Research on Engineering Structures & Materials
www.jresm.org
Research Group

Research Article

Optimizing process performance through statistical design of experiments: A data-driven engineering approach

Shobhalatha G [1,a], Bhuvaneswara Prasad R [1,b], Charankumar Ganteda [*,2,c], Rajyalakshmi K [3,d]

[1]*Department of Mathematics, Sri Krishnadevaraya University, Ananthapuramu-515003, India*
[2]*Department of Mathematics, Siddhartha Academy of Higher Education, Deemed to be University, Vijayawada-520007, Andhra Pradesh, India*
[3]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Green fields, Vaddeswaram, Guntur, India*

Article Info

Article History:

Received 02 July 2025
Accepted 29 Oct 2025

Keywords:

Modified Taguchi design;
Response surface methodology;
Support vector regression;
XGBoost;
Random Forest;
Parameter optimization

Abstract

Over recent years, the design of experiments has emerged as a dynamic research field, attracting significant attention from scholars and practitioners. Experimental outcomes inherently exhibit variability due to measurement errors and the complex, non-linear behavior of system responses influenced by unidentified input factors. Within this context, the Taguchi method—with its use of orthogonal arrays—offers an effective framework for identifying optimal input parameters with a reduced number of experiments, typically validated through empirical test data. Conventional statistical techniques, such as the modified Taguchi model and response surface methodology, remain widely employed for parameter estimation and optimization. However, recent advances in machine learning present powerful alternatives. In this study, support vector regression, random forest regression, and XGBoost regression models were compared with traditional approaches to assess their relative efficiencies. The machine learning–based methodologies demonstrated superior predictive accuracy while significantly reducing experimental costs, preserving essential process insights, and minimizing performance variability. Among these models, the XGBoost regression approach delivered the most reliable performance, exhibiting the lowest prediction error and an exceptionally high coefficient of determination ($R^2 = 0.99$).

1. Introduction

Taguchi introduced a structured approach using orthogonal arrays that significantly reduces the number of experiments required while still capturing essential information from a full factorial design. Taguchi introduced a structured approach using orthogonal arrays that significantly reduce the number of experiments required while still capturing essential information from a full factorial design. The Taguchi method is one of the most widely used statistical techniques for improving the quality of manufactured products. According to Taguchi, the selection of appropriate control factors can effectively nullify noise factors. Each test run in the Taguchi method is classified into low (−1), medium (0), and high (1) levels of input process parameters.

Bhattacharya [1] explored the use of Response Surface Methodology (RSM) in pharmaceutical applications, providing valuable insights relevant to the present work. Cui et al. [2] optimized the physical and mechanical properties of spline surfaces using the Taguchi approach. Mounika et al.

*Corresponding author: charankumarganteda@gmail.com
[a]orcid.org/ 0000-0003-4964-7502; [b]orcid.org/ 0009-0008-8952-9969; [c]orcid.org/0000-0003-1680-2078; [d]orcid.org/0000-0001-6833-622x
DOI: http://dx.doi.org/10.17515/resm2025-998ml0702rs

[3] optimized friction welding parameters using modifications in Taguchi, fuzzy logic, and response surface methodologies. Panigrahi et al. [4] employed an integrated Taguchi and machine-learning-based optimization approach to determine the optimal design parameters of a trapezoidal solar cooker while minimizing heat loss. Rajyalakshmi et al. [5] performed a comparative evaluation of the Taguchi method, Box–Behnken Design (BBD), and Central Composite Design (CCD) for optimizing process parameters involving four factors at three levels. Their findings show that although the Taguchi method is cost-effective, BBD and CCD offer superior accuracy and precision. Rajyalakshmi et al. [5] also identified optimal parameters for Chemical Oxygen Demand (COD) reduction and decolorization efficiency using multi-objective optimization in Fenton oxidation processes, supported by strong correlations between empirical model predictions and experimental data.

Romelin et al. [6] reported that Taguchi's L9 array effectively identified optimal design points for hydraulic ram pump performance. Singhavavelu et al. [7] reported the use of an L9 orthogonal array for four factors at three levels. Panigrahi et al. [8] presented employed Taguchi and ML in solar cooker optimization .Ross [9] reported the use of an L9 orthogonal array for four factors at three levels. Rajyalakshmi [10] presented a simple methodology and validated it using existing experimental results based on the modified Taguchi method. Rajyalakshmi and Nageswara Rao [11] applied a modified Taguchi methodology to identify the optimum parameters influencing weld dilution in ST-37 plates. Samuel et al. [12] demonstrated significant improvements in the mechanical properties of biomaterials using heat-treatment optimization through Taguchi and machine-learning techniques.

Athreya and Venkatesh [13] provided foundational work upon which the present study is built. Varalakshmi et al. [14] focused on optimizing responses using various statistical design methods. Varalakshmi et al. [15] investigated multi-response optimization of agricultural residues using a modified Taguchi approach combined with Chauvenet's criterion. Rajyalakshmi and Nageswara Rao [16] .Lestari et al. [17] studied the influence of fused deposition modeling (FDM) parameters on printing time and socket weight in transtibial prosthetic components using the Taguchi method. Jou et al. [18] developed predictive defect models in die-casting using Artificial Neural Networks, Support Vector Machines, and Random Forests combined with Taguchi methods.

Traditional statistical techniques often fail to adequately capture nonlinear behaviour, motivating the present investigation. While traditional statistical techniques such as linear models and lower-order interaction frameworks have been widely used, they often fall short in capturing the intricate nonlinear relationships present in complex systems. To address these limitations, this research extends previous models by integrating advanced machine learning algorithms capable of modeling higher-order and nonlinear interactions among variables. This study systematically compares traditional statistical models (Taguchi, RSM) with modern machine learning algorithms (SVR, XGBoost, Random Forest) on a unified dataset. This hybrid approach is rarely presented in machining literature with such rigor and reproducibility. Industrial experiments often operate under resource constraints, limiting the number of design points. By demonstrating how different models perform on a 27-run factorial design, provide practical guidance for real-world applications where data is limited but precision is critical.

The experimental framework utilizes the test data originally developed by Srinivasa Athreya and Venkatesh, with modifications aligned to the suggestions of Rajyalakshmi and Nageswararao. A comparative analysis was conducted across several predictive and optimization techniques, including Response Surface Methodology (RSM), Support Vector Regression (SVR), XGBoost, and Taguchi-based linear models. Each method was evaluated based on its predictive accuracy, quantified through metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). Graphical representations were employed to visualize the performance differences and error margins between predicted and actual values.

Among the models tested, XGBoost emerged as the most accurate, demonstrating the lowest prediction error and the highest R^2 value. This indicates its superior capability in capturing complex patterns and delivering precise predictions. The study highlights the effectiveness of machine learning techniques in enhancing process modelling and optimization, especially when traditional statistical methods are constrained by linear assumptions. By integrating these advanced approaches into the existing analytical framework, the research offers a robust methodology for improving predictive performance in engineering and scientific applications.

2. Methodology

2.1 Taguchi Method

Taguchi designed a method to improve the quality of manufactured goods. According to him, way of selecting appropriate control factors can nullify the noise factors. Ross(9) suggested an appropriate orthogonal array (OA) to perform experiments is as follows.

$$N_T=1+n_P(n_L-1) \tag{1}$$

N_T: Total number of experimental runs required; n_p: Number of parameters (or factors) bei ng studied; n_L: Number of levels for each parameter; 1: Represents the baseline or control condition (reference level).

2.2 Response Surface Methodology (RSM)

Response Surface Methodology (RSM) is a statistical technique that is used to model and optimize complex systems. It is commonly used in the fields of engineering, science, and business to design and improve products or processes. RSM provides a suitable approximation using the following form.

$$y = a_0 + \sum_{i=1}^{k} a_i x_i + \sum_{i=1}^{k} a_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=i+1}^{k} a_{ij}\, x_i\, x_j + e \tag{2}$$

a_0: Constant, a_i: coefficient of the variable x_i, $\sum a_i x_i$: linear term for each variable, $\sum_{i=1}^{k} a_{ii} x_i^2$: pure quadratic terms, $\sum_{i=1}^{k} \sum_{j=i+1}^{k} a_{ij}\, x_i\, x_j$: interaction terms (combined effects of two variables acting together), e; random error.

The basic idea behind RSM is to create a mathematical model that relates the response of a system to its input variables. This model can then be used to identify the optimal settings for the input variables to achieve the desired response. RSM involves a series of experimental designs that are used to collect data on the response of the system to different combinations of input variables. The data is then analyzed using statistical techniques to create a mathematical model that describes the relationship between the input variables and the response. It is used to improve the accuracy and to explore curvature and interactions efficiently. Once the model is created, it can be used to optimize the system by identifying the input variable settings that will produce the desired response. This can help to improve product quality, reduce costs, or increase efficiency. Overall, RSM is a powerful tool for designing and optimizing complex systems. It can help to reduce the time and cost associated with trial-and-error experimentation and can lead to significant improvements in product or process performance.

2.3 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost)is a machine learning algorithm for solving complex nonlinear relationships with more accuracy, speed, and ability. It is mainly used to handle nonlinear interactions, regularization to reduce overfitting and high performance with small to large datasets. In the present paper, integrated XGBoost with Taguchi design to enhance the accuracy. It helpful to predict the unseen combination of parameter enabling virtual experimentation. It gives rank to the input parameters based on their influence on response variable. Hence, the experimental data has trained using XGBoost Regressor model. The implementation of XGBoost, involves Python libraries such as xgboost and scikit-learn. For XGBoost, the XGBRegressor class is commonly used, with key hyperparameters including n_estimators, max_depth learning_rate, subsample and regularization terms like reg_alpha and reg_lambda. These settings help control overfitting and improve model generalization.

2.4 Support Vector Regression (SVR)

Support Vector Regression (SVR)is a supervised machine learning algorithm that extends Support Vector Machines (SVM) to predict continuous numerical values. Unlike traditional regression, SVR tries to find a function that approximates the data within a specified margin of tolerance (ε) while minimizing model complexity and penalizing errors beyond the margin. It is particularly useful when you want accurate predictions with high generalization ability even on non-linear data, by using kernel functions. SVR is implemented using sklearn.svm.SVR, where important hyperparameters include kernel (e.g., 'rbf', 'linear'), C (regularization strength), epsilon (margin of tolerance), and gamma (kernel coefficient). These parameters influence the model's flexibility and error tolerance.

2.5 Random Forest Regression

Random Forest Regression is an ensemble learning method that creatively integrates the predictions of multiple decision trees to produce a reliable, accurate, and high-performance regression model. By combining the outputs of several individually weak learners, it constructs a powerful predictive framework capable of capturing complex and non-linear patterns within data. Random Forest Regression is built using sklearn. ensemble. Random Forest Regressor, with hyperparameters such as n_estimators, max_depth, max_features, min_samples_split, and bootstrap. These control the ensemble size, tree complexity, and feature sampling.

This study adopts a range of integrated methodologies including Taguchi-based linear models, Response Surface Methodology (RSM), Support Vector Regression (SVR), XGBoost, and Random Forest regression to compare their performance metrics. The objective is to identify the approach that yields the lowest prediction error, thereby reducing both the experimental cost and the time required for testing. To divide the dataset into subsets for model training and performance testing. This ensures that the model is trained on one part of the data and tested on unseen data to measure generalization ability. Hyper parameter tuning for all three models is often performed using grid search or randomized search to optimize performance metrics like MAE, RMSE, and R^2.

3. Data Acquisition

Traditional methods like Taguchi and RSM are limited to linear and low-order interactions. Machine learning models (SVR, XGBoost, Random Forest) capture nonlinear patterns and high-order interactions, improving prediction accuracy. With only 27 runs from a 3^3 factorial design, these models extract maximum insight without requiring large datasets.

This hybrid strategy blends the strengths of classical design with modern predictive power ideal for optimizing manufacturing processes, reducing variability, and achieving consistent quality. The foundational experiments conducted by Srinivasa Athreya and Venkatesh (2012) aimed to identify optimal machining parameters, namely cutting speed (X_1), depth of cut (X_2), and feed rate (X_3) to improve surface roughness during facing operations on mild steel components. These parameters were categorized into three levels: low (−1), medium (0), and high (1), and evaluated using a lathe machine equipped with a portable surface tester. Building on this work, Rajyalakshmi and Nageswara Rao (2018) introduced a novel enhancement to the Taguchi method by incorporating a fictitious parameter (X_4), which allowed for a more comprehensive analysis of variability in the output response. Their study employed Analysis of Variance (ANOVA) to quantify the contribution of each factor to surface roughness. The results revealed that cutting speed (X_1) contributed 71.49%, depth of cut (X_2) 4.18%, feed rate (X_3) 11.16%, and the fictitious parameter (X_4) 13.17% to the total variation. Notably, the error percentage was found to be zero when the fictitious parameter was included, whereas its exclusion resulted in a 13.17% error equivalent to its contribution highlighting its significance in the model. According to their findings, the optimal surface roughness value was 2.62 µm, achieved at a cutting speed of 960 rpm, depth of cut of 0.3 mm, and feed rate of 130 mm/min. The expected range of surface roughness under these conditions was estimated between 2.349 and 2.768 µm.

Expanding upon these prior studies, the present research integrates both statistical and machine learning methodologies including Taguchi design, Response Surface Methodology (RSM), Support

Vector Regression (SVR), XGBoost, and Random Forest regression—to evaluate and compare their effectiveness in optimizing process performance. Each method was assessed based on its predictive accuracy and ability to model complex interactions among machining parameters. This integrated approach provides a robust framework for parameter optimization, offering deeper insights into the relationships between input variables and surface quality outcomes.

4. Results and Analysis

This study integrates multiple Modeling techniques to establish a comprehensive predictive framework. Specifically, it compares the performance of five regression models: Taguchi-based linear model, Response Surface Methodology (RSM), Support Vector Regression (SVR), XGboost Regression, and Random Forest Regression. A paired t-test was conducted to compare the experimental (test) data and the corresponding values predicted by each model. The calculated t-statistic was smaller than the critical t-value. These results indicate that there is no statistically significant difference between the experimental values and the predicted values of the mentioned models. In other words, the model's predictions are statistically consistent with the observed data, demonstrating that successfully captured the underlying relationship within the dataset. Although five different modelling approaches - XGboost, Support Vector Regression (SVR), Random Forest (RF), Response Surface Methodology (RSM), and Taguchi method were compared, no statistically significant differences were found among them based on the selected error indicators. Figures 1 through 5 illustrate the comparison between actual and predicted values for each model: XGboost (Figure 1), SVR (Figure 2), Random Forest (Figure 3), RSM (Figure 4), and the Taguchi-based model (Figure 5). Among all models, XGboost shows an exceptional alignment with the actual experimental data, indicating its superior predictive accuracy. Table I further highlights the experimental versus predicted values, demonstrating the effectiveness of each model, with XGboost outperforming the rest. Table (2) Provides the error analysis for each model.

4.1 Sensitivity Analysis

The correlation coefficient (r) measures the strength of the linear relationship between the model predictions and the experimental (test) data.

- XGBoost (r = 0.99) shows an *almost perfect linear relationship* with the test data — it predicts values that very closely follow the observed trend.
- RF (r = 0.95) and SVR (r = 0.8913) also demonstrate strong positive correlations, suggesting reliable predictive ability.
- RSM (r = 0.8974) performs comparably to SVR, though slightly less correlated.
- Taguchi method (r = 0.59) shows only moderate correlation, implying that it captures the general trend but with considerable deviation from the actual test data.

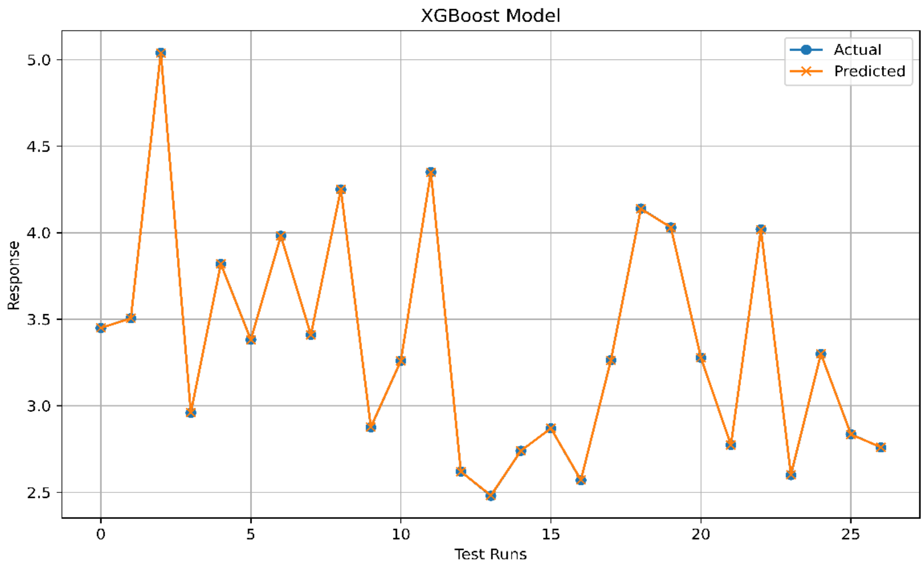
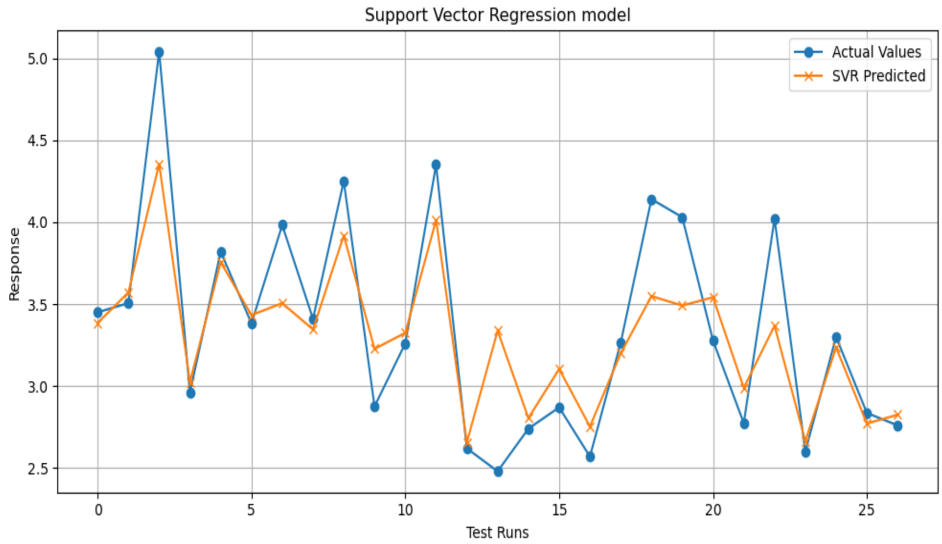


Fig. 1. XGBoost model



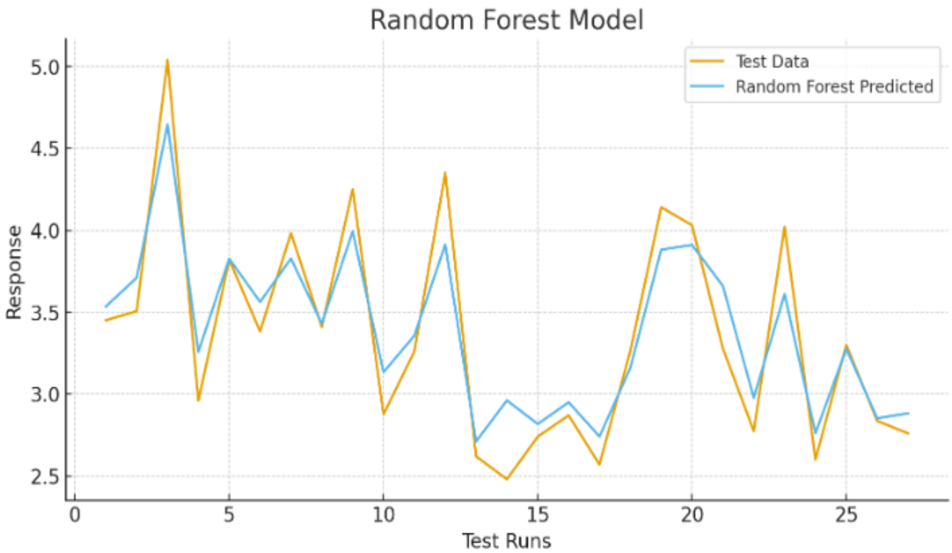


Fig. 3. Random forest regression

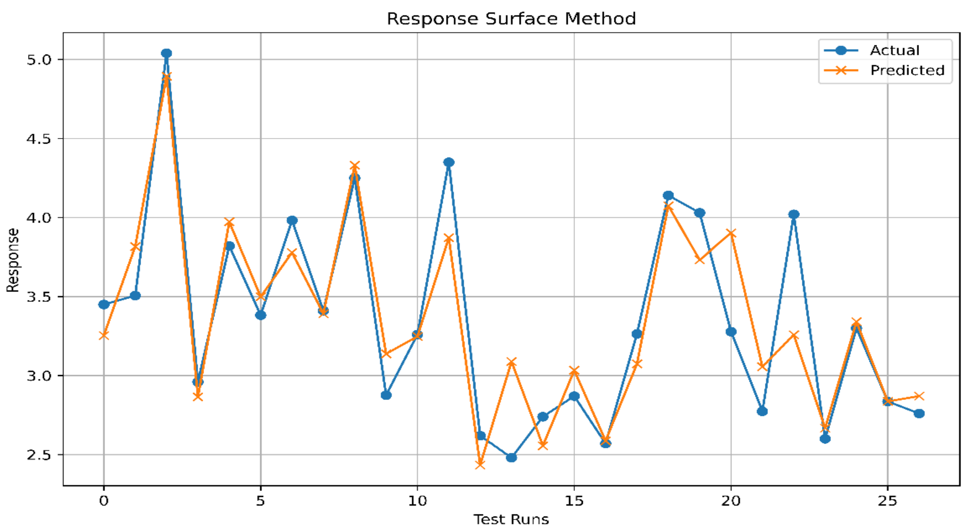


Fig. 4. Response surface methodology

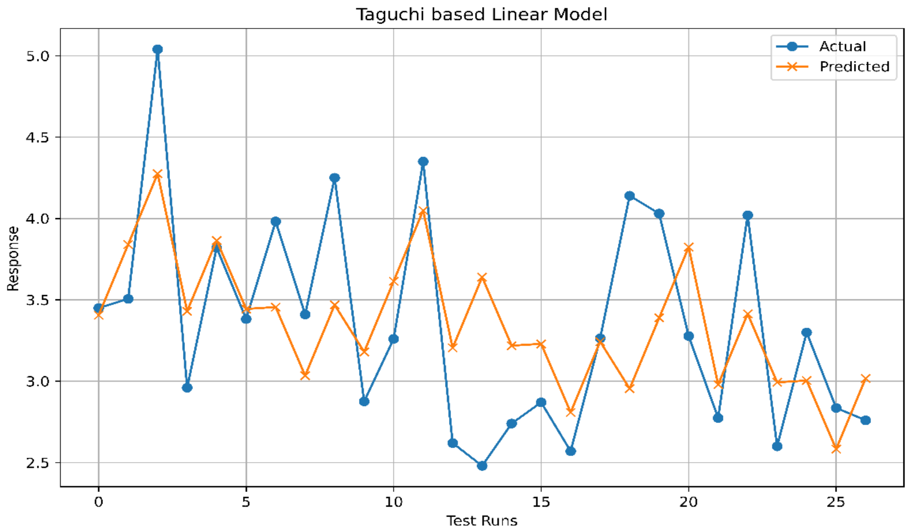


Fig. 5. Taguchi based linear model

The error values for each model (XGBoost, Support Vector Regression, Random Forest, Response Surface Methodology, and Taguchi-based Linear Model) were computed using the test dataset, based on the difference between the predicted and actual values. Each model performance was evaluated using the following error metrics,

$$Mean\ Absolute\ Error\ (MAE) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

where, y_i: observed values, $\hat{y}_i$: predicted values, n: number of observations.

$$Mean\ Squared\ Error\ (MSE) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

$$Root\ Mean\ Squared\ Error\ (RMSE) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$Mean\ absolute\ percentage\ error\ (MAPE) = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{6}$$

$$\text{Coefficient of determination } (R^2) = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{7}$$

Table 1. Experimental and predicted values using different regression models

S.NO.	X1	X2	X3	X4	TEST DATA	PREDICTED (1) XGBOOST	PREDICTED (2) Support Vector Regression	PREDICTED (3) Random Forest Regression	PREDICTED (4) Response Surface Method	PREDICTED (5) Taguchi based linear model
1	-1	-1	-1	-1	3.45	3.450144	3.385071	3.5339	3.253852	3.405593
2	-1	-1	0	0	3.506	3.505383	3.570623	3.7098	3.816685	3.840481
3	-1	-1	1	1	5.04	5.037933	4.354874	4.6458	4.893519	4.27537
4	-1	0	-1	0	2.96	2.96144	3.024591	3.2573	2.864519	3.430481
5	-1	0	0	1	3.82	3.819308	3.755149	3.8261	3.972685	3.86537
6	-1	0	1	-1	3.382	3.380852	3.43307	3.562	3.498352	3.444259
7	-1	1	-1	1	3.982	3.981829	3.505373	3.8274	3.778519	3.45537
8	-1	1	0	-1	3.41	3.410325	3.34553	3.4301	3.391519	3.034259
9	-1	1	1	0	4.25	4.250471	3.918938	3.992	4.330352	3.469148
10	0	-1	-1	-1	2.876	2.877862	3.227187	3.133	3.13763	3.179926
11	0	-1	0	0	3.26	3.260162	3.324519	3.3564	3.24763	3.614815
12	0	-1	1	1	4.35	4.34951	4.012648	3.9123	3.87163	4.049704

13	0	0	-1	0	2.62	2.619565	2.656429	2.7106	2.49363	3.204815
14	0	0	0	1	2.48	2.482483	3.337663	2.9611	3.088963	3.639704
15	0	0	1	-1	2.74	2.739819	2.804515	2.8173	2.556296	3.218593
16	0	1	-1	1	2.87	2.87007	3.104081	2.9498	3.032963	3.229704
17	0	1	0	-1	2.57	2.571641	2.752938	2.741	2.58763	2.808593
18	0	1	1	0	3.264	3.262257	3.199616	3.1598	3.07363	3.243481
19	1	-1	-1	-1	4.14	4.13794	3.548494	3.8808	4.074519	2.954259
20	1	-1	0	0	4.03	4.030339	3.491223	3.9103	3.731685	3.389148
21	1	-1	1	1	3.278	3.278382	3.541257	3.6593	3.902852	3.824037
22	1	0	-1	0	2.774	2.773647	2.989197	2.9778	3.055852	2.979148
23	1	0	0	1	4.02	4.017954	3.368106	3.6101	3.258352	3.414037
24	1	0	1	-1	2.6	2.603139	2.66463	2.7612	2.667352	2.992926
25	1	1	-1	1	3.3	3.300311	3.23555	3.2762	3.340519	3.004037
26	1	1	0	-1	2.836	2.834259	2.771423	2.852	2.836852	2.582926
27	1	1	1	0	2.76	2.76097	2.824781	2.8824	2.870019	3.017815

The error analysis confirms that XGBoost is the most accurate and robust model among all considered. Its ability to consistently predict values very close to the actual test data across all conditions makes it the best choice for this dataset. Traditional linear or response surface methods show significant limitations, especially in handling nonlinearities or interactions in the input features.

- XGBoost clearly outperforms all other models, achieving the lowest error (MAE: 0.007, RMSE: 0.009) and a near-perfect R^2 score (~0.9999), indicating that it almost perfectly fits the data. Highly sensitive to input variation but robust against noise.
- Support Vector Regression (SVR) and Random Forest Regression also show strong predictive capabilities with high R^2 scores and relatively low errors, making them reliable alternatives. Acceptable performance but moderately less robust under varying conditions.
- Response Surface Method (RSM) performs moderately well but not as accurately as SVR or XGBoost. It Performs reasonably well but limited in modeling nonlinear effects.
- The Taguchi-based Linear Model shows the least accuracy, indicating that linear assumptions may not capture the complexity of the data.
- The dataset contains only 27 design points. XGBoost's boosting mechanism is better suited to extracting signal from limited data, whereas Random Forest may suffer from variance and instability due to its reliance on random sampling.

Table 2. Comparison of prediction errors for different modelling approaches

S.NO.	X1	X2	X3	X4	TEST DATA	ERROR (1) XG BOOST	ERROR (2) Support Vector Regression	ERROR (3) Random Forest Model	ERROR (4) RSM	ERROR (5) Taguchi based linear model
1	-1	-1	-1	-1	3.45	0.00417	1.881992	2.4313	5.6855	1.2872
2	-1	-1	0	0	3.506	0.01759	1.843218	5.8118	8.8615	9.5403
3	-1	-1	1	1	5.04	0.04101	13.593771	7.8218	2.9064	15.171
4	-1	0	-1	0	2.96	0.04866	2.182137	10.044	3.2257	15.895
5	-1	0	0	1	3.82	0.01813	1.69768	0.1592	3.997	1.1877
6	-1	0	1	-1	3.382	0.03393	1.510066	5.3223	3.4403	1.8409
7	-1	1	-1	1	3.982	0.00429	11.969532	3.882	5.11	13.225
8	-1	1	0	-1	3.41	0.00954	1.890602	0.5894	0.542	11.019
9	-1	1	1	0	4.25	0.01109	7.789693	6.0701	1.8906	18.373
10	0	-1	-1	-1	2.876	0.06474	12.210941	8.9374	9.097	10.568
11	0	-1	0	0	3.26	0.00498	1.97912	2.9571	0.3795	10.884

12	0	-1	1	1	4.35	0.01126	7.755219	10.061	10.997	6.9034
13	0	0	-1	0	2.62	0.01662	1.390413	3.4595	7.1134	22.321
14	0	0	0	1	2.48	0.10011	34.583173	19.399	24.555	46.762
15	0	0	1	-1	2.74	0.00662	2.354546	2.8204	6.7045	17.467
16	0	1	-1	1	2.87	0.00243	8.156142	2.7812	5.6782	12.533
17	0	1	0	-1	2.57	0.06384	7.118204	6.6553	0.686	9.2838
18	0	1	1	0	3.264	0.05341	1.972538	3.1918	5.8324	0.6286
19	1	-1	-1	-1	4.14	0.04976	14.287578	6.2609	1.5817	28.641
20	1	-1	0	0	4.03	0.00841	13.369167	2.9692	7.4024	15.902
21	1	-1	1	1	3.278	0.01165	8.031029	11.632	19.062	16.658
22	1	0	-1	0	2.774	0.01273	7.75765	7.3453	10.16	7.3954
23	1	0	0	1	4.02	0.05089	16.216281	10.197	18.946	15.074
24	1	0	1	-1	2.6	0.12073	2.485776	6.2008	2.5905	15.113
25	1	1	-1	1	3.3	0.00941	1.953034	0.7224	1.2278	8.9686
26	1	1	0	-1	2.836	0.0614	2.277031	0.5642	0.03	8.9236
27	1	1	1	0	2.76	0.03513	2.347153	4.4355	3.9862	9.3411

Table 3. Comparison of key performance metrics

S. No.	MODEL	MAE	MSE	RMSE	MAPE	R^2
1	Taguchi based Linear Model	0.4292	0.2690	0.5186	13%	0.355
2	Response Surface Method	0.2101	0.0819	0.2862	6.36%	0.8037
3	Support Vector regression	0.244	0.115	0.338	7.06%	0.7247
4	XGboost Regression	0.001	0.000002	0.001326	0.0323%	0.9999
5	Random Forest Regression	0.5463	0.4304	0.6561	17.58%	0.1755

5. Conclusion

The results clearly demonstrate that XGBoost Regression offers the best predictive performance among the evaluated models, with exceptionally low error values (MAE: 0.001, MSE: 0.000002, RMSE: 0.001326, MAPE: 0.0323%) and an R^2 of 0.9999, indicating near-perfect correlation with actual values. Importantly, the predicted optimum value from XGBoost is 2.48, which falls well within the target range of 2.349 to 2.768 established by previous studies. This value is also closer to the ideal surface roughness compared to prior research, which reported an optimum of 2.62. This suggests that XGBoost not only provides a more accurate model but also identifies more optimal conditions all'3' factors at medium levels for minimizing surface roughness. Furthermore, although other models (like RSM and SVR) also produced predictions within the acceptable range, some yielded slightly higher optimum response values than desired. XGBoost can be considered the most accurate model for this specific dataset, showing the strongest agreement with the experimental results. However, its generalization to new or unseen data cannot be guaranteed without further validation steps such as:

- Cross-validation or external test datasets,
- Sensitivity and robustness analyses, and
- Hyperparameter optimization with regularization to prevent overfitting.

From an engineering point, while XGBoost offers high accuracy and flexibility, it should be applied with caution in experimental studies involving limited data. Incorporating larger datasets, applying regularization techniques, and conducting sensitivity or uncertainty analyses are essential to ensure reliable predictions. Future work should focus on integrating hybrid models and ensemble averaging to achieve a balance between prediction accuracy and physical interpretability.

Acknowledgment

This research has been funded by RUSA at Sri Krishnadevaraya University through project number SKURUSA 2.0-04.

References

[1] Bhattacharya S. Central composite design for response surface methodology and its application in Pharmacy. London: IntechOpen; 2021. https://doi.org/10.5772/intechopen.95835

[2] Cui F, Su Y, Xu S, Liu F, Yao G. Optimization of the physical and mechanical properties of a spline surface fabricated by high-speed cold roll beating based on Taguchi theory. Math Probl Eng. 2018;2018:1-12. https://doi.org/10.1155/2018/8068362

[3] Mounika G, Rajyalakshmi K, Rajkumar GVS, Sravani D. Prediction and optimization of process parameters using design of experiments and fuzzy logic. Int J Interact Des Manuf. 2024;18(4):2333-2343. https://doi.org/10.1007/s12008-023-01446-x

[4] Nikhila Sri D, Kottapalli R, Pavani A, Ganteda C, Gouthami E, Abd-Elmonem A, et al. Comparison between response surface methodology and Taguchi method for dyeing process parameters optimization in fabric manufacturing by empirical planning. Sci Rep. 2025;15(1):1-10. https://doi.org/10.1038/s41598-025-94919-w

[5] Rajyalakshmi K, Medatati V, Boggarapu NR. Optimal Fenton Process Using the Modified Taguchi Approach. Thail Stat. 2025;23(2):377-392.

[6] Rajyalakshmi K, Nageswara Rao B. Modified Taguchi approach to trace the optimum GMAW process parameters on weld dilution for ST-37 steel plates. ASTM Int J Test Eval. 2019;47(4):3209-3223. https://doi.org/10.1520/JTE20180617

[7] Romelin C, Zahedi, Nusantara B. Comparative Analysis of Response Surface Methodology (RSM) and Taguchi Method: Optimization Hydraulic Ram Pump Performance. Oper Res Forum. 2024;5:85. https://doi.org/10.1007/s43069-024-00359-z

[8] Panigrahi BN, Nayak J, Bhatra J, Sahoo SS, Gountia D, Choudhury BK. Experimentation, Taguchi and machine learning-based optimisation of parameters concerning heat losses from the trapezoidal box-type solar cooker. Energy Sources Part A. 2024;46(1):60-76. https://doi.org/10.1080/15567036.2024.2416084

[9] Ross PJ. Taguchi techniques for quality engineering. 2nd ed. New York: Tata McGraw Hill; 2005.

[10] Rajyalakshmi K, Boggarapu NR. Expected range of the output response for the optimum input parameters utilizing the modified Taguchi approach. Multidiscip Model Mater Struct. 2018;15(2):508-522. https://doi.org/10.1108/MMMS-05-2018-0088

[11] Rajyalakshmi K, Nageswara Rao B. Modified Taguchi approach to trace the optimum GMAW process parameters on weld dilution for ST-37 steel plates. ASTM Int J Test Eval. 2019;47(4):3209-3223. https://doi.org/10.1520/JTE20180617

[12] Samuel BO, Alabi AA, Lawal SA, Peter E, Ibrahim TK. Optimizing the effect of heat treatment on the mechanical properties (tensile Strength and hardness) of Hyphaene Thebaica nut; A machine learning and Taguchi approach. Heliyon. 2024;10(19):e38899. https://doi.org/10.1016/j.heliyon.2024.e38899

[13] Athreya S, Venkatesh YD. Application of Taguchi method for optimization of process parameters in improving the surface roughness of lathe facing operation. Int Ref J Eng Sci. 2012;1(3):13-19.

[14] Varalakshmi M, Rajyalakshmi K, Nageswara Rao B. Optimization of process parameters using different statistical designs. J Eng Sci Technol. 2022;17(1):523-533.

[15] Varalakshmi M, Rajyalakshmi K. Optimization of responses using balanced ternary designs. Int J Adv Sci Technol. 2020;29(5):4771-4775.

[16] Varalakshmi M, Rajyalakshmi K, Charankumar G, Vu VD, Nguyen VD, Ngoc-Hung, et al. Multiresponse optimization of agricultural residues using modified taguchi approach and statistical conditioning of the data. Int J Agricult Stat Sci. 2021;17(1):1387-1394.

[17] Lestari WD, Adyono N, Faizin AK, Haqiyah A, Sanjaya KH, Nugroho A, et al. Optimization of 3D printed parameters for socket prosthetic manufacturing using the taguchi method and response surface methodology. Results Eng. 2024;21:101847. https://doi.org/10.1016/j.rineng.2024.101847

[18] Jou YT, Silitonga RM, Sukwadi R. A study on the construction of die-casting production prediction model by machine learning with Taguchi methods. J Chin Inst Eng. 2023;46(5):540-550. https://doi.org/10.1080/02533839.2023.2204880