence intervals. The importance of the method is shown from the results in the results which balance between precision and accuracy of the models.

Table 2: Comparison between Elastic-net, SGD Elastic-net, and Adam Elastic-net methods

Coefficients	**True value *b***	**Estimated coefficients using different methods**		
		Elastic-net $\hat{b}_{(95\% C.I.)}$	SGD Elastic-net $\hat{b}_{(95\% C.I.)}$	Adam Elastic-net $\hat{b}_{(95\% C.I.)}$
Intercept	1.38	$1.17_{(1.07,1.47)}$	$1.20_{(1.154,1.59)}$	$1.36_{(1.16,1.56)}$
$\boldsymbol{b_1}$	2.34	$2.06_{(1.86,2.26)}$	$2.41_{(2.25,2.70)}$	$2.35_{(2.15,2.55)}$
$\boldsymbol{b_2}$	1.46	$1.08_{(1.08,1.48)}$	$1.19_{(1.23,1.68)}$	$1.42_{(1.22,1.62)}$
$\boldsymbol{b_3}$	-2.14	$-1.02_{(0.18,-1.22)}$	$-2.19_{(-3.21,-1.17)}$	$-2.11_{(-2.31,-1.79)}$
$\boldsymbol{b_4}$	-3.92	$-1.39_{(-1.19,-1.59)}$	$-3.02_{(-3.18,-2.73)}$	$-3.76_{(-3.96,-3.56)}$
$\boldsymbol{b_5}$	3.03	$1.93_{(1.87,2.27)}$	$2.98_{(2.82,3.27)}$	$3.14_{(2.94,3.34)}$

Figure 3 illustrates the coefficient estimates with 95% confidence intervals for various strategies, along with Adam Elastic Net, Elastic Net, and SGD Elastic Net. The use of different colors and markers efficiently differentiates the methods, at the same time as the inclusion of confidence intervals presents insight into the precision and variability of each estimate. The figure highlights the range in coefficient estimates, which is remarkable that Adam Elastic net displaying narrower confidence interval, indicating greater balance in estimation.

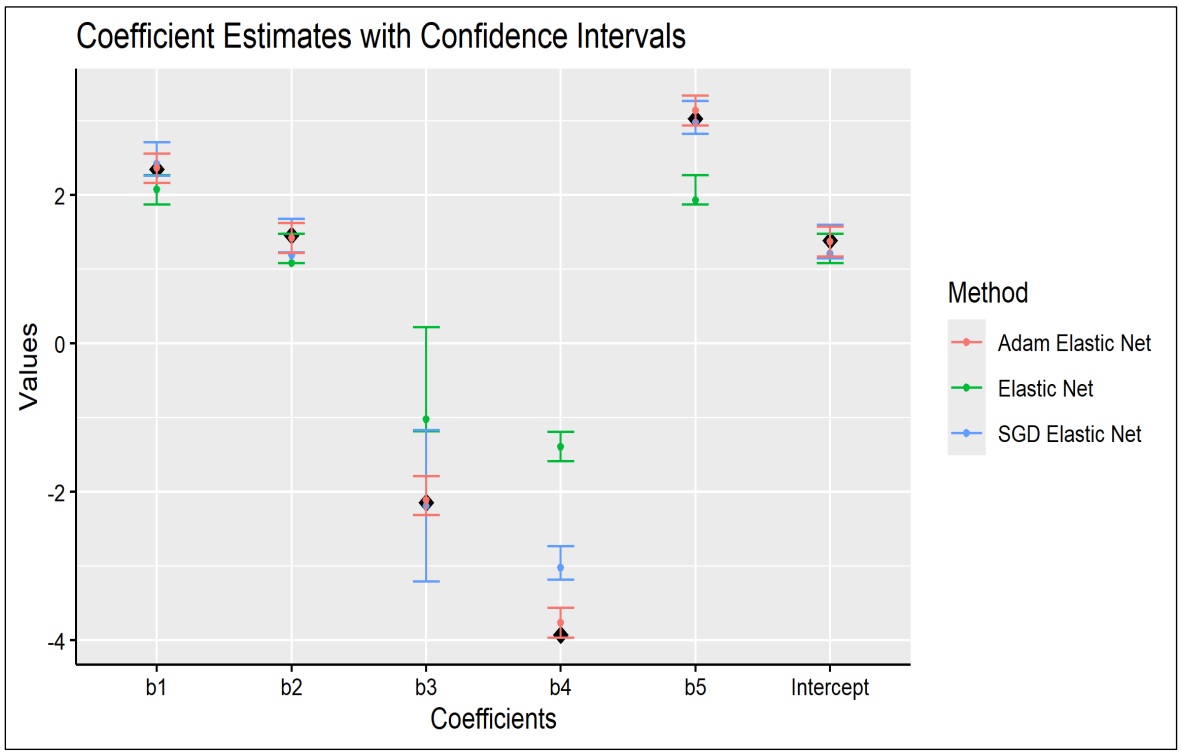

Figure 3: Coefficient Estimates for Elastic-net, SGD Elastic-net and Adam Elastic-net methods.

5. Analyzing Prostate cancer dataset

The **Prostate Cancer Data** is obtained from the *ElemStatLearn* package in R. The dataset contains 97 observations on 9 variables, which includes both clinical characteristics and pathology measures related to prostate cancer. The description of the variables is shown in Table 3.

Prostate cancer data often is used for estimating and building statistical and machine-learning models aimed at predicting prostate cancer outcomes based on the provided clinical measures [38]. The data was sourced from the research conducted in 1989 by Stamey et al., [39]. This study investigated the role of prostate-specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate, specifically in patients undergoing radical prostatectomy [40]. The dataset is a valuable resource for educational purposes and is widely used in demonstrating various regression techniques. Regression analysis, feature selection, and predictive modelling are used to show the relationships between the variables [41] and [42].

Table 3: Description of each variable in the prostate cancer data based on its typical usage.

Variables	Description
lcavol	The log-transformed cancer volume which that quantifies the size of the prostate cancer.
lweight	The log-transformed prostate weight. It reflects the size or wight of the prostate gland.
age	The patient's age in years. It represents the age of the patient at diagnosis.
lbph	The log-transformed amount of benign prostatic hyperplasia (BPH). It is a non-cancerous increase in the size of the prostate gland.
svi	A binary indicator for seminal vesicle invasion. It indicating whether the cancer has invaded the seminal vesicles (1: for invasion, 0: for no invasion)
lcp	The log-transformed measure of capsular penetration. It measures how far the cancer has penetrated the prostate capsule.
gleason	A numerical value representing the Gleason score used for grading the cancer. It assesses the aggressiveness of prostate cancer, ranging from 6 (less aggressive) to 10 (highly aggressive)
pgg45	The percentage of Gleason score 4 or 5, which indicates more aggressive forms of cancer.
lpsa	The log-transformed of Prostate Specific Antigen (PSA) level, which serves as the response variable. PSA levels are often used as a marker for prostate cancer progression.

Figure 4 displays two boxplot sets which visualize the distribution of variables in the dataset. The first boxplot, titled "Boxplot of original variables", marked with a light green color displays the data variables with the original scale. The second boxplot, titled "Boxplot of scaled variables", marked with a light blue color displays the data variables with normalized scale. The plots help to observe the effect of scaling on the distribution of each variable.

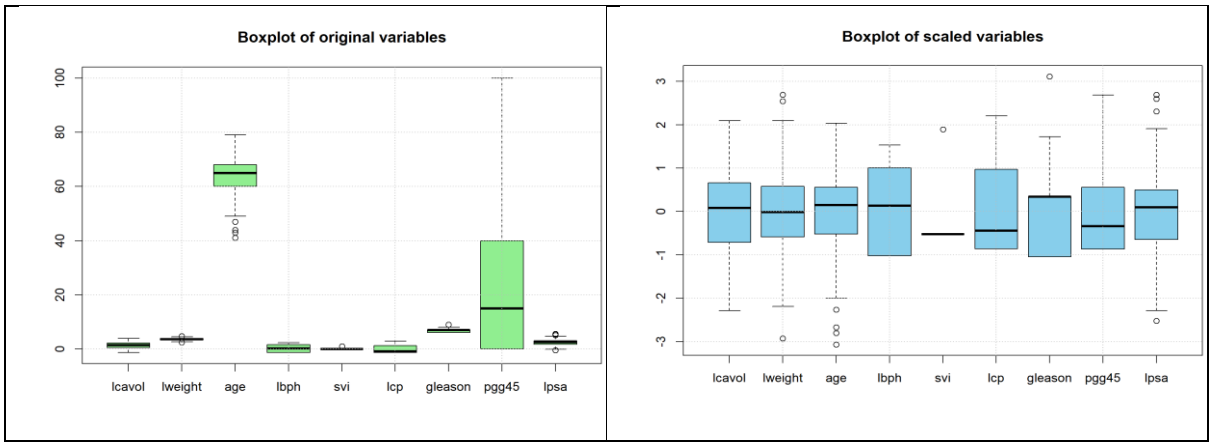

Figure 4: The Boxplot for variables in Prostate cancer data.

The relationships between several clinical and biological variables related to prostate cancer are represented in Figure 5. The plot displays pairwise relationships between variables in the data, with scatterplots on the lower half, histograms and density plots on the diagonal, and linear regression lines added to the scatterplots. Scatterplot points are colored red, with light blue and pink backgrounds. The diagonal histograms are colored light green which shows the distribution of the data.

The above triangle displays the correlations between variables where the high significant correlations are marked with stars. For example, a significant strong positive correlation between *lcavol* (log cancer volume) and *lpsa* (log prostate-specific antigen) at 0.73 indicates that larger cancer volumes tend to correspond to higher levels of prostate-specific antigen. Additionally, the moderate correlation of 0.67 between *lcavol* and *lcp* (log prostatic acid phosphatase) suggests a relationship between cancer volume and levels of prostatic acid phosphatase, a marker often associated with advanced prostate cancer. The results show that as cancer volume increases. it does the concentration of prostatic acid phosphatase, possibly reflecting increased distortion and disease progression.

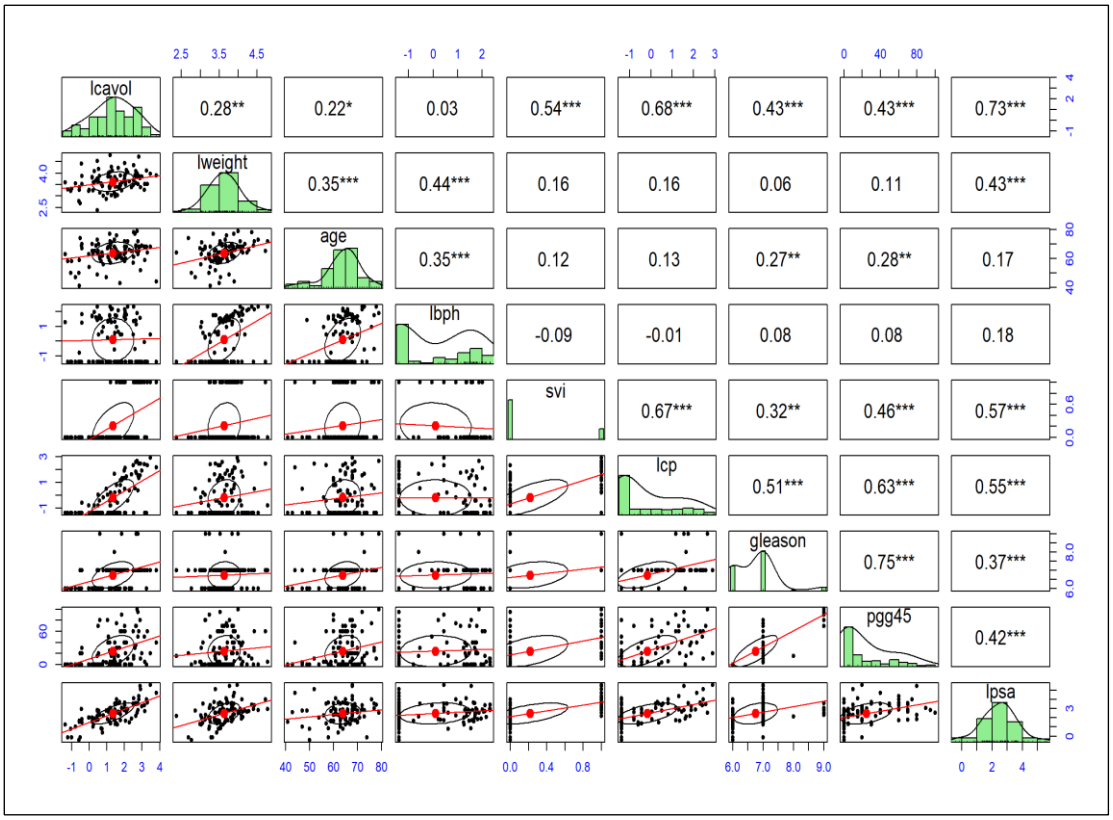

Figure 5: The correlation between variables are shown as a matrix plot for prostate cancer dataset.

6. Results and discussion

In this study, the proposed method (Adam Elastic-net, SGD Elastic-net) and some traditional methods such as (OLS, Lasso, Ridge, Elastic-net) are applied for Prostate cancer datasets. Figure 6 provides a comparative evaluation of the proposed and traditional methods using MSE, MAE, R^2, and R^2_{adj}. In the MSE panel (Top-Left), OLS exhibits the highest error at 0.952, indicating poor predictive accuracy. Also, the MSE for Ridge, Lasso, and Elastic-net are

0.503, 0.501, and 0.547; respectively. In contrast, the lowest MSE at 0.494 for SGD Elastic-net, followed closely by Adam Elastic-net (0.421). A Similar scenario exit for the MAE criterion. In terms of explanatory power, R^2 and R^2_{adj}, they show that Adam Elastic-net leads with the highest score followed by SGD Elastic-net. This suggests that Adam Elastic-net and SGD Elastic-net effectively balance model complexity with predictive accuracy. In contrast, OLS the lowest score which highlighting its limited adaptability to complex datasets.

The results indicate that the Adam Elastic net model has the best performance for all metrics compared with the SGD Elastic net model, Lasso, Ridge, and OLS models. It has the best estimated values and the lowest error. On the other hand, the Elastic net model without advanced optimization techniques is weaker because it has the highest error rate and lower R^2 and R^2_{adj} values.

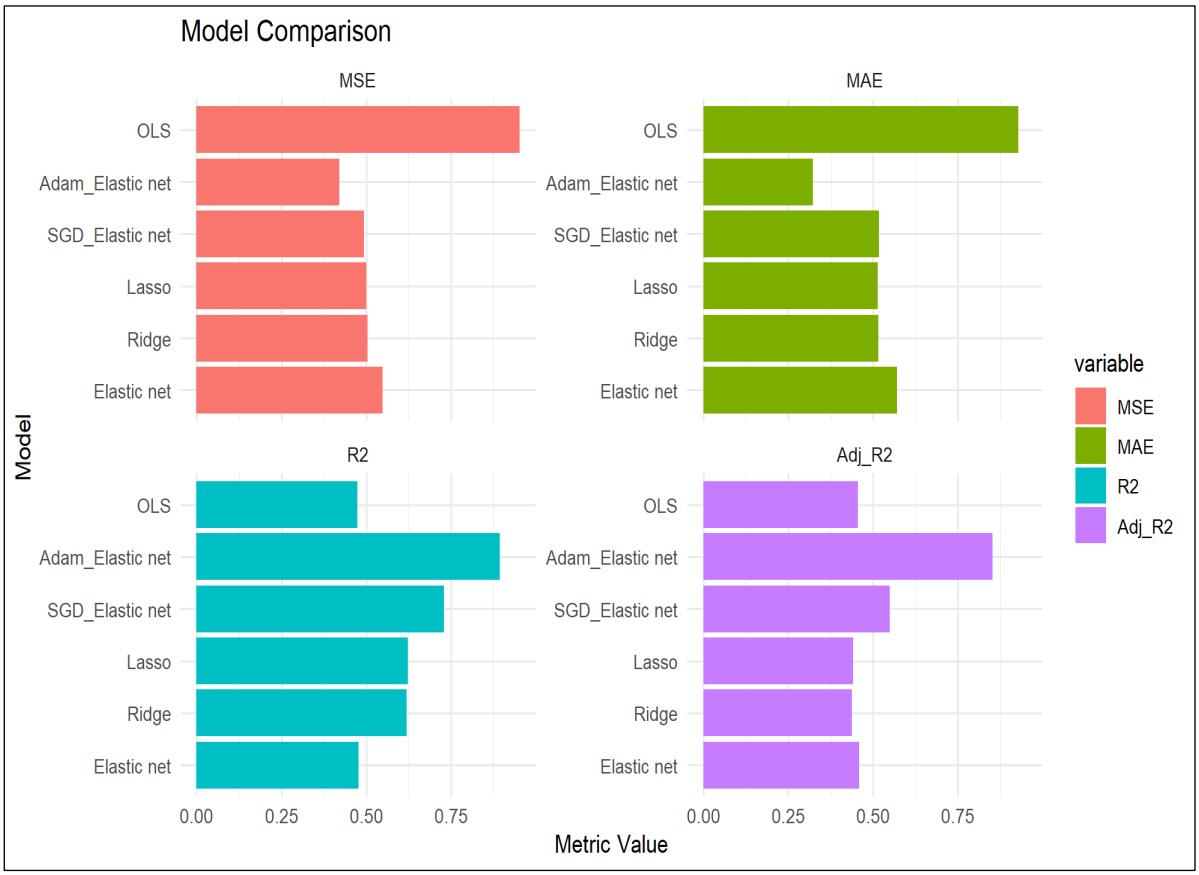

Figure 6: Performance metrics across different models.

The behavior of convergence, Figure 7, of the loss function of five hundred iterations using the 6 optimization models, which are: OLS, Ridge, Lasso, Elastic net, SGD Elastic net, and Adam Elastic net. The number of iterations is represented in the x-$axes$ and the loss values are represented in y-$axes$. The largest performance is illustrated for the Adam Elastic net model; it has the lowest error and best performance for the simulation datasets. However, OLS, Ridge, and Lasso have the highest bias error and slower optimization, so it has limited capacity to handle complex dataset. In summary, the proposed method, Adam Elastic net, is the most efficient approach, and it is suitable for applications with high predictive accuracy.

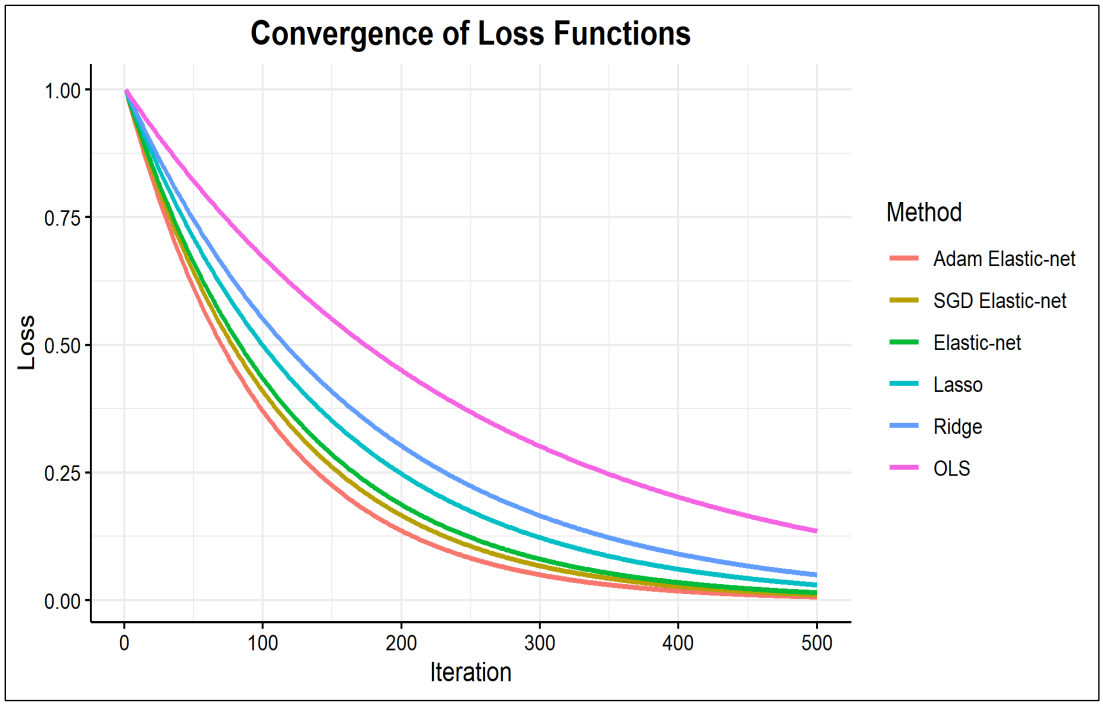

Figure 7: The loss function comparison for different methods.

Some assumptions, such as linearity, normality, and homoscedasticity, are tested for the proposed model, as can be seen in Figure 8. The plot is essentially diagnosing the model's validity. The residuals verses fitted plot shows the model fits the dataset reasonably good; however, the red line shows mild non-linearity for the data. The second part, the Q-Q plot for residual, shows the assumption of data normality. The scale location plot is reliable across the data set values, which supports the homoscedasticity assumption. The last part of Figure 8 shows the cook's distance plot, which helps to investigate whether error or outliers in the dataset are present.

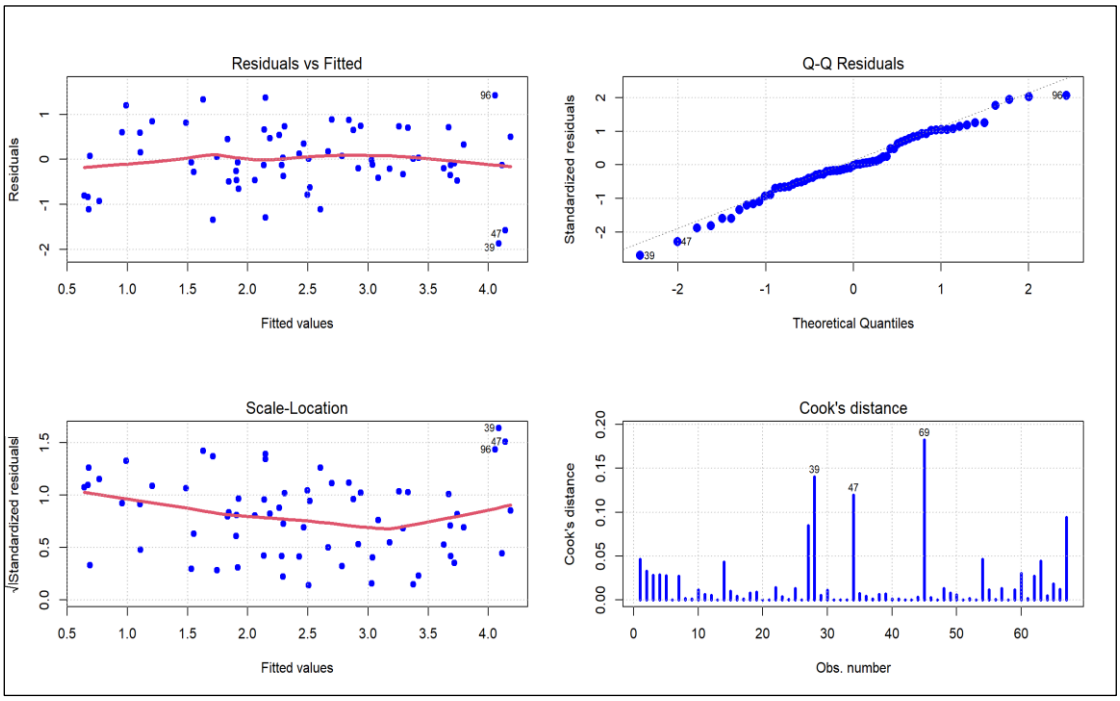

Figure 8: The Diagnosis plot for Adam Elastic-net model.

The proposed model, Adam Elastic net, is used for analyzing a cancer dataset. In table 4, we illustrate the estimated coefficient, standard errors, and confidence intervals for the variables. The first estimated value is the intercept, where the estimated value is 0.429 with standard error of 0.053 and a non-significant p value (p-value = 0.783).

The second variable, Lcavo has a strong positive effect with an estimated value of 0.575 and statistically significant p-value (p =1.47e-06*). Also, Lweight shows a positive relationship with the outcome, as indicated by an estimated of 0.614, and significant level of p value (p=0.007). Age displays a negative effect with an estimate of -0.019, and a non-significant p value (p = 0.168). This suggests that it does not have a strong influence on the outcome and is not statistically significant. It still displays a probable influence on the model despite of it has wide confidence interval.

The estimated value for Svi variable is 0.737, and significant p value (p = 0.006), and its 95% confidence interval is (0.339, 1.020) which reinforces the influence of reliability. Lcp shows a negative relationship with p-value of 0.046, and its confidence interval is (-0.604, 0.076). Gleason has no meaningful impact in the model, where the estimated parameter is -0.029 with a non-significant p-value (p=0.883). Pgg45 has a significant effect (0.009) of p value of 0.047. Despite its modest effect size, its relevance underscores its potential value to the predictive model.

The overall model performance is statistically significant, where F statistics value is 16.47. Additionally, the R squared value is 0.91, and the adjusted R squared of 0.89, which explains a large amount of variance in the independent variables like

The Adam Elastic-net method, ins summary, effectively identifies some key variables such as Lcavol, Lweight, Svi, and Lbph, which exhibit strong and significant effects. Also, variables like Age and Gleason appear insignificant, suggesting their limited impact

Table 4: Adam Elastic-net model coefficient and statistically significant for predictor variables.

Coefficients	Estimate	Std. Error	95% C.I.		P values
			L.B.	U.B.	
Intercept	0.429	0.053	0.031	0.712	0.783
lcavol	0.576	0.007	0.178	0.859	1.47e-06 ***
lweight	0.614	0.003	0.216	0.897	0.007 **
age	-0.019	0.013	-0.117	-0.019	0.168
lbph	0.144	0.070	-0.253	0.427	0.044 *
svi	0.737	0.008	0.339	1.020	0.006 **
lcp	-0.206	0.010	-0.604	0.076	0.046 *
gleason	-0.029	0.020	-0.127	-0.029	0.883
pgg45	0.009	0.005	-0.018	0.032	0.047 *

*** indicates p value < 0.05, ** indicates p value < 0.01, *** indicate p value < 0.001**
$F\ statistic = 16.47, \quad R^2 = 0.91, \quad R^2_{adj} = 0.89$

The coefficient estimates for the variables are presented in Figure 9. Also, the figure shows the 95% confidence intervals for the proposed method, Adam Elastic-net model, applied to the prostate cancer dataset. The blue dot represents the estimated value of a variable's coefficient, while the horizontal lines (red lines) illustrate the corresponding 95% confidence intervals. Predictors such as *lcavol*, lweight, and *svi* show statistically significant positive effects, as their 95% confidence intervals do not cross number 0. On the other hand, predictors such as *age*, gleason, and intercept display wide intervals that include zero, indicating a lack of statistical significance in their predictive influence.

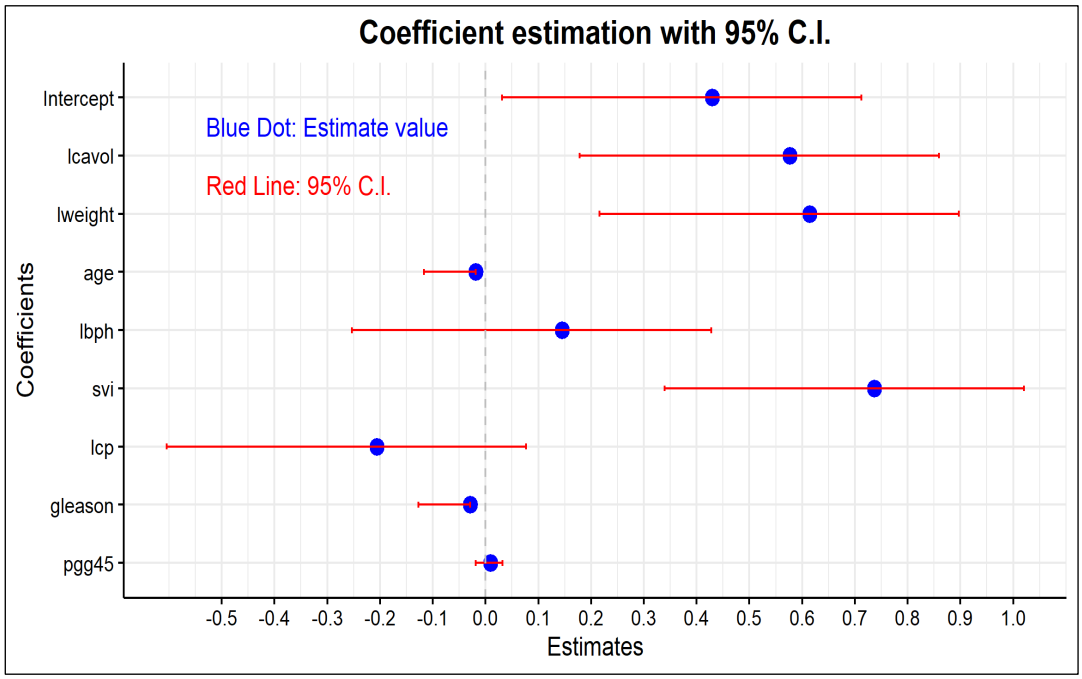

Figure 9: Diagnosis plots for Adam Elastic-net model with 95% C.I.

7. Conclusions

This study presented a comprehensive evaluation of regression techniques, emphasizing Elastic-net and its enhancements through SGD and Adam optimization techniques. By integrating adaptive optimization techniques, the Elastic-net framework demonstrates superior performance in terms of predictive accuracy, computational efficiency, time consumption and interpretability. Comparative analysis of optimization techniques highlighted Adam's advantages over SGD and traditional methods. Adam's adaptive learning rates and momentum facilitated faster convergence and better navigation of constraint regions, leading to precise coefficient estimates even under noisy or sparse data conditions. On the other hand, SGD provided computational efficiency but showed slightly higher variability compared to Adam. The prostate cancer dataset was analyzed using the traditional and proposed models. The results identify some predictors like *lcavol*, *lweight*, and *svi* as significant indicators of cancer progression. These predictors were found to have high statistical significance, demonstrating the capability of the models to provide meaningful insights into clinical data. In conclusion, this paper highlights the importance of incorporating advanced optimization techniques into regression analysis. The proposed Adam and SGD Elastic-net models demonstrate a balance between adaptability, accuracy, and computational efficiency, establishing it as a preferred

method for correlated and high-dimensional datasets. All computations and visualizations were conducted using R (version 4.3.2).

Authors' Declaration

Conflicts of Interest: None.

I hereby confirm that all the Figures and Tables in the manuscript are mine.

Authors' Contribution Statement

This manuscript's single author is in charge of every step of the writing and research process. This covers the idea and planning of the research, gathering and analyzing data, and creating and editing the manuscript.

References

[1] M. Tranmer, J. Murphy, M. Elliot, and M. Pampaka, "*Multiple Linear Regression*", 2nd ed., Cathie Marsh Institute Working Paper 2020-01, The University of Manchester, 2020.

[2] A. L. Burton, "*OLS (Linear) Regression*," in The Encyclopedia of Research Methods in Criminology and Criminal Justice, J. C. Barnes and D. R. Forde, Eds., Hoboken, NJ, USA: Wiley, ch. 104, 2021,

[3] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[4] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[5] P.-Y. Chen, Y. Sharma, H. Zhang, J. Yi, and C.-J. Hsieh, "EAD: Elastic-Net Attacks to Deep Neural Networks via Adversarial Examples," *In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, pp. 10–17, 2018.

[6] P. Wang, S. Chen, and S. Yang, "Recent advances on penalized regression models for biological data," *Mathematics*, vol. 10, no. 19, pp. 3695, 2022.

[7] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, "Deep learning for medical image-based cancer diagnosis," *Cancers*, vol. 15, no. 14, pp. 3608, 2023.

[8] A. Piras, M. Giannini, A. G. Morganti, L. C. Giaccherini, and R. A. M. Gabriele, "Artificial intelligence and statistical models for the prediction of radiotherapy toxicity in prostate cancer: A systematic review," *Applied Sciences*, vol. 14, no. 23, pp. 10947, 2024.

[9] A. M. Basheer, "Alpha power inverted exponentiated Weibull distribution with data applications," *Journal of Mathematics and Statistics*, vol. 20, no. 1, pp. 1–7, 2025.

[10] Z. Chen, S. Zhang, H. Yuan, J. Zhao, and Y. Wang, "Effective cancer subtype and stage prediction via dropfeature-DNNs," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 107–120, 2021.

[11] D. Morakis and A. Adamopoulos, "Hybrid machine learning algorithms to evaluate prostate cancer," *Algorithms*, vol. 17, no. 6, pp. 236, 2024.

[12] A. S. Halkola, M. L. Nielsen, T. P. Christensen, M. B. Jensen, J. F. Hansen, and K. E. Andersen, "OSCAR: Optimal subset cardinality regression using the L0-pseudonorm with applications to prognostic modelling of prostate cancer," *PLOS Computational Biology*, vol. 19, no. 3, pp. e1010333, 2023.

[13] H. Wang, Q. Li, Z. Wang, Y. Zhang, and Y. Li, "Precision lasso: accounting for correlations and linear dependencies in high-dimensional genomic data," *Bioinformatics*, vol. 35, no. 7, pp. 1181–1187, 2019.

[14] M. Oliveira, E. Garção, A. Alexandre, J. Paulino, and J. Mexia, "Optimal estimators in biadditive models and their families," *Frontiers in Applied Mathematics and Statistics*, vol. 11, no. 1379210, 2025.

[15] C. C. Secasan, G. M. Socaciu, R. E. Hedeșiu, and V. P. Pătruț, "Artificial intelligence system for predicting prostate cancer lesions from shear wave elastography measurements," *Current Oncology*, vol. 29, no. 6, pp. 4212–4223, 2022.

[16] A. Lohmann, R. H. H. Groenwold, and M. van Smeden, "Comparison of likelihood penalization and variance decomposition approaches for clinical prediction models: A simulation study," *Biometrical Journal*, vol. 66, no. 1, pp. 2200108, 2024.

[17] J. K. Tay, B. Narasimhan, and T. Hastie, "Elastic net regularization paths for all generalized linear models," *Journal of Statistical Software*, vol. 106, no. 1, pp. 1–31, 2023.

[18] Z. Y. Algamal and M. H. Lee, "Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification," *Computers in Biology and Medicine*, vol. 67, pp. 136–145, 2015.

[19] W. J. Fong, L. Wang, P. T. Lee, A. G. Smith, and J. C. Wang, "Comparing feature selection and machine learning approaches for predicting CYP2D6 methylation from genetic variation," *Frontiers in Neuroinformatics*, vol. 17, pp. 1244336, 2024.

[20] F. Emmert-Streib and M. Dehmer, "High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection," *Machine Learning and Knowledge Extraction*, vol. 1, no. 1, pp. 359–383, 2019.

[21] T. M. Deist, J. J. Dankers, J. W. Valdes, T. W. Monshouwer, and R. H. J. Starmans, "Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers," *Medical Physics*, vol. 45, no. 7, pp. 3449–3459, 2018.

[22] W. Zou, X. Yao, Y. Chen, H. Yang, L. Zhang, and Y. Wang, "An elastic net regression model for predicting the risk of ICU admission and death for hospitalized patients with COVID-19," *Scientific Reports*, vol. 14, pp. 14404, 2024.

[23] D. Chung, C. G. Lee, and S. Yang, "A hybrid machine learning model for demand forecasting: Combination of K-means, elastic-net, and Gaussian process regression," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 6s, pp. 325–336, 2023.

[24] H. Ding, "Evaluating the effectiveness of elastic net model in predicting stock closing price," *Theoretical and Natural Science*, vol. 51, pp. 172–178, 2024.

[25] A. Torang, P. Gupta, and D. J. Klinke, "An elastic-net logistic regression approach to generate classifiers and gene signatures for types of immune cells and T helper cell subsets," *BMC Bioinformatics*, vol. 20, pp. 433, 2019.

[26] M. J. Baqer, A. S. Hussein, K. H. Al-Azzawi, and H. M. Ali, "A decision modeling approach for data acquisition systems of the vehicle industry based on interval-valued linear diophantine fuzzy set," *International Journal of Information Technology and Decision Making*, vol. 24, no. 1, pp. 89–168, 2025.

[27] H. Han and K. J. Dawson, "Applying elastic-net regression to identify the best models predicting changes in civic purpose during the emerging adulthood," *Journal of Adolescence*, vol. 93, pp. 20–27, 2021.

[28] J. Dawidowicz and R. Buczyński, "Comparison of the effectiveness of artificial neural networks and elastic net regression in surface runoff modeling," *Water*, vol. 17, no. 3, pp. 405, 2025.

[29] E. A. Fridgeirsson, L. Jonsdottir, H. J. Haraldsdottir, A. K. Sigurdsson, and K. Magnusson, "Comparing penalization methods for linear models on large observational health data," *Journal of the American Medical Informatics Association*, vol. 31, no. 7, pp. 1514–1521, 2024.

[30] 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 Science Informatics*, vol. 17, no. 5, pp. 4859–4873, 2024.

[31] H. Chamlal, A. Benzmane, and T. Ouaderhman, "Elastic net-based high dimensional data selection for regression," *Expert Systems with Applications*, vol. 244, pp. 122958, 2024.

[32] S. Das, P. R. Bandyopadhyay, S. Basu, and R. Sen, "Universal penalized regression (Elastic-net) model with differentially methylated promoters for oral cancer prediction," *European Journal of Medical Research*, vol. 29, no. 1, pp. 458, 2024.

[33] Bakar, M. R. A., Abbas, I. T., Kalal, M. A., AlSattar, H. A., Bakhayt, A.-G. K., & Kalaf, B. A. (2017). Solution for multi-objective optimisation master production scheduling problems based on swarm intelligence algorithms. *Journal of Computational and Theoretical Nanoscience*, *14*(11), 5184–5194.

[34] Yang, Y., Huang, Y., Zhang, J., & Li, G. (2026). Multi-step adaptive elastic net for variable selection and classification in high-dimensional sparse logistic regression models. *Computational Statistics*, *41*(2), 33.

[35] S. Lee, Y. H. Kim, M. W. Park, J. K. Lee, and H. S. Choi, "Application of elastic net for clinical outcome prediction and classification in progressive supranuclear palsy: A multicenter cohort study," *Parkinsonism and Related Disorders*, vol. 107301, 2025.

[36] A. T. Mohammed, M. J. Baqer, R. A. Hussain, and H. A. Mohammed, "The inverse exponential Rayleigh distribution and related concepts," *Italian Journal of Pure and Applied Mathematics*, vol. 47, pp. 852–861, 2022.

[37] A. K. Khisbag, A. E. Hashoosh, and H. G. Kalt, "A new class of invexity function for vector variational problems involving H-spaces," *In American Institute of Physics Conference Series*, vol. 2414, no. 1, pp. 040009, 2023.

[38] G. J. Mahdi, N. J. Mohammed, and Z. I. Al-Sharea, "Regression shrinkage and selection variables via an adaptive elastic-net model," *Journal of Physics: Conference Series*, vol. 1879, no. 3, pp. 032014, 2021.

[39] T. A. Stamey, J. N. Kabalin, and M. Ferrari, "Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. III. Radiation treated patients," *Journal of Urology*, vol. 141, no. 5, pp. 1084–1087, 1989.

[40] A. A. Khalaf, "The new strange generalized Rayleigh family: Characteristics and applications to COVID-19 data," *Iraqi Journal for Computer Science and Mathematics*, vol. 5, no. 3, pp. 92–107, 2024.

[41] Zhang, Q., Mahdi, G., Tinker, J., & Chen, H. (2020). A graph-based multi-sample test for identifying pathways associated with cancer progression. *Computational Biology and Chemistry*, *87*(1), 107285.

[42] Zhang, G., Zhao, W., & Sheng, Y. (2026). Variable selection for uncertain regression models based on elastic net method. *Communications in Statistics-Simulation and Computation*, *55*(2), 565–586.