INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 16-10-2023

Accepted: 28-10-2023

Published: 26-11-2023

Citation: Kumar D, Kumari P, Singh MV, Devi S, Kumar Y (2023) Prediction of Material Surface Area using Multiple Linear Regression Algorithm with Independent Variables Time, T emperature and Quantity. Indian Journal of Science and Technology 16(44): 4048-4053. https://doi.org/10.17485/IJST/v16i44.2624

* **Corresponding author**.

yk.physics@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN

Print: 0974-6846

Electronic: 0974-5645

Prediction of Material Surface Area using Multiple Linear Regression Algorithm with Independent Variables Time, T emperature and Quantity

Dinesh Kumar[1], Poonam Kumari[1], Man Vir Singh[2], Sarita Devi[3], Yogesh Kumar[4]*

1 Department of Mathematics, Singhania University, Jhunjhunu, 333515, Rajasthan, India
2 Department of Chemistry, Dev Bhoomi University, Uttarakhand, 248007, India
3 Department of Chemistry, Govt. College Sec-12, Palwal, 121102, Haryana, India
4 Department of Physics, Govt. College Sec-12, Palwal, 121102, Haryana, India

Abstract

Objective: In this paper, the surface area of MnO_2 was predicted by multiple linear regression algorithm (MLRA) in python Jupyter notebook with accuracy and best correlation between parameters. **Methods:** A data set was collected from different recent research papers. The dataset underwent data processing and factor analysis, involving the removal of unnecessary data. The surface area was predicted by the different experimentally synthesis parameters (Temperature, Material quantity, reaction time and experimental evaluated surface area). For prediction of surface area the total data set of 11 different research pater data was divided in two part as 72% for training dataset and 28% kept for test dataset. **Findings:** The surface area of the material can be enhanced from 92 to 196.67 m^2/g by changing the synthesis parameters during hydrothermal process. For prediction, the best relationship was found between temperature and surface area in heatmap. The predictions yielded accuracy levels was 94%. **Novelty:** The surface area of MnO_2 was first time predicted by MLRA with high accuracy by tuning the hydrothermally synthesis parameters.

Keywords: Multiple Linear Regression Algorithm (MLRA); Machine Learning; Python; Heatmap; Prediction Of Surface Area

1 Introduction

The surface area of a material is a critical factor influencing performance and effectiveness in divers field[1]. It plays a role in the adsorption of pollutants and contaminants, contributing to environmental remediation efforts[2]. In the field of electronics, materials with optimized surface properties are essential for device performance and functionality[3,4]. The surface area can be enhanced by tuning the hydrothermally synthesis parameter of material[5]. It is not easy to experimentally synthesis of material by changing parameters[6,7]. Hence, the machine learning approach can be used for find the best parameter for higher surface area of the material[8]. The MLRA Jupyter Notebook gives the correlation between parameters (independent variables)[9]. It serves as a comprehensive package encompassing various data structures and general algorithm[10].

Understanding and manipulating surface area characteristics are fundamental in advancing materials science and technology across various industries. The multiple linear regression model was used to load forecasting by changing parameters like heat, last day burden, and weekly burden etc and the prediction accuracy levels was 95%[11]. Another study focused on predicting coconut palm production through the application of multiple linear regression. In this model, the dependent variable (y) represents coconut palm production, while independent variables (x) include month, rainfall, land width, tree count, bunch quantity, and average weight. External data from the Meteorology, and Climatology Agency were incorporated. The training statistics constituted 80%, and the test statistics comprised 20% of the training set[12–14]. The mean absolute % error was 14.28%[15]. Another studies were focused on predicting the power output of solar by multiple linear regression. The predictive equations for power output were explained by Nurazizah & Winarno. Correlation coefficient for the dataset in Factory 1 was 0.52, indicating that external weather and conditions influenced 52% of the power generated. In Factory 2, the correlation coefficient was 0.92, signifying a 92% impact of external factors on power generation[16]. Abobaker et al., predict the rainfall by developing MLRA for Khartoum state, Sudan and concluded that the average mean square error changed when training and testing ratio change[17]. The Precipitation and Visibility were predicted by MLRA to evaluate the accuracy of the models by a large weather data with 4018 samples and concluded that it can provide more accurate predictions[18]. Hence MLRA was used to predict the surface area of MnO_2 from the experimental data.

In this study, data collection was accomplished from different experimentally evaluated value from research articles. Surface area of material was dependent on the physical parameters like temperature, time, and synthesis material quantity of potassium per magnate. Jupyter Notebook was used from an open-source web application enabling the creation and sharing of documents integrating live code, equations, visualizations, and narrative text.

2 Methodology

The prediction of surface area of material (MnO_2) evaluated by the steps as shown in Figure 1. In this research Anaconda Jupyter notebook in python programming language was used to predict the surface area[19]. Identify variables as either free or bound. Collect necessary data, grouping variables accordingly. For instance, free variables like $KMnO_4(x_1)$, time (x_2), and temperature (x_3), and bound variable data such as harvest results (Y). For prediction of surface area, data was collected from experimentally evaluated results from recent reputed articles with parameter temperature, time, material quantity used during hydrothermally synthesis of the MnO_2 and surface area evaluated by Brunauer-Emmett-Teller (BET)[20]. After completing the data collection stages, over 10 papers contributed datasets specifically focused on material MnO_2. The dataset was utilized for constructing the prediction model comprises four key components: surface area, temperature, time, and the concentration of $KMnO_4$ during the reaction stage. Subsequently, this curated dataset underwent training using the multiple linear regression method. In this method, temperature, time, and $KMnO_4$ act as independent variables or predictors, while the dependent variable or target was the resulting surface area. as shown in Table 1.

Table 1. Data collection for prediction of Surface area

Sr. No.	Nano Material	Temp in c^0	Surface Area in m^2 /g	Time in hour	$KmNO_4$ in gram	Ref.
1	MnO_2	80	78.56	12	0.2	[21]
2	MnO_2	140	33.54	12	0.237	[22]
3	MnO_2	180	226	48	1.264	[23]
4	MnO_2	200	160.4	48	0.1	[24]
5	MnO_2	160	198	12	1.5	[25]
6	MnO_2	140	29.2	3	0.948	[26]
7	MnO_2	140	150	12	0.5	[27]
8	MnO_2	110	108.6	6	1.262	[28]
9	MnO_2	110	86.7	3	1.262	[28]
10	MnO_2	110	92.8	12	1.262	[28]
11	MnO_2	110	52	24	1.262	[28]

Using multiple linear regression models, the forecast of the surface area of MnO_2 at the destination in order to increase surface area. Two different types of variables were used in this training. Second variable bound (Y) and first independent (X). The dependent variable was results prediction Y, while the independent variables were time (2), temperature (3), and the amount of $KMnO_4$ (1). Converted raw boolean data types into numeric or integer forms for machine learning compatibility. Steps include: A. Change Data Type : Converted data types for machine learning system readability. B. Normalization: Ensure variables were the same value range. C. Divided data into training and testing sets with a ratio of generally 80:20 (train-test split). The effect of all the selected attributes along with temperature, time and quantity of $KmNO_4$ were taken into as input variables, and surface area as the output variable were chosen to train and test the model. Data was modelled for the training dataset using multiple linear regression and predictions were made for the test data.

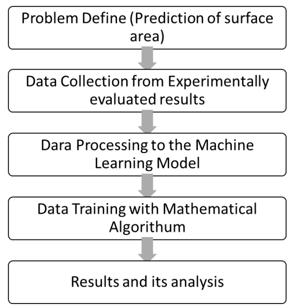

Fig 1. Steps for prediction of surface area

The data was used to make this prediction, thus the level of accuracy is still insufficient. The analysis uses Python-based Jupyter notebook implementations of built-in machine learning models from the sci-kit learn library. Linear regressive, as seen in the example below, simple linear regression created a linear equation or relationship between the input (x) and the output (y):

$y = a + b.x + e$

In this case, x standed for the input variable, y for the output variable, 'a' for the y-intercept, 'b' for the slope, and 'e' for the error [29]. The linear relationship between two or more input variables and a single output was using multiple linear regression (MLR) result in the expression

$Y = a_1.x_1 + a_2.x_2 \ldots\ldots\ldots a_n.x_n + e$

Where a_1, a_2 , and so on were the variable coefficients of the factors x_1, x_2, and so forth. The non-linear relationship between the input variables and the conditional mean y, denoted as E(y/x), was captured by polynomial regression . In general, the predicted value of y could be represented as an nth degree polynomial, which results in the general polynomial regression model shown in the equation below:

$Y = A_0 + A1X_1{}^1 + A2X^2{}_2 \ldots\ldots AnX^n{}_n \ldots\ldots + E$

where A1, A2,.... were the coefficients of X_1, X_2 ... correspondingly [30]

Performance measure

For measuring the performance of the model of multiple linear regression following formula were used

$$MAE = \frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i|}{N}$$

$$MSE = \frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}{N}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)}{N}}$$

R -Squared $R^2 = 1$ -

$$\frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N} (y_i -)^2}$$

where $(\widehat{y}_i)$ were the predicted output, (yi) were the actual output value, y was the mean of the actual values, and N was the number of data points.

The next stage entails were starting the training procedure to create a predictive model after the dataset collected. By using independent and explanatory factors, regression acted as a method for predicting future data. In contrast to predictive classification, which dealt with discrete score variables, regression focused on mapping the function of learning data elements to predict continuous score variables, offering valuable insighted into real-world predictions. This systematic approach guided the analysis, ensuring a thorough understanding of relationships and predictive capabilities within the data.

3 Results and Discussion

The surface area of MnO_2 was predicted by a data set taken from 11 different research papers that was divided into two part as 72% for training dataset and 28% kept for test dataset. The MLRA was trained by data set and the outcome of the training was represented by equation

$$y = -12.5730 + 0.7369x1 + 1.1951 \times 2 + 3.4753 \times 3 \tag{1}$$

Where y was the surface area, x1, x2 and x3 are $KmNO_4$ quantity, time and temperature. Equation 1 depicts the relation between surface area and synthesis parameters which conclude that the surface area is highly dependent on the temperature as compared to the other parameters. The goodness of the fit of the model was analysed by R-squared score and found 4.1891. The accuracy of the model was obtained by R2 score and it was 94.95%. While Nurazizah & Winarno (2022) predict the catfish yield to fulfil community needs using multiple linear regression algorithm method the accuracy of the model was 75%[(31)]. Dabhade et, al., (2021) predicted students' academic performance using machine learning algorithms and the goodness of fit obtained using MLR for the available dataset with R2 score was 83.27%[(32)].

In Figure 2 , the heat map revealed the insightful information. The visualization of the variable in heatmap was illustrating the correlation between value of x1, x2, x3 and y. The heat map provided a clear and informative visualization, allowing for a nuanced understanding of the relationships within the variables under consideration. The value of heat map correlation varies from -1 to 1[(33)]. When the value of boxes approached to 1, it indicates a highly robust relationship between the variables. The correlation of the temperature and time with surface area were 0.62 and 0.59 respectively. The value 0.37 on the plot were shown a corelation between quantity of $KMnO_4$ and surface area, which conclude that the surface area was less dependent on the material quantity. This implies that both temperature and time play an important role in tuning the surface area[(34)].

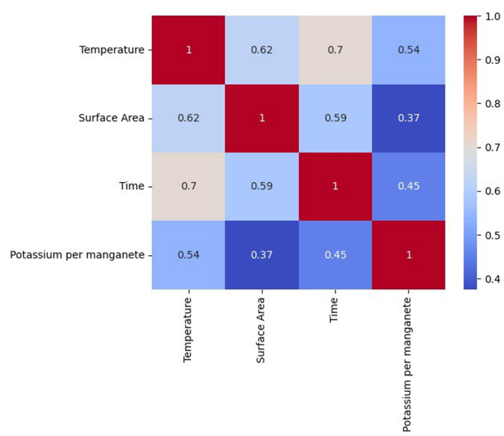

Fig 2. Correlation plot of variables

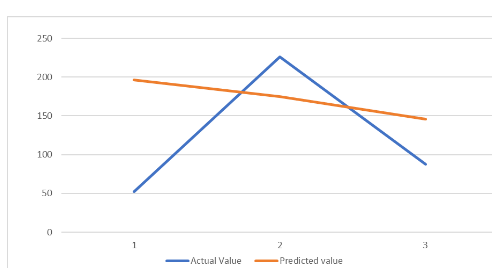

Fig 3. Chart comparing actual and predicted data

Figure 2 gives the comparison between the original data and the predicted data. Which shows a huge gape between predicted and actual surface area. By tuning the synthesis parameters, the surface area of the material could be enhanced from 92 to 196.67 m^2/g. As the data size increases, there might be the possibility to obtain better accuracy in the result.

4 Conclusion

In the present study, the surface area of MnO_2 was first time predicted by MLRA Jupyter notebook with high accuracy of 94% by tuning the hydrothermally synthesis parameters. The correlation between the surface area and temperature was highest and lowest was surface area and material quantity. The surface area of the MnO_2 can be enhanced from 92 to 196.67 m^2/g by tuning the synthesis parameters. Hence, the MLRA may be used for optimising the surface area of the material and high surface area MnO_2 can be synthesised experimentally.

Acknowledgement

The author Dinesh Kumar would like to acknowledge the financial support from the University Grant Commission, New Delhi, India for providing the Junior Research Fellowship under the scheme of NET JRF.

References

1) Ulusoy U. A Review of Particle Shape Effects on Material Properties for Various Engineering Applications: From Macro to Nanoscale. *Minerals*. 2023;13(1):91. Available from: https://doi.org/10.3390/min13010091.
2) Rai PK. Novel adsorbents in remediation of hazardous environmental pollutants: Progress, selectivity, and sustainability prospects. *Cleaner Materials*. 2022;3:100054. Available from: https://doi.org/10.1016/j.clema.2022.100054.
3) Gupta MK, Kumar Y, Sonnathi N, Sharma SK. Synthesis of MnO2 nanostructure and its electrochemical studies with ratio optimization of ZnO. *Ionics*. 2023;29(7):2959–2968. Available from: https://doi.org/10.1007/s11581-023-04998-w.
4) Komal, Kumar A, Kumar Y, Shukla VK. Parthenium hysterophorus derived activated carbon for EDLC device application. *Journal of Materials Science: Materials in Electronics*. 2023;34(27):1880. Available from: https://doi.org/10.1007/s10854-023-11309-6.
5) Rani S, Bansal L, Tanwar M, Bhatia R, Kumar R, Sameera I. Role of precursor concentration in tuning the electrochemical performance of MoS2 nanoflowers. *Materials Science and Engineering*. 2023;292:116436. Available from: https://doi.org/10.1016/j.mseb.2023.116436.
6) Morey GW. Hydrothermal Synthesis. *Journal of the American Ceramic Society*. 1953;36(9):279–285. Available from: https://doi.org/10.1111/j.1151-2916.1953.tb12883.x.
7) Kumar Y, Uke SJ, Kumar A, Merdikar SP, Gupta M, Thakur AK, et al. Triethanolamine–ethoxylate (TEA-EO) assisted hydrothermal synthesis of hierarchical β-MnO₂ nanorods: effect of surface morphology on capacitive performance. *Nano Express*. 2021;2(4):040008. Available from: https://doi.org/10.1088/2632-959X/abef21.
8) Selvaratnam B, Koodali RT. Machine learning in experimental materials chemistry. *Catalysis Today*. 2021;371:77–84. Available from: https://doi.org/10.1016/j.cattod.2020.07.074.
9) Hunter-Zinck H, De Siqueira AF, Vásquez VN, Barnes R, Martinez CC. Ten simple rules on writing clean and reliable open-source scientific software. *PLOS Computational Biology*. 2021;17(11):e1009481. Available from: https://doi.org/10.1371/journal.pcbi.1009481.
10) Raschka S, Patterson J, Nolet C. Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information*. 2020;11(4):193. Available from: https://doi.org/10.3390/info11040193.
11) Dhaval B, Deshpande A. Short-term load forecasting with using multiple linear regression. *International Journal of Electrical and Computer Engineering (IJECE)*. 2020;10(4):3911. Available from: https://doi.org/10.3390/info11040193.
12) Rashid M, Bari BS, Yusup Y, Kamaruddin MA, Khan N. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. *IEEE Access*. 2021;9:63406–63439. Available from: https://doi.org/10.1109/ACCESS.2021.3075159.
13) Ibañez SC, Monterola CP. A Global Forecasting Approach to Large-Scale Crop Production Prediction with Time Series Transformers. *Agriculture*. 2023;13(9):1855. Available from: https://doi.org/10.3390/agriculture13091855.
14) Vishweshwar S, Meti S, Champa BV, Nagaraja M. Climate Based Coconut Yield Analysis in Chanrayapatna Taluk of Hassan District of Karnataka, India. *International Journal of Current Microbiology and Applied Sciences*. 2019;8(07):2867–2877. Available from: https://doi.org/10.20546/ijcmas.2019.807.357.
15) Prasetyo A, Salahuddin S, Amirullah A. Prediksi Produksi Kelapa Sawit Menggunakan Metode Regresi Linier Berganda. *Jurnal Infomedia*. 2021;6(2):76. Available from: https://doi.org/10.30811/jim.v6i2.2343.
16) Sahin G, Isik G, Van Sark WGJHM. Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression. *Energy Reports*. 2023;10:2837–2849. Available from: https://doi.org/10.1016/j.egyr.2023.09.097.
17) Ahmed HAY, Mohamed SWA. Rainfall Prediction using Multiple Linear Regressions Model. *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*. 2021;p. 1–5. Available from: https://doi.org/10.1109/ICCCEEE49695.2021.9429650.
18) Harun GS. Multiple Linear Regression Based Analysis of Weather Data for Precipitation and Visibility Prediction. In: Communications in Computer and Information Science. Springer Nature Switzerland. 2023;p. 60–71. Available from: https://doi.org/10.1007/978-3-031-37940-6_6.
19) Bloice MD, Holzinger A. A Tutorial on Machine Learning and Data Science Tools with Python. In: Lecture Notes in Computer Science. Springer International Publishing. 2016;p. 435–480. Available from: https://doi.org/10.1007/978-3-319-50478-0_22.
20) Uke SJ, Chaudhari GN, Bodade AB, Mardikar SP. Morphology dependant electrochemical performance of hydrothermally synthesized NiCo2O4 nanomorphs. *Materials Science for Energy Technologies*. 2020;3:289–298. Available from: https://doi.org/10.1016/j.mset.2019.11.004.
21) Kumar Y, Chopra S, Gupta A, Kumar Y, Uke SJ, Mardikar SP. Low temperature synthesis of MnO2 nanostructures for supercapacitor application. *Materials Science for Energy Technologies*. 2020;3:566–574. Available from: https://doi.org/10.1016/j.mset.2020.06.002.
22) Li L, Nan C, Lu J, Peng Q, Li Y. α-MnO2 nanotubes: high surface area and enhanced lithium battery properties. *Chemical Communications*. 2012;48(55):6945. Available from: https://doi.org/10.1039/C2CC32306K.
23) Zhao JG, Yin JZ, Yang SG. Hydrothermal synthesis and magnetic properties of α-MnO2 nanowires. *Materials Research Bulletin*. 2012;47(3):896–900. Available from: https://doi.org/10.1016/j.materresbull.2011.11.023.
24) Wang C, Bongard HJ, Weidenthaler C, Wu Y, Schüth F. Design and Application of a High-Surface-Area Mesoporous δ-MnO2 Electrocatalyst for Biomass Oxidative Valorization. *Chemistry of Materials*. 2022;34(7):3123–3132. Available from: https://doi.org/10.1021/acs.chemmater.1c04223.

25) Li D, Wu X, Chen Y. Synthesis of Hierarchical Hollow MnO2 Microspheres and Potential Application in Abatement of VOCs. *The Journal of Physical Chemistry* . 2013;117(21):11040–11046. Available from: https://doi.org/10.1021/jp312745n.

26) Subramanian V, Zhu H, Vajtai R, Ajayan PM, Wei B. Hydrothermal Synthesis and Pseudocapacitance Properties of MnO2 Nanostructures. *The Journal of Physical Chemistry B*. 2005;109(43):20207–20214. Available from: https://doi.org/10.1021/jp0543330.

27) Su D, Ahn HJ, Wang G. Hydrothermal synthesis of α-MnO2 and β-MnO2 nanorods as high capacity cathode materials for sodium ion batteries. *Journal of Materials Chemistry A*. 2013;1(15):4845. Available from: https://doi.org/10.1039/C3TA00031A.

28) Chen K, Noh YD, Li K, Komarneni S, Xue D. Microwave–Hydrothermal Crystallization of Polymorphic MnO2 for Electrochemical Energy Storage. *The Journal of Physical Chemistry C*. 2013;117(20):10770–10779. Available from: https://doi.org/10.1021/jp4018025.

29) Rong S, Bao-Wen Z. The research of regression model in machine learning field. *MATEC Web of Conferences*. 2018;176:01033. Available from: https://doi.org/10.1051/matecconf/201817601033.

30) Yağcı M. Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*. 2022;9(1):11. Available from: https://doi.org/10.1186/s40561-022-00192-z.

31) Nurazizah S, Winarno S. Prediction of Catfish Yield To Fulfill Community Needs Using Multiple Linear Regression Algorithm Method. *Devotion Journal of Community Service*. 2022;3(13):2146–2153. Available from: https://doi.org/10.36418/dev.v3i13.270.

32) Dabhade P, Agarwal R, Alameen KP, Fathima AT, Sridharan R, Gopakumar G. Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*. 2021;47:5260–5267. Available from: https://doi.org/10.1016/j.matpr.2021.05.646.

33) Srivastava AK, Safaei N, Khaki S, Lopez G, Zeng W, Ewert F, et al. Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Scientific Reports*. 2022;12(1):3215. Available from: https://doi.org/10.1038/s41598-022-06249-w.

34) Musić A, Telalović JH, Đulović D. The Influence of Stringency Measures and Socio-Economic Data on COVID-19 Outcomes. In: J HT, M K, editors. Communications in Computer and Information Science;vol. 2021. Springer International Publishing. 2021;p. 39–54. Available from: https://doi.org/10.1007/978-3-030-72805-2_3.