

Prediction and Optimization of MIG Welding Parameters using Random Forest Regression and Pareto Front Analysis

Subhajit Bhattacharya¹, Sourav Das¹, Diya Mondal¹, Manohar Mahato¹, Debkanta Ghosh¹

¹Department of Mechanical Engineering, Dr. B. C. Roy Engineering College, Durgapur 713206

Abstract

MIG welding is widely used in industries and optimization and selection of correct set of welding parameters is very much essential for quality welds to be produced. In the current experiments firstly machine learning algorithm random forest regression has been applied for prediction and optimization of welding parameters. For comparison and better performance Pareto front analysis was performed on top of it. The performance and results has been explained through visualization for better understanding and gaining insight. It was found that random forest regression performed quite well despite limited number of dataset.

Keywords: Machine Learning, GMAW, Random forest, Pareto front, prediction, optimization.

1. Introduction

For optimizing welding parameters various techniques have been applied by various searchers throughout the world. For this purpose techniques like multiple linear regressions (MLR) have been used extensively. Jayant et al [1] applied analytic hierarchy process for selection of best welding process for welding of high pressure vessels. Twelve competing parameters were chosen for the multi criteria decision making process. Jeet et al [2] applied TOPSIS for optimizing the multi response process of GMAW while welding disparate metals SS202 and AISI 1018. They also developed a predictive model by genetic algorithm and applied simulated annealing for optimization. Bhattacharya et al [3] applied AHP for optimizing process parameters of MIG welding for welding high carbon steel. Zhu et al [4] first applied Douglas-Puke algorithm to extract features points for weld quality. For online monitoring of weld quality they relied on random forest regression (RFR) ensemble learning and also compared the result with ID3 algorithm of decision tree. They found that accuracy of RFR was quite higher than ID3 or CART algorithm. Munghate et al [5] predicted bead geometry using machine learning algorithm while TIG welding austenitic steel. They applied classification models RFR to predict the bead geometry fittingly. Proper correlation was also established between flux deposition, process parameters and bead geometry. Based on random forest approach Babaiyan et al [6] developed a hybrid mishmash of SVM (support vector machine) and RVM (relevance vector machine) for predicting precise weld bead geometry of GMAW process. Their proposed model was highly

accurate and more than that could be obtained by SVM or RVM alone. Rajesh et al [7] optimized for ultimate tensile strength of dissimilar metal welding of duplex 2304 stainless steel and Inconel 718 by laser beam welding (LBW). The variables parameters were focal position, welding speed and laser power. The optimum value of UTS was determined by running as random forest model. Upon validation test they confirmed an error of 0.814 % compared with the predicted result of the model. Nampoothiri et al [8] employed various machine learning techniques including ANN, RFR, MLR and ANFIS (adaptive neuro-fuzzy inference system) to predict micro-hardness and hot crack length while welding Inconel 625 by ultrasonically assisted TIG process. The variables of welding were current, ultrasonic vibration, gas flow rate and filler metal. Amongst the model RFR performed the best with highest accuracy regarding together training and testing data on both crack length and microhardness. Zhang et al [9] performed real time detection of weld defects by taking arc spectrum while robotic arc welding of aluminium alloy. They first extracted 50 features of the arc spectrum and after pre-processing performed quantitative index of feature importance through mean decrease Gini and mean decrease accuracy to obtain six spectral features of importance. Based on random forest they established a defect identification model and important features of the arcs were found out. Han et al [10] employed a visual sensing technology for online monitoring of defects in MIG welding. With the help of SDM method they extracted the edge features of images of molten pool. For identifying and classifying penetration status they applied random forest on the features like molten pool area, half-length, width and back width. The accuracy of the model on classification was 89.8 % which is quite acceptable. Chandra et al [11] took data from a thermal camera while welding 304L steel by TIG welding process. The TIG welding process was equipped with activated flux and it was called activated TIG or A-TIG. The applied five machine learning algorithms of supervised learning category for predicting surface temperature of the welding and to arise with a predictive model. Variable of consideration was amount of activated flux. Decision tree and ensemble learning methods were implemented effectively. Among these random forests yielded least error. Kumar et al [12] experimented on MIG welding with varying welding current, voltage and gas flow rate. Their target features were tensile strength and hardness. To evaluate the test factor impacts they implemented regression models, neural networks, ridge, and lasso methods. Gradient boosting and random forest regression proved to be efficiently predicting with higher R^2 values and lower errors. Mezher et al [13] performed dissimilar metal welding of austenitic stainless steel and grade-2 titanium alloy by resistance spot welding. They observed for shear force, micro-hardness, and failure mode for 100 test samples varying with process parameters like welding current, time, pressure, squeeze time, pulse welding and holding time. For accurately analyse the experimental data they used models of GBR, CatBoost, and RFR along with ANN. ANN model was proved to be the best with Polak-Ribiere training function. The second best was RFR. Zhang et al [14] investigated the seam strength of laser welded of aluminium lithium alloy by extreme gradient boosting (XGBoost) decision tree and optical spectrum. Superior complementarity was demonstrated between RF (random forest) and PCA (principal component analysis). They proposed a novel regression model, namely RFPCA-XGBoost. It was found that the proposed model performed the best with

R2 value of 0.9383. Rameshkumar et al [15] experimented on SMAW process with voltage, current, and sound sensors. The sensors data was classified with good and bad welding. They carried out signal processing and statistical features in time domain were extracted. For statistical modelling they employed ML algorithms like classification and regression tree (CART) and SVM. With quadratic kernel function SVM was trained and the model generated an accuracy of 99%. Yang et al [16] investigated for making an expert system for robotic GMAW automation where taking the user input the system by its own will decide the welding parameters for optimum welding. They combined the XGBoost algorithm and database technology for this purpose.

Grujicic et al[17] had developed a multi-physics based model where they used six modules of different physical characteristics of the process. They upgraded the system by incorporating functional relationship between the model outputs and welding parameters. Zhao et al [18] proposed a grey wolf algorithm which is adaptive in nature to optimize a four factor, five level experimental data of GMAW. The algorithm performed well and to attain required quality of geometry they used TOPSIS algorithm and analysed the data. Finally from Pareto front they arrived at their optimum condition. Thakar et al [19] experimented on welding of S690Q1 steel by GMAW process. They utilized metal-cored filler wire. Their input features were voltage, current, and gas flow rate. Target features selected were penetration depth, width, reinforcement, and HAZ width. They designed the experiment by Box-Behnken of response surface methodology. They applied mathematical regression models and heat transfer search algorithm. For estimation of critical bead geometry HAS algorithm was proven to be very precise and that was validated by experimental data. Data envelopment analysis is a popular linear programming method for optimization. Rocha et al [20] combined DEA with RSM for better performance. Five quality characteristic of GMAW was considered and they optimized those correlated responses to arrive at a singular objective function. They also employed Taguchi multi response design in conjunction with principal component analysis. Both the methods were very close in terms of the solution. However they concluded that DEA performed better.

2. Research Methodology

- (I) Experimental data were obtained and cleaned first.
- (II) Pre-processing of the cleaned data was performed. This included converting the categorical data into quantifiable form to be used by the algorithms.
- (III) Correlation matrix of the experimental data was obtained for gaining some idea of importance of features.
- (IV) Random forest regression (RFR) was applied first and optimization was performed.
- (V) One search algorithm was carried out for finding a new set of optimal data and validation test was performed.

- (VI) Pareto front analysis was carried out as the problem poses a multi criteria decision problem and optimality points were found.

3. Experimental work

The MIG welding experiments were performed using ESAB INDIA LIMITED make AUTOK 400 model. It has 0-400 ampere range and 0-75 voltage range. The wire feed motor is made by Matsushita Industrial Equipment Co. Limited with model number PM 12 MA 14 K. The parameters that were varied are welding current, voltage and speed. Work piece material was selected as medium carbon steel. The tests that were performed are dye penetration test, visual test and bend-rebend test. Dye penetration test were performed for penetration and blow holes. Spatter and deposition were tested visually and categorically described. Blow holes also were categorically described. Quantitative values were taken for bending load of the weld through bend-rebend test in universal testing machine.

The experimental data after cleaning is shown in table-1.

Table-1: Experimental data

Sl. No.	Speed (mm/min)	Voltage (V)	Current (A)	Spatter	Penetration	Blow Hole	Deposition	Crack Description	Strength(kN)
E1	370.5	25	140	Large	Less	Large	Bad weld	Transverse, HAZ, longitudinal, under-bead	8.8
E2	370.5	25	150	Less	Good	Large	Thin weldment deposition	Toe crack	15.8
E3	370.5	25	160	Less	Good	Less	Thin deposition	Transverse, longitudinal	9.2
E4	370.5	30	140	No	Very good	Very less	Continuous deposition	Longitudinal	7.8
E5	370.5	30	150	No	Very good	Very less	Continuous	HAZ crack	9

							deposition		
E6	370.5	30	160	No	Desirable	No	Thick and continuous	Transverse	10.2
E7	475.75	25	140	Some	Less	Less	Discontinuous deposition	Toe crack	7
E8	475.75	25	150	Little	Less	Less	Not smooth deposition	Transverse, longitudinal, under-bead	7.8
E9	475.75	25	160	No	Good	No	Thin deposition, better	Transverse, longitudinal, Root	6.5
E10	475.75	30	140	No	Good	No	Desirable weld,	No crack	13
E11	475.75	30	150	No	Good	No	Desirable weld, more undercut	No crack	13.4
E12	475.75	30	160	No	Very good	No	Desirable weld,	No crack	16

It can be seen that excepting the bending load all the output features are categorical in nature. In a systematic procedure those categorical output variables were converted into quantitative terms by defining some functions in python. The functions were defined by the user.

The preprocessed data is shown in table-2.

Table-2: Preprocessed Experimental Data

Experiment No	Weld Speed	Weld Voltage	Weld Current	Spatter	Penetration	Blow Hole	Metal Deposition	Crack	Bending Load
E1	370.5	25	140	3	1	3	0	3	8.8
E2	370.5	25	150	1	2	3	2	1	15.8
E3	370.5	25	160	1	2	1	1	1	9.2
E4	370.5	30	140	0	3	1	2	1	7.8
E5	370.5	30	150	0	3	1	2	1	9
E6	370.5	30	160	0	2	0	3	1	10.2
E7	475.75	25	140	2	1	1	0	1	7
E8	475.75	25	150	1	1	1	0	3	7.8
E9	475.75	25	160	0	2	0	2	3	6.5
E10	475.75	30	140	0	2	0	0	0	13
E11	475.75	30	150	0	2	0	0	0	13.4
E12	475.75	30	160	0	3	0	0	0	16

The statistical description of the data is shown in table-3.

Table-3: Statistical description of the data

ex	Weld Speed	Weld Voltage	Weld Current	Spatter	Penetration	Blow Hole	Metal Deposition	Crack	Bending Load
count	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0
μ	423.125	27.5	150.0	0.667	2.0	0.917	1.0	1.25	10.375
σ	54.965	2.612	8.528	0.9847	0.7385	1.084	1.1282	1.1381	3.338
min	370.5	25.0	140.0	0.0	1.0	0.0	0.0	0.0	6.5
25%	370.5	25.0	140.0	0.0	1.75	0.0	0.0	0.75	7.8
50%	423.125	27.5	150.0	0.0	2.0	1.0	0.5	1.0	9.1
75%	475.75	30.0	160.0	1.0	2.25	1.0	2.0	1.5	13.1
max	475.75	30.0	160.0	3.0	3.0	3.0	3.0	3.0	16.0

4. Results and discussion

For gaining insight about the data pair plot of all the variables are constructed and shown in figure-1. From this pair plots frequency distribution of all the input and output variables are visualized.

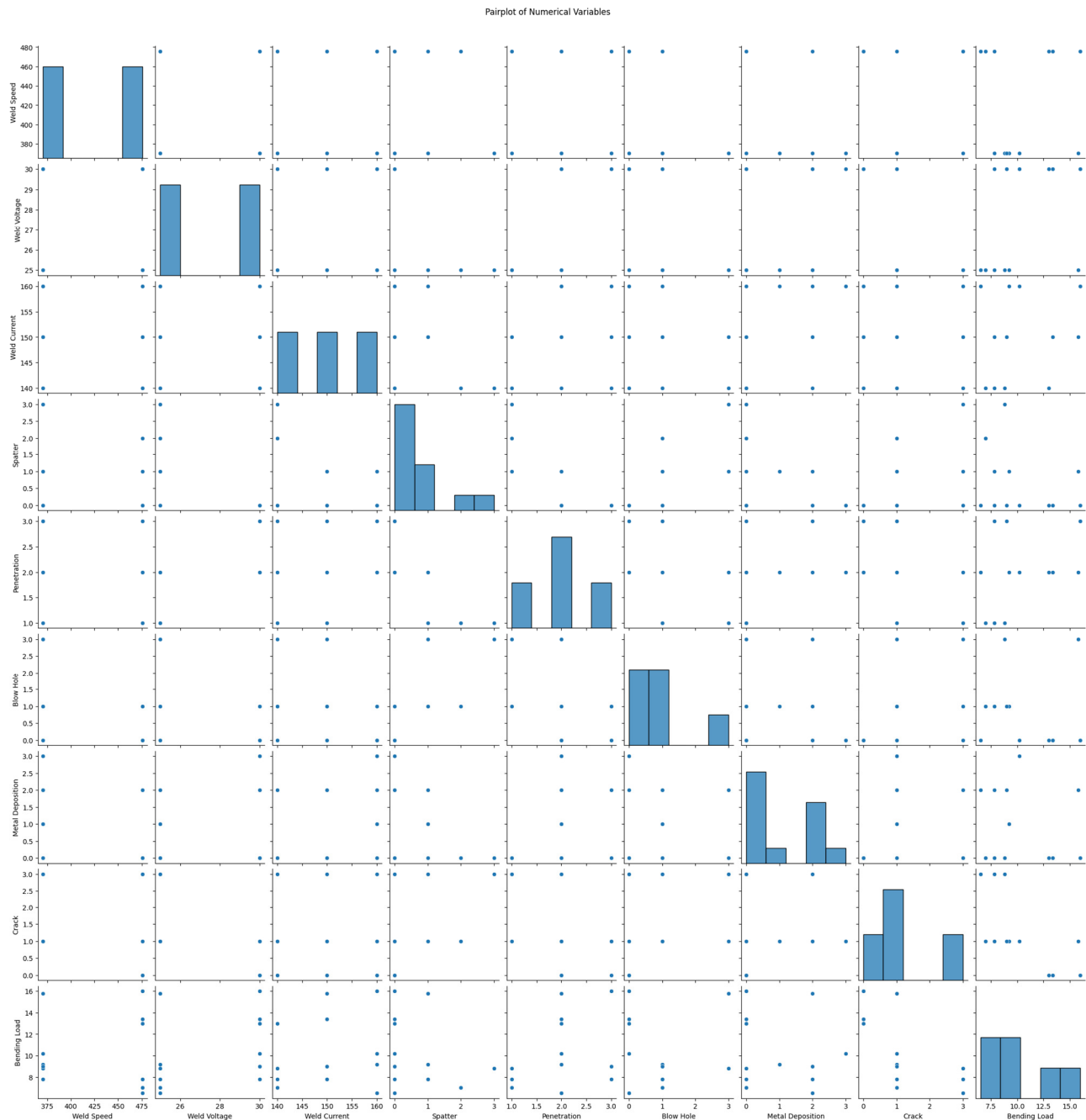


Figure-1: Pair plots of numerical data

Now for better understanding of the statistical significance of the variables, the distribution of variables is shown in figure-2 to figure-7.

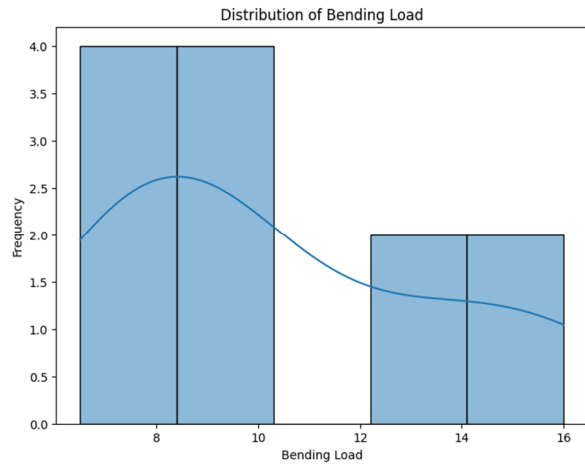


Figure-2: Distribution of bending load

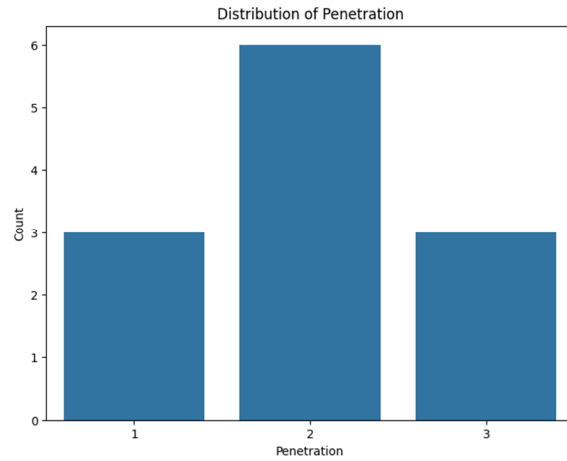


Figure-3: Distribution of penetration

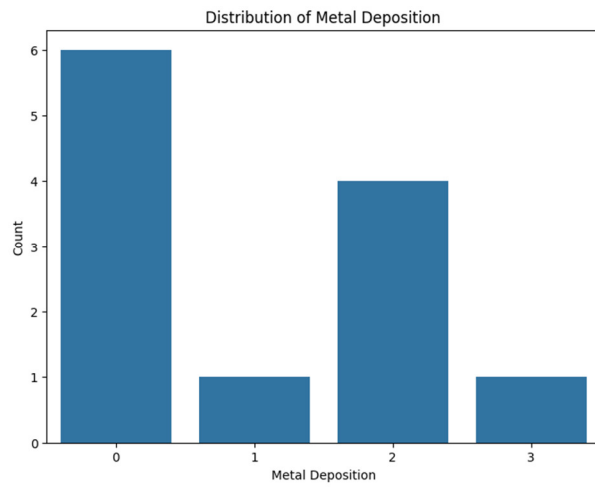


Figure-4: Distribution of Deposition

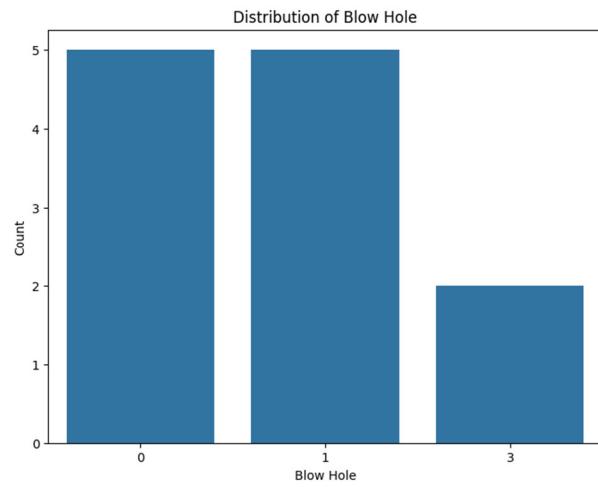


Figure-5: Distribution of Blow Hole

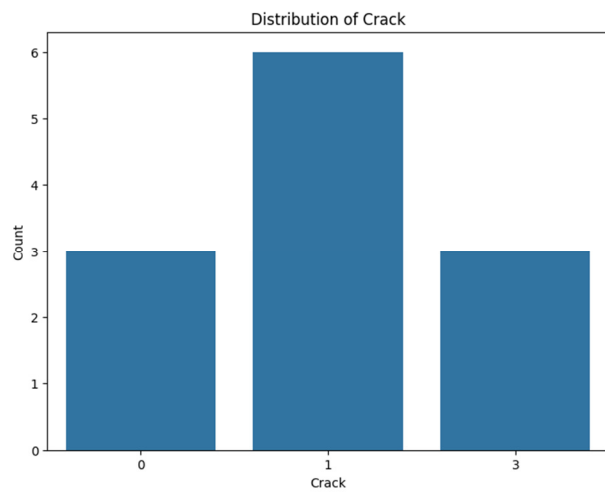


Figure-6: Distribution of Crack

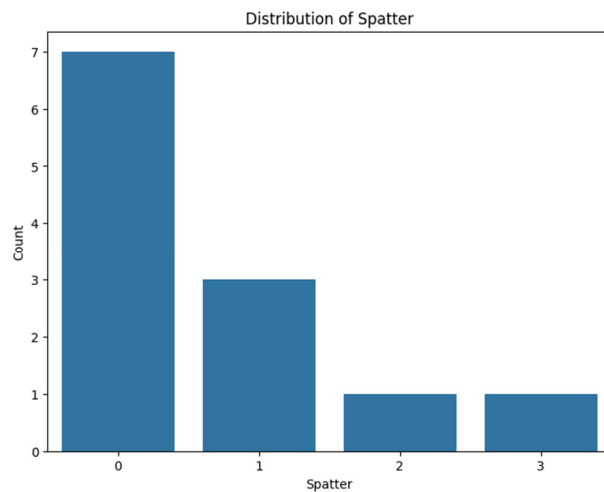


Figure-7: Distribution of Spatter

For further analysis and getting direction correlation among the variables are required to be derived. The correlation matrix is presented in figure-8.

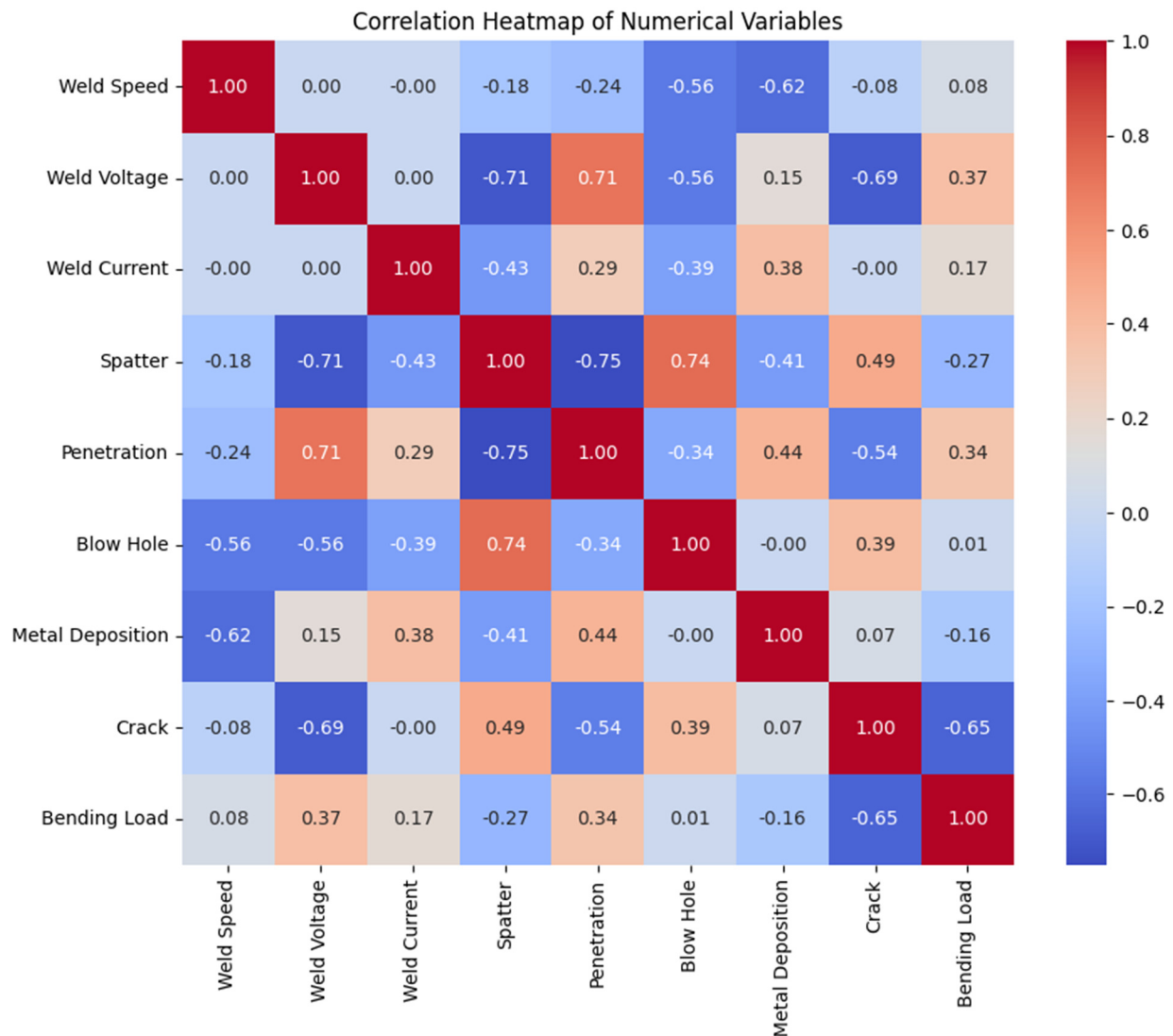


Figure-8: Correlation matrix

From this correlation matrix it can be concluded that Penetration and welding voltage, whereas welding speed contributes less. There is high correlation between welding voltage and spatter and blow hole. As a consequence spatter and blow hole are also correlated, that is if spatter is seen on the surface of the weld metals, it can be concluded that there is higher probability of blow hole occurring inside the weldment. There is also moderate correlation between welding current and spatter and blow hole. There is substantial correlation between welding voltage and crack. Metal deposition is to some extent is correlated to welding current.

The most important factor of the welding is the welding strength which was determined by the bending load. To draw the significance of different input features on the weld strength the visualization is shown in figures -9 to 11.

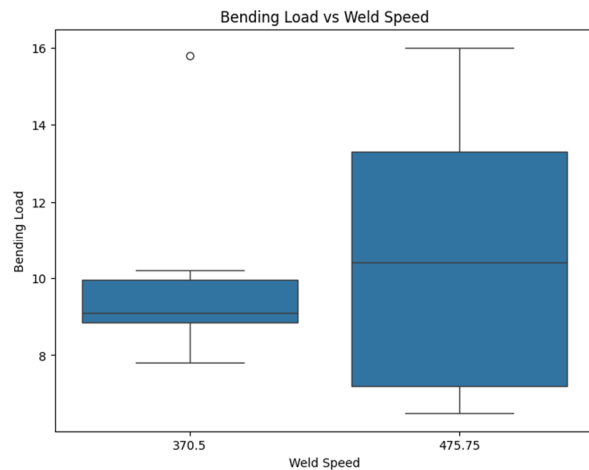


Figure-9: Bending load vs weld speed voltage

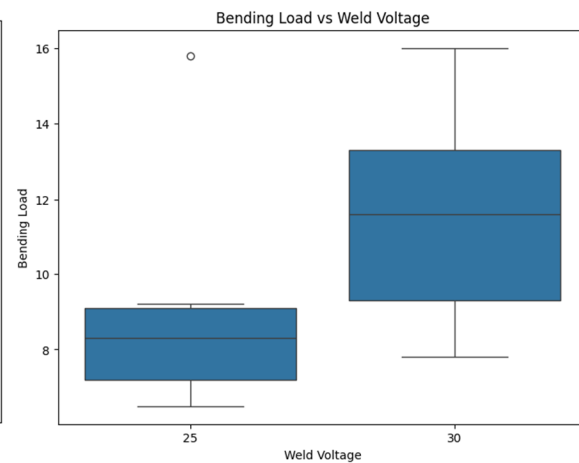


Figure-10: Bending load vs weld

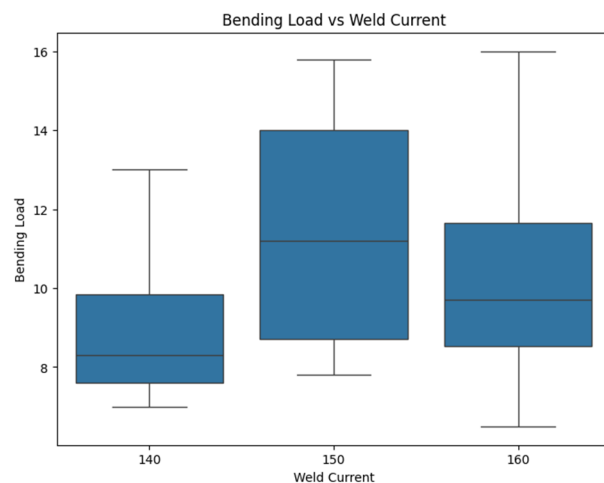


Figure- 11: Bending load vs welding current

From these plots the spread of bending load can be visualized for different parametric settings.

Now for optimization the output variables were categorized as positive attributes and negative attributes depending on whether we want to maximize or minimize. Bending load, penetration and deposition were taken as positive attributes, and blow hole, crack and spatter were considered as negative attributes. The processed experimental data was modified accordingly by incorporating negativities in the columns of negative attributes. Then outputs were scaled for a meaningful composite score. Here a composite score is being defined as sum of scaled maximized outputs - sum of scaled minimized outputs. Bending Load and Penetration are generally maximized, but there might be tradeoffs. Metal Deposition being 'Desirable' is also

positive. For simplicity, all 3 were considered as to be maximized (after appropriate scaling/mapping) and spatter, blow hole, crack as to be minimized. The random forest regressor was trained and then a grid search algorithm was run. The model generated a mean squared error (MSE) as: 2.803 which is quite low. The grid search was run with the aim of highest composite score and the result is tabulated in table-4.

Table-4: parametric setting and output with composite score.

index	Weld Speed	Weld Voltage	Weld Current	Spatter	Penetration	Blow Hole	Metal Deposition	Crack	Bending Load	Composite Score
11	475.75	30	160	0	3	0	0	0	16.0	4.986208022167752
5	370.5	30	160	0	2	0	3	1	10.2	3.6169456878143125
4	370.5	30	150	0	3	1	2	1	9.0	2.765990548321702
10	475.75	30	150	0	2	0	0	0	13.4	2.75844274373932
9	475.75	30	140	0	2	0	0	0	13.0	2.6332809412692684

Schematically the distribution composite score on original data is shown in figure-12.

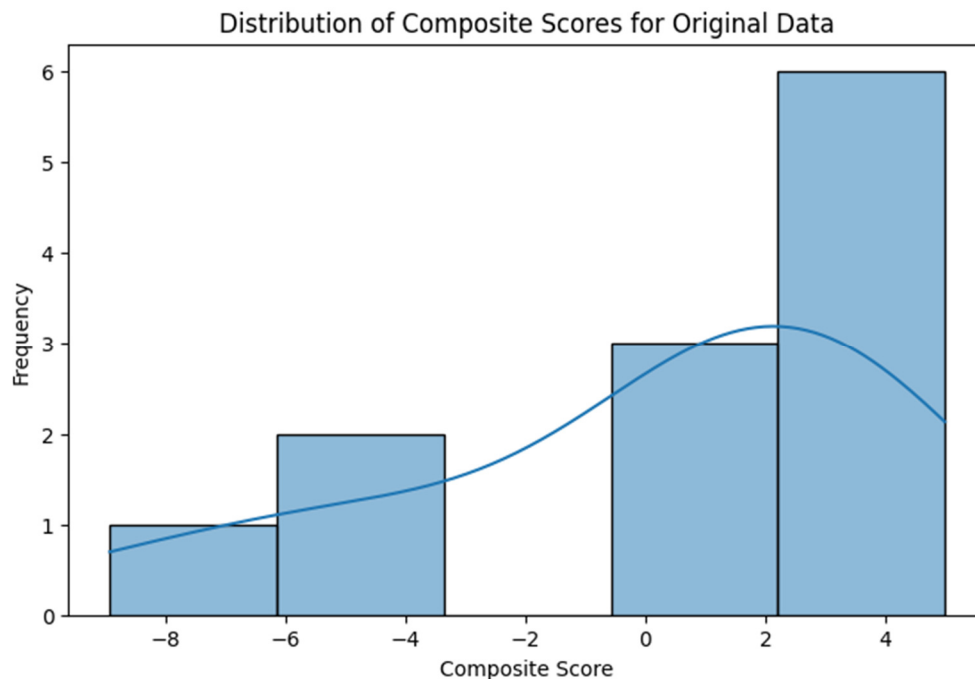


Figure-12: Distribution of composite score for original data.

Experiment number 11 proved to be best and it is also validated from the experimental point of view intuitively. However, that may vary with user choice of desirability of outputs. To overcome this there was a generation of new data set by the machine learning algorithm of RFR and an iterative search was performed. After 1000 iterations the model came up with the following optimized result.

Best composite score found: 0.07954667810877378

Optimal parameters suggested by search:

Weld Speed 406.610170

Weld Voltage 29.386625

Weld Current 143.921477

Predicted outputs for the suggested optimal parameters:

Spatter	Blow Hole	Crack	Bending Load	Penetration	Metal Deposition
0.600741	0.874074	1.382963	9.677111	2.054074	1.251111

The above set of parameters and output variables were generated by the algorithm and this is to be validated through experimentation.

Additionally Pareto front analysis was performed as the problem is of the kind of multi objective decision making. With the absence of experimental validation this is performed and compared for performance of the algorithm on the data.

Table -5 indicates the result obtained by Pareto front.

Table-5: Pareto Front Data (Optimal trade-offs observed in the data):

index	Weld Speed	Weld Voltage	Weld Current	Spatter	Penetration	Blow Hole	Metal Deposition	Crack	Bending Load
1	370.5	25	150	1	2	3	2	1	15.8
4	370.5	30	150	0	3	1	2	1	9.0
5	370.5	30	160	0	2	0	3	1	10.2
11	475.75	30	160	0	3	0	0	0	16.0

It is evident that the result obtained from Pareto front is well correlated to that of RFR.

It was observed in the correlation matrix that with bending load crack and spatter was correlated and was conflicting attribute. Pareto front of bending load with spatter and crack is shown in figure 13 and 14.

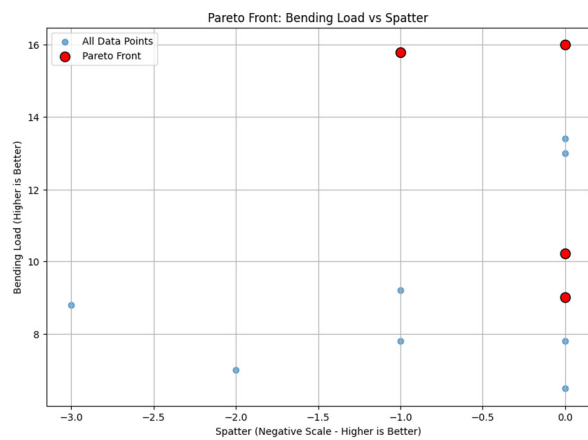


Figure-13: Pareto front-bending load vs spatter crack

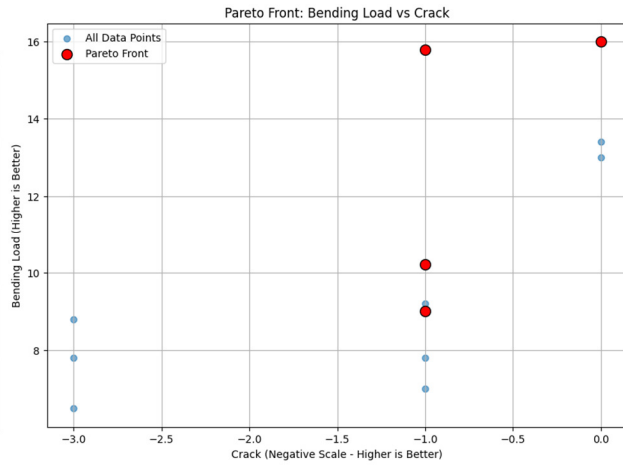


Figure-14: Pareto front-bending load vs crack

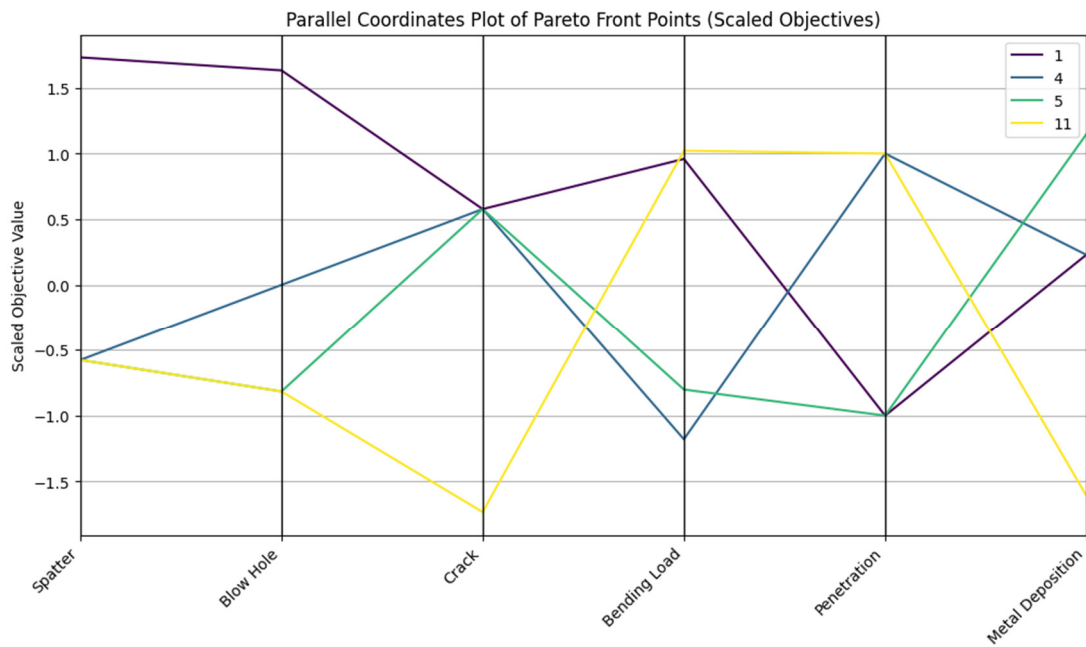


Figure-15: Parallel coordinates plot of Pareto Front points

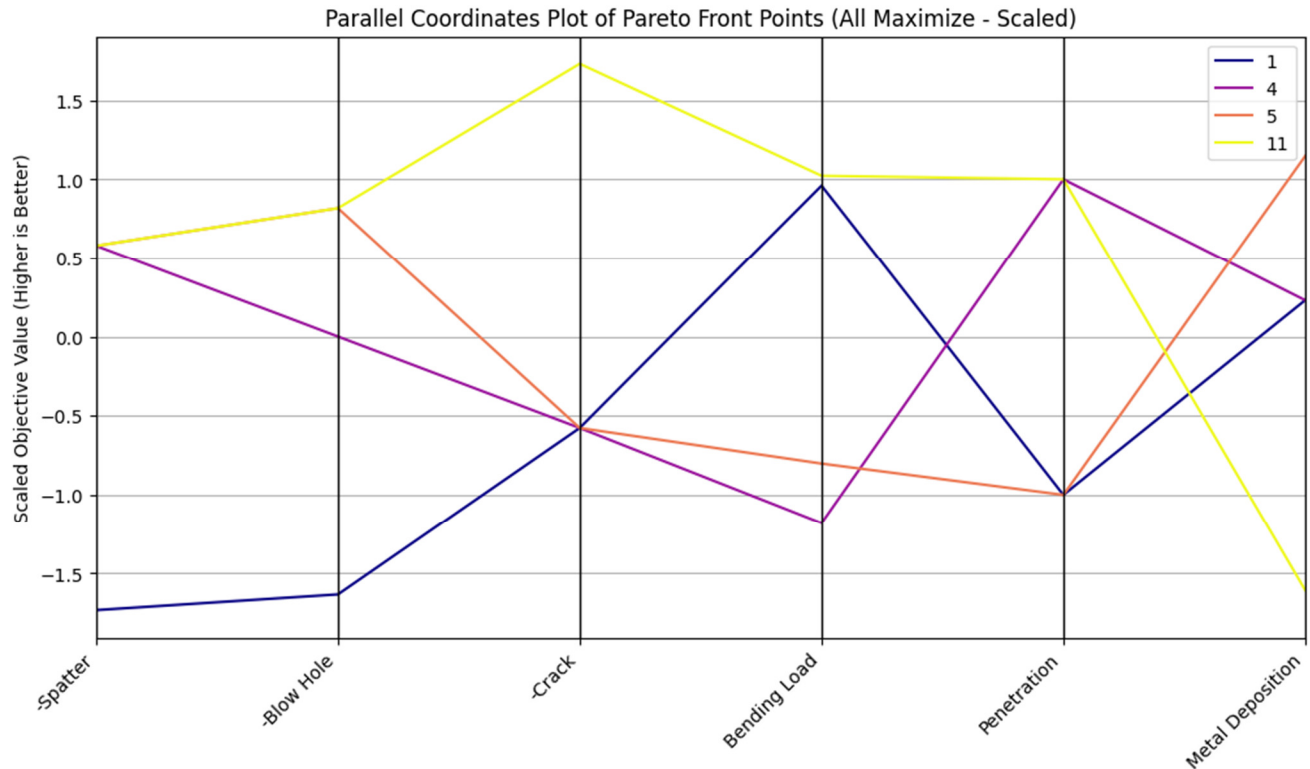


Figure-16: Parallel coordinates plot of Pareto front points (all maximize)

The plot corresponding to figure16 significantly expresses the performance of various Pareto fronts conforming to the solutions 1,4,5 and 11 across all the scaled objectives. All the objectives were scaled and normalized in such a manner that higher value indicates better. The best solution is indicated by the yellow line, that is, solution11. And this is exactly the solution provided by the random forest regression. So we can conclude that our machine learning model performed well and is matching with Pareto front analysis. So we may consider the optimized solution provided by RFR.

5. Conclusion

The random forest regression model which is a kind of ensemble machine learning model performed well in case of MIG welding experimental data. Here the number of data points was very low. Nonetheless the algorithm worked well and can be implemented in industrial scenario, where having multitudes of data the ensemble machine learning regression model is expected to perform very well. In the present case the best solution among the experimental data experiment no 11 proved to be best and that was supported by Pareto front analysis. If we look closely the output variables associated with this experiment, we find that it performed the best in blow hole and crack resistance. It had also high performance with respect to bending load and spatter. However with regard to metal deposition it did not perform well, which indicates a tradeoff. To

overcome this it was desired to search for optimal solution, which may consists of parametric combination out of the experimentally considered ones. And the grid search of RFR finally arrived at an optimal solution considering all the output variables to be satisfied optimally. Bearing in mind the success of the algorithm the optimal solution may be considered. However there is scope for experimental validation.

References

1. Jayant, Arvind, and M. S. Dhillon. "Use of analytic hierarchy process (AHP) to select welding process in high pressure vessel manufacturing environment." *International Journal of Applied Engineering Research* 10, no. 8 (2015): 586-595.
2. Jeet, Siddharth, Abhishek Barua, Biswajit Parida, Bibhuti Bhusan Sahoo, and Dilip Kumar Bagal. "Multi-objective optimization of welding parameters in GMAW for stainless steel and low carbon steel using hybrid RSM-TOPSIS-GA-SA approach." *Int J Tech Innov Mod Eng Sci* 4 (2018): 683-692.
3. Bhattacharya, Subhajit, Kazi Sabiruddin, and Santanu Das. "Optimal selection of metal active gas welding parameters in joining high carbon steel: though the AHP." *Indian Science Cruiser* (2021): 27-36.
4. Zhu, Caixia, Haitao Yuan, and Guohong Ma. "An active visual monitoring method for GMAW weld surface defects based on random forest model." *Materials Research Express* 9, no. 3 (2022): 036503.
5. Munghate, Abhinav Arun, and Shivraman Thapliyal. "Process parameters based machine learning model for bead profile prediction in activated TIG Welding using random forest machine learning." *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 238, no. 12 (2024): 1761-1768.
6. Babaiyan, Vahide, Nader Mollayi, Morteza Taheri, and Majid Azargoman. "An Ensemble Random Forest Model to Predict Bead Geometry in GMAW Process." *Journal of Advanced Manufacturing Systems* 21, no. 02 (2022): 393-410.
7. Rajesh, S., Gobikrishnan Udhayakumar, M. Bhuvaneswari, R. Arun Chakravarthy, M. Arun, and C. Jayabalan. "Prediction of ultimate tensile strength of dissimilar substances welding using random forest regression." *In AIP Conference Proceedings*, vol. 3204, no. 1. AIP Publishing, 2025.
8. Nampoothiri, Jayakrishnan. "Leveraging Machine Learning to Predict Welding Quality in Ultrasonically Assisted TIG Welded Inconel 625 Joints." *Journal of Materials Engineering and Performance* (2025): 1-15.
9. Zhang, Zhifen, Zhe Yang, Wenjing Ren, and Guangrui Wen. "Random forest-based real-time defect detection of Al alloy in robotic arc welding using optical spectrum." *Journal of Manufacturing Processes* 42 (2019): 51-59.
10. Han, Junfeng, Zhiqiang Feng, Ziquan Jiao, and Xiangxi Han. "The research of real-time welding quality detection via visual sensor for MIG welding process." *In Transactions on Intelligent Welding Manufacturing: Volume III No. 4 2019*, pp. 75-86. Springer Singapore, 2021.

11. Chandra, Mukesh, Shatrudhan Kumar, Kumar Ankit, Sonu Kumar Rajak, and Sonu Rajak. "A Machine learning approach for prediction of surface temperature of the weld region in A-TIG welding." *Transactions of the Indian Institute of Metals* 77, no. 3 (2024): 907-917.
12. Kumar, Kaushal, Puja Acharya, Subhav Singh, Deekshant Varshney, Umank Mishra, Prawar, Rishabh Arora, Gulab Singh Chauhan, and Annap Narain Singh. "Analyse the performance characteristics of mild steel plates at varying weld parameters by using artificial intelligence approaches." *Welding International* (2025): 1-12.
13. Mezher, Marwan T., Alejandro Pereira, Rusul Ahmed Shakir, and Tomasz Trzepieciński. "Application of machine learning and neural network models based on experimental evaluation of dissimilar resistance spot-welded joints between grade 2 titanium alloy and AISI 304 stainless steel." *Heliyon* 10, no. 24 (2024).
14. Zhang, Zhifen, Yiming Huang, Rui Qin, Wenjing Ren, and Guangrui Wen. "XGBoost-based on-line prediction of seam tensile strength for Al-Li alloy in laser welding: Experiment study and modelling." *Journal of Manufacturing Processes* 64 (2021): 30-44.
15. Rameshkumar, K., A. Vignesh, P. Gokula Chandran, V. Kirubakaran, J. Sankaran, and A. Sumesh. "Machine learning models for weld quality monitoring in shielded metal arc welding process using arc signature features." *SAE International Journal of Materials and Manufacturing* 15, no. 4 (2022): 347-366.
16. Wang, Xuewu, Qian Chen, Hao Sun, Xiuwei Wang, and Huaicheng Yan. "GMAW welding procedure expert system based on machine learning." *Intelligence & Robotics* 3, no. 1 (2023): 56-75.
17. Grujicic, M., S. Ramaswami, J. S. Snipes, R. Yavari, C-F. Yen, and B. A. Cheeseman. "Optimization of gas metal arc welding (GMAW) process for maximum ballistic limit in MIL A46100 steel welded all-metal armor." *Journal of Materials Engineering and Performance* 24 (2015): 229-244.
18. Zhao, Yun-tao, Wei-gang Li, and Ao Liu. "Optimization of geometry quality model for wire and arc additive manufacture based on adaptive multi-objective grey wolf algorithm." *Soft Computing* 24, no. 22 (2020): 17401-17416.
19. Thakar, Hemenkumar H., Mrunalkumar D. Chaudhari, Jay J. Vora, Vivek Patel, Subhash Das, Din Bandhu, Manish Gupta, and V. Suryaprakash Reddy. "Performance optimization and investigation of metal-cored filler wires for high-strength steel during gas metal arc welding." *High Temperature Materials and Processes* 42, no. 1 (2023): 20220305.
20. Rocha, Luiz CS, Anderson P. Paiva, Emerson J. Paiva, and Pedro P. Balestrassi. "Comparing DEA and principal component analysis in the multiobjective optimization of P-GMAW process." *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 38 (2016): 2513-2526.