Chapter 4

RESULTS

The results obtained for machine learning algorithms which include Naïve Bayes, PART, SMO, Jrip, and Random Forest are presented in section 4.1 and statistical techniques (PCA & SSE) has been illustrated in section 4.2.

4.1 Results for prediction performance of pillar stability using Machine Learning

The entire experiment was performed based on the percentage split of 80:20. The confusion matrix for Machine learning classification tools Navie Bayes, Jrip, Multilayer perception, PART, SMO, and Random Forest models have presented in Tables 4.1-4.10.

Table 4.1: ***Naive Bayes*** model's confusion matrix based on percentage split of 80:20 on the original dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	12	2	0
Unstable cases (1)	9	1	7	1
Failed cases (2)	13	1	4	8

Table 4.2: ***SMO*** model's confusion matrix based on percentage split of 80:20 on the original dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	12	1	1
Unstable cases (1)	9	1	7	1
Failed cases (2)	13	1	3	9

Table 4.3:***Jrip*** model's confusion matrix based on percentage split of 80:20 on the original dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	11	1	2
Unstable cases (1)	9	2	7	0
Failed cases (2)	13	1	4	8

Table 4.4: ***PART*** model's confusion matrix based on percentage split of 80:20 on the original dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	7	6	1
Unstable cases (1)	9	1	8	0
Failed cases (2)	13	1	2	10

Table 4.5: ***RF*** model's confusion matrix based on percentage split of 80:20 on the original dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	12	2	0
Unstable cases (1)	9	2	7	0
Failed cases (2)	13	1	3	9

Table 4.6: ***Naive Bayes*** model's confusion matrix based on percentage split of 80:20 on reduced dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	13	0	1
Unstable cases (1)	9	1	7	1
Failed cases (2)	13	1	4	8

Table 4.7: ***SMO*** model's confusion matrix based on percentage split of 80:20 on the reduced dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	13	0	1
Unstable cases (1)	9	1	7	1
Failed cases (2)	13	1	3	9

Table 4.8:***Jrip*** model's confusion matrix based on percentage split of 80:20 on the reduced dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	11	1	2
Unstable cases (1)	9	2	7	0
Failed cases (2)	13	1	3	9

Table 4.9: ***PART*** model's confusion matrix based on percentage split of 80:20 on the reduced dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	13	0	1
Unstable cases (1)	9	2	7	0
Failed cases (2)	13	2	3	8

Table 4.10: ***RF*** model's confusion matrix based on percentage split of 80:20 on the reduced dataset

Pillar Stability Condition	Actual cases	Predicted cases		
		Stable cases (0)	Unstable cases (1)	Failed cases (2)
Stable cases (0)	14	12	1	1
Unstable cases (1)	9	2	7	0
Failed cases (2)	13	1	1	11

Table 4.11: Performance metrics for different ML tools on the original dataset.

ML tools	**Accuracy (%)**	**AUC**	**MCC**
Naive Bayes	75.0	0.890	0.652
Jrip	72.2	0.810	0.593
SMO	77.8	0.845	0.674
PART	69.4	0.813	0.581
RANDOM FOREST	77.8	0.934	0.691

Table 4.12: Performance metrics for different ML tools on the reduced dataset

ML tools	**Accuracy**	**AUC**	**MCC**
Naïve Bayes	77.8	0.912	0.675
Jrip	75.0	0.827	0.628
SMO	80.6	0.854	0.711
PART	77.8	0.877	0.674
RANDOM FOREST	83.3	0.920	0.748

To rank the features based on their importance in the classification process, a feature ranking algorithm was used. The number of characteristics in the order of most significant to least significant was then varied in steps of some constant in trials on classification algorithms. Based on the results of these tests, it was feasible to identify the redundant and non-informative features and, as a consequence, eliminate them from the input feature set. When the number of features in the input feature vector is big, this strategy is particularly efficient.

Table 4.13: Ranking of different features based on Fuzzy rough features evaluator

Features	**Ranks**
(w/h) pillar width/pillar height	0.04185
($\boldsymbol{\sigma_p}/\boldsymbol{\sigma_c}$) pillar strength / pillar stress	0.02747
UCS(MPa)$\boldsymbol{\sigma_c}$	0.00745
(MPa)$\boldsymbol{\sigma_p}$	0.00566

On the other hand, the visual representation of various machine learning algorithms performed based on AUC were obtained for both original and reduced datasets are presented in Figure 4.1 and Figure 4.2. Higher the AUC value i.e. close to 1, best will be the predictor algorithm.

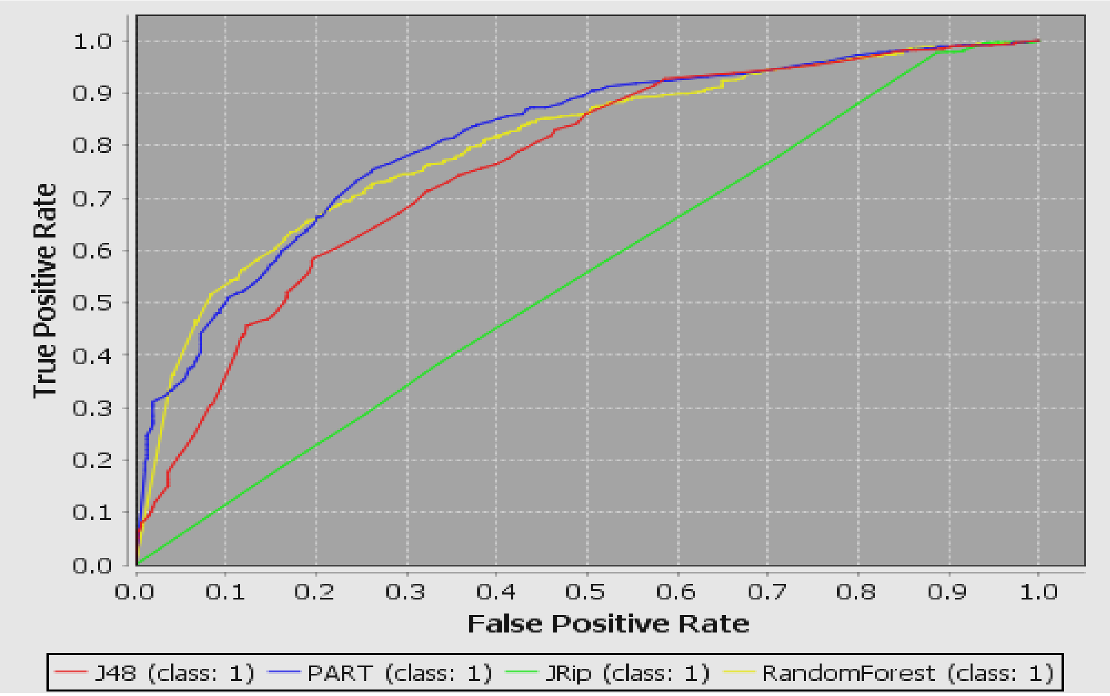

Figure 4.1. AUC for various machine learning algorithms on original dataset

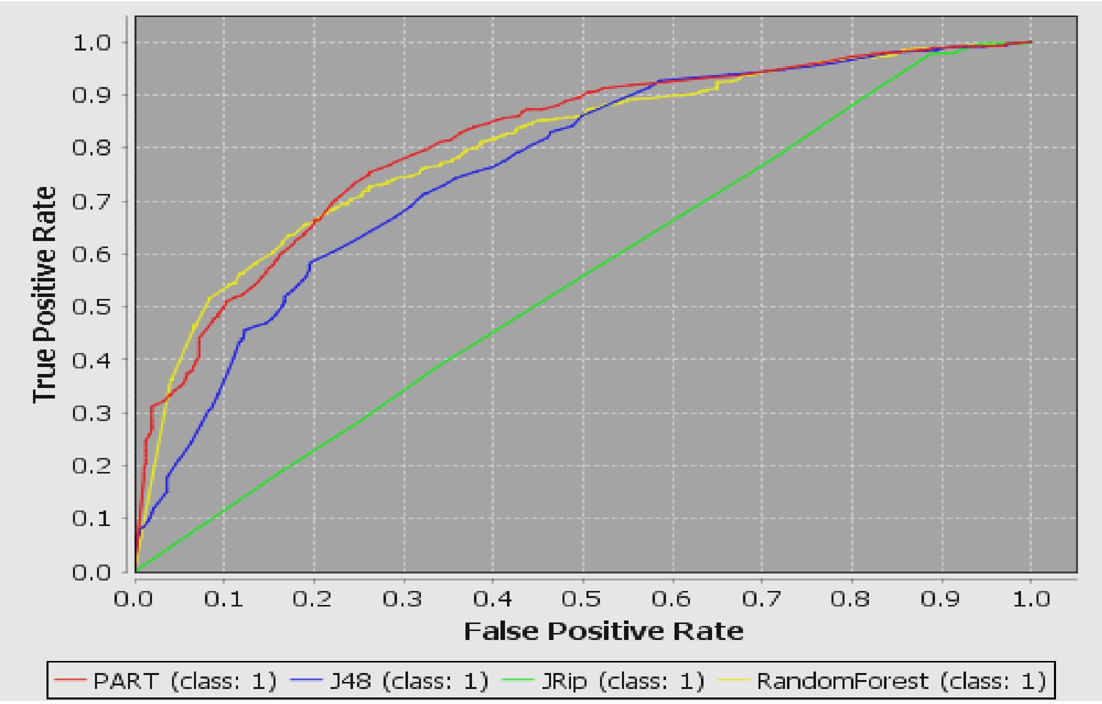

Figure 4.2. AUC for various machine learning algorithms on reduced dataset.

4.2 Results of PCA and SSE Techniques

The results obtained by both PCA and SSE techniques has been discussed in the following section one by one:

4.2.1. Results of PCA Technique

Table 4.14 presents the minimum, maximum and mean values of the seven input and an output parameter, Pillar Strength (PS) together with their respective symbols. The results of the present study have been discussed in the following sections:

Table 4.14: Descriptive information of the pillar stability data set indicated in annexure B.

	Sl. No.	**Parameters**	**Symbols**	**Min.**	**Max.**	**Mean**
Input	1	Depth (m)	D	42	300	133.20
	2	Gallery Width (m)	B	3.0	4.2	3.52
	3	Pillar Width (m)	W	9.0	35.5	19.26
	4	Pillar Height (m)	H	1.8	3.0	2.63
	5	Pillar Width/Pillar Height ratio	W/H	3.0	16.1	7.63
	6	Uniaxial Compressive Strength (MPa)	UCS	18.3	85.0	47.18
	7	Pillar Load (MPa)	PL	1.8	42.6	13.08
Output		Pillar Strength (MPa)	PS	7.9	48.0	20.71

PCA resulted in the formation of one principle component group with an eigenvalue greater than 1.0, which accounted for 79.80% of the total variance. Table 4.15 depicts the data matrix that explains the overall variance as well as the number of principal component groups. A total of one principle component group with an eigenvalue greater than one has been identified.

Table 4.15: Data matrix explaining total variance

Principal Component Group	**Initial Eigenvalues**		
	Total	% of Variance	Cumulative %
1	5.586	79.802	79.802
2	0.953	13.610	93.412
3	0.329	4.697	98.109
4	0.102	1.451	99.560
5	0.017	0.246	99.806
6	0.010	0.142	99.948
7	0.004	0.052	100.000

The P-P plot for the datasets used in the present study has been presented in figure 4.3. The R^2 value has been found as 0.989, which clearly reveals that the observed datasets used to predict the pillar strength (PS) is closely related to the predicted values. The graph shown in figure 4.3 indicates the normal distribution curve against the observed and predicted datasets, which is closely related to each other (as indicated in figure 4.3).

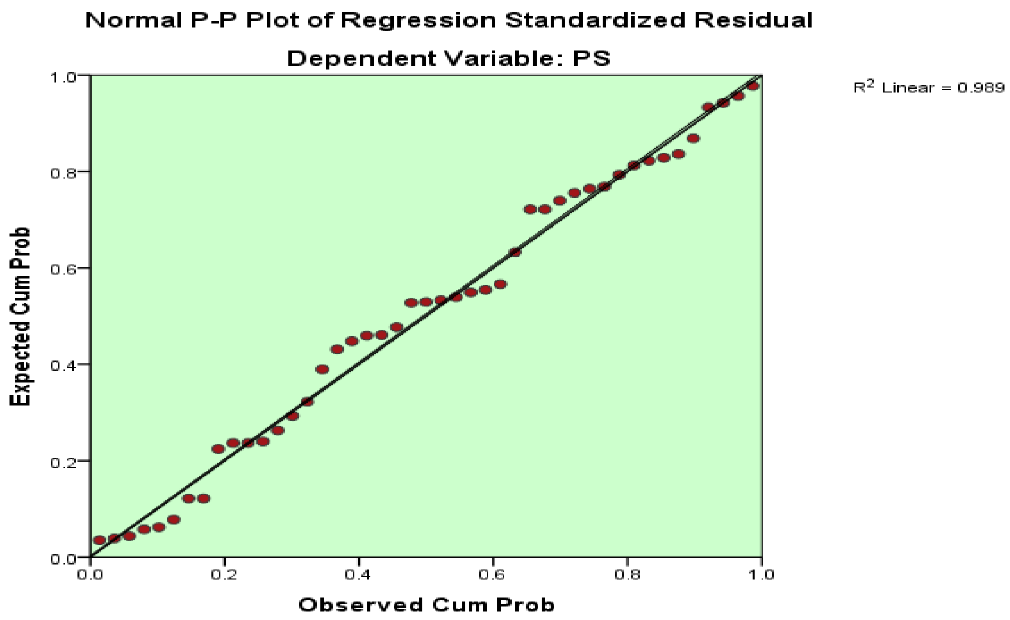

Figure 4.3: Probability plot for the datasets by PCA.

Table 4.16 illustrates the one principal component groups with the coefficient of determination (R^2) for all the pillar design parameters. The principal components of each component group were determined based on a greater R^2 value. A total number of five PCs were identified and extracted from Table 4.16, namely D, W, W/H, UCS, and PL, and kept separately in Table 4.17.

Table 4.16: Identification of principal components.

Input parameters	**Component groups with a coefficient of determination, R^2**
D	0.991
B	0.894
W	0.986
H	-0.460
W/H	0.991
UCS	0.935
PL	0.971

Table 4.17: List of principal components extracted by PCA.

Principal Component Group-1	D, W, W/H, UCS, PL

MLR analysis for Pillar strength was subsequently carried out by following the two steps as listed below:

1. For the prediction of pillar stability, MLR was carried out for all the identified five PCs (D, W, W/H, UCS, and PL).

2. After eliminating multi-collinearity, MLR analysis for PS prediction was carried out for the retained two PCs (W/H and UCS). Table 4.18 summarizes the MLR analysis for all the identified five PCs. It has been found that there is multi-collinearity among some of the principle components.

Table 4.18: MLR results for all the identified 5 PCs.

R	R^2	Adjusted R^2	Significance	SEE (Error)
0.933	0.870	0.835	0.00	0.65401

From the analysis, it was found that three out of five PCs were having multi-collinearity (VIF>10). Therefore, these three PCs were rejected. The multi-collinearity (VIF>10) among various pillar design parameters is represented in Table 4.19.

Table 4.19: Parameters with VIF>10.

Sl. No.	Input parameters	VIF values
1	D	53.12
2	W	32.51
3	PL	21.15

As a result, the new prediction model was created after deleting the three PCs that contained multi-collinearity. Thus, two PCs (W/H and UCS) were selected, free from multi-collinearity. Table 4.20 lists the pillar design parameters with VIF <10.

Table 4.20: Parameters with VIF<10

Sr.no.	Input parameters	VIF values
1	W/H	7.12
2	UCS	5.75

The MLR was applied on the selected parameters to develop the model, subsequently the unstandardized coefficients of the selected parameters together with their significance value and standard error was derived. Table 4.21 presents the unstandardized coefficients with the significance, standard error and collinearity statistics of the selected parameters.

Table 4.21: For developing the equation for PS.

Model	Unstandardized Coefficients		Significance	Collinearity Statistics	
	β	Std. Error		Tolerance	VIF
(Constant)	-7.392	0.532	.001		
W/H	2.338	0.054	.0001	0.406	7.12
UCS	0.217	0.016	.001	0.406	5.75

Finally, the model in the form of an equation Eq. 4.1 has been developed using β-value associated with the retained PCs (Table 4.22). The MLR analysis results for the developed model is presented in Table 4.23.

$$PS = 2.338\frac{W}{H} + 0.217\,UCS - 7.392 \qquad \ldots (4.1)$$

Table 4.22: MLR results for the two PCs without multi-collinearity for predicting PS

R	R^2	Adjusted R^2	Significance	SSE (Error)	F value
0.927	0.860	0.848	0.001	0.112	3609.11

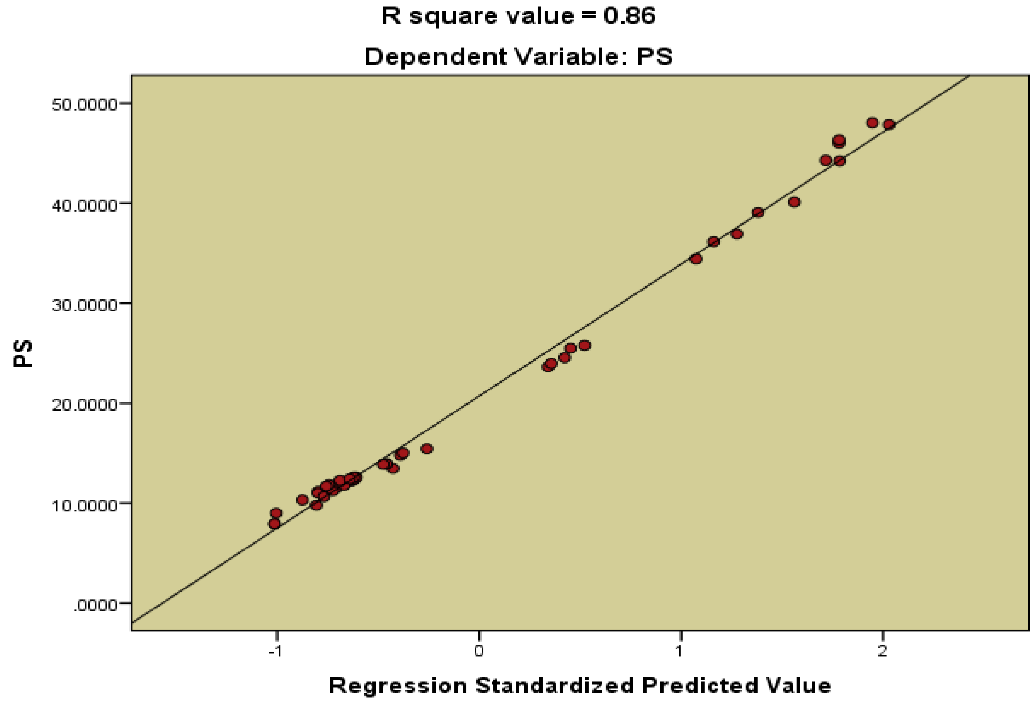

Figure 4.4: Regression plot between observed and predicted values of PS by PCA

It is clearly illustrated from the figure 4.4, that the value of (R^2) for the developed model for the prediction of PS by PCA has been found as 0.86 and adjusted (R^2) is 0.84. This shows that the two retained PCs used in the regression analysis had an 86 percent accuracy in predicting PS. The significance level is 0.001 and the F-value has been observed as 3609.117, which strengthens the obtained results.

4.2.2 Results of SSE techniques

The bivariate correlation approach with Pearson's correlation was used to identify the significantly associated pillar design parameters that exhibited significance (2-tailed) ≤ 0.05. Table 4.23 illustrates the correlation matrix as well as the parameter significance values.

Table 4.23:Correlation values and significance level of the parameters.

Parameters	**PS**	**Significance**	**N**
	Pearson’s correlation	(2-tailed)	
D	0.983	0.004	45
B	0.836	0.001	45
W	0.965	0.001	45
H	-0.469	0.000	45
W/H	0.984	0.000	45
UCS	0.860	0.001	45
PL	-0.696	0.001	45

The P-P plot for the datasets used in the present study by SSE has been presented in figure 4.5. The R^2 value has been found as 0.987, which clearly reveals that the observed datasets used to predict the pillar strength (PS) is closely related to the predicted values. The graph shown in figure 4.5 indicates the normal distribution curve against the observed and predicted datasets, which is closely related to each other (as indicated in figure 4.5).

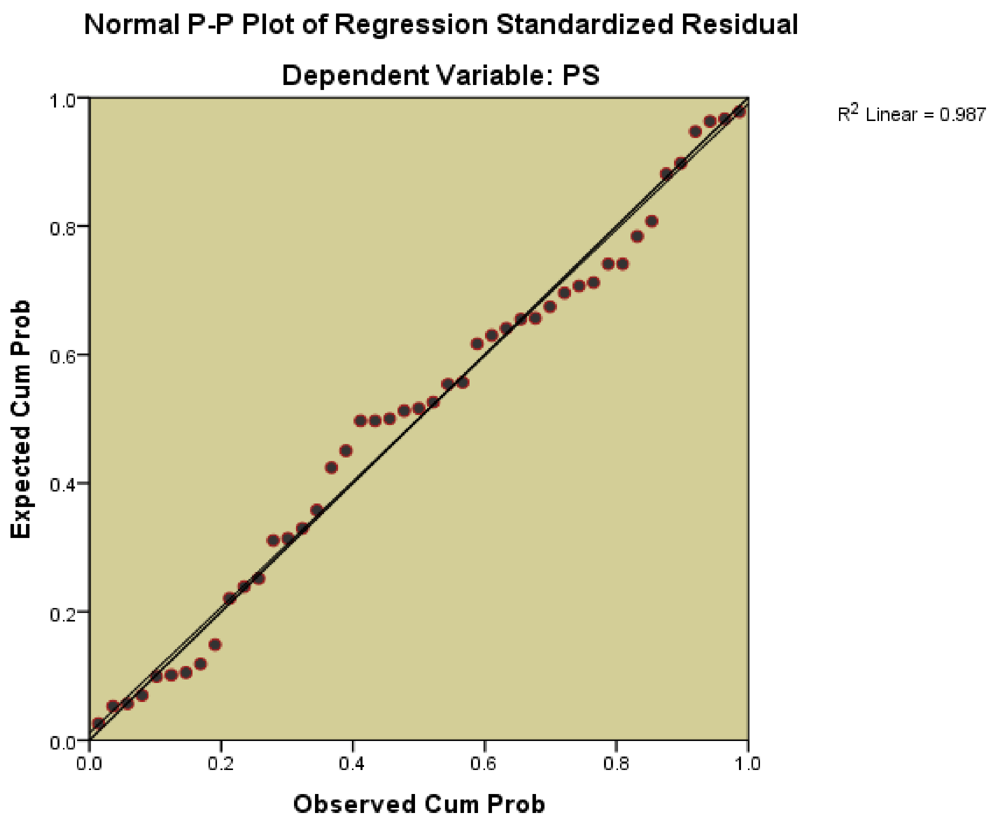

Figure 4.5: Probability plot for the datasets by SSE

Seven parameters (D, B, W, H, W/H, UCS and PL) have been selected having significance ≤ 0.05, from Table 4.10. MLR for PS prediction has been performed on the seven correlated parameters and the results are illustrated in Table 4.24.

Table 4.24: MLR results for predicting PS using all the identified parameters

R	R Square	Adjusted R Square	Significance	SSE (Error)
0.934	0.871	0.855	0.00	0.54463

MLR analysis has revealed that multi-collinearity existed in four out of seven parameters. The values of VIF for the four parameters, which showed multi-collinearity are given in Table 4.25.

Table 4.25: Parameters with VIF>10

Sl. No.	Parameters	VIF values
1	D	16.73
2	B	145.23
3	W/H	19.35
4	PL	69.52

As a result of selecting and eliminating parameters from the prediction model, multi-collinearity was eliminated, and three parameters (W, H, and UCS) were found as having no multi-collinearity. Table 4.26 presents the list of selected parameters along with their VIF values.

Table 4.26 Pillar design parameters with VIF<10

Sl. No.	Parameters	VIF values
1	W	4.028
2	H	1.323
3	UCS	3.561

The MLR has been applied on the selected parameters to develop the model, subsequently the unstandardized coefficients of the selected parameters together with their significance value and standard error have been derived. Table 4.27 presents the unstandardized coefficients with the significance, standard error and collinearity statistics of the selected pillar design parameters.

Table 4.27: Table for developing equation for PS.

Model	Unstandardized Coefficients		Significance	CollinearityStatistics	
	β	Std. Error		Tolerance	VIF
(Constant)	13.164	2.356	.000		
W	0.813	0.033	.000	.281	2.178
H	-8.489	0.054	.000	.248	1.816
UCS	0.302	0.851	.000	.756	6.996

Finally, the model in the form of an equation has been developed using β-value associated with retained PCs (Table 4.27), which is presented in Eq.4.2. The MLR analysis results for the developed model is presented in Table 4.28.

$$\text{PS (Pillar strength)} = 0.813W + 0.302UCS - 8.489H + 13.164 \qquad \ldots(4.2)$$

Table 4.28**:** MLR results for three PCs without multi-collinearity for predicting Pillar Strength (PS)

R	R Square	Adjusted R Square	Sig.	SSE (Error)	F value
0.920	0.846	0.818	0.00	0.123	2001.36

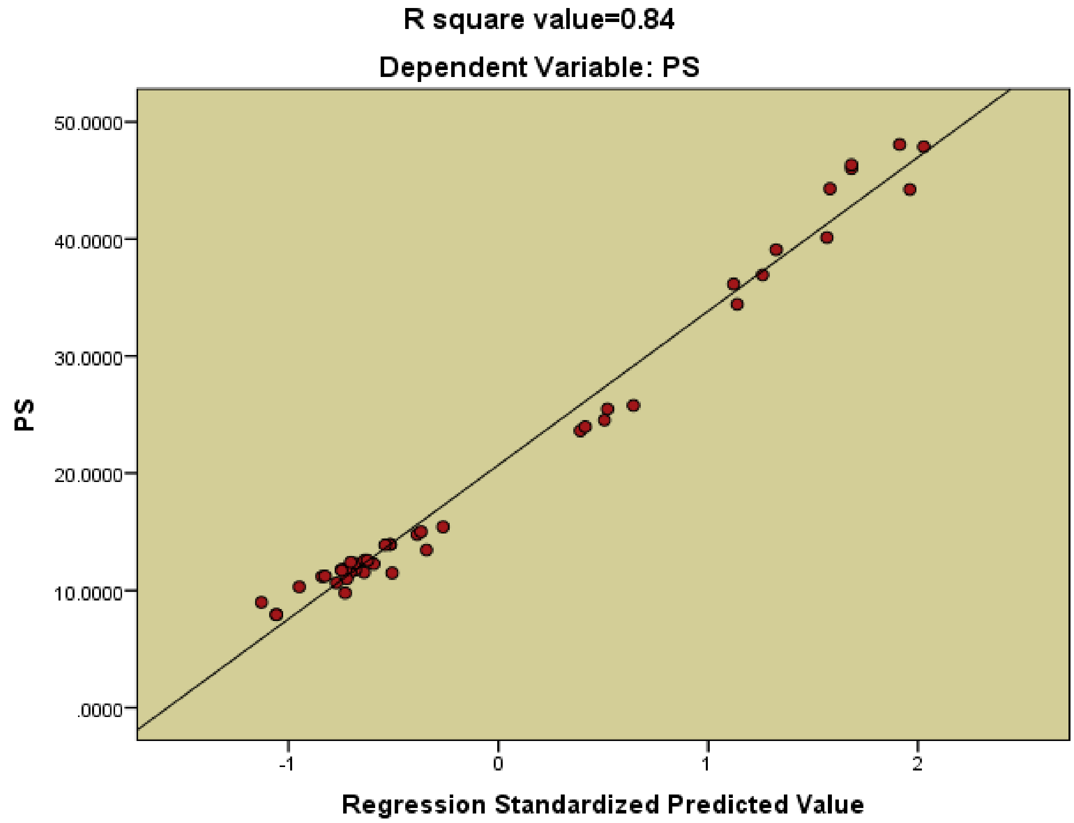

Figure 4.6: Regression plot between observed and predicted values of PS by SSE

It is clearly illustrated in the figure 4.6, that the value of (R^2) for the developed model for the prediction of PS by SSE has been found as 0.84 and adjusted (R^2) is 0.81. The value of R^2 was found to be 0.84 implying that with three selected parameters having mo any multi-collinearity, the PS has been predicted with 84 percent accuracy. Apart from the R^2 value, the F-ratio has been found as 2001.36 which is much > 4 and the significance level was 0.00, which strengthens the obtained results.