Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 6, pp. 1185–1198
DOI : 10.47974/JSMS-1230

Prediction of solar energy forecasting by using linear and logistic regression : A review and geographical comparative analysis in Indian context

Siddharth Wadehra *
Jagadeesh Varanasi †
Ambuj Anand §
Department Information Systems and Business Analytics
IIM Ranchi
Jharkhand
India

Abstract

Power generation from renewable resources is now-a-days recognized as an necessary basic skill to improve the operation of power system. Forecasting of the renewable energy is the important factor. This study on renewable energy forecasting Techniques gives the more knowledge about how the renewable energy is forecasted and how the renewable energy is converted to the electrical power. And it tells various techniques and methods for the forecasting of renewable energy resources. India is geographically diversified due to its different climatic conditions and we have collected data from solar energy sources from 6 different locations.

It comprises the data collected over a period of around one year solar plant output, solar irradiations and temperature of the PV plant. The datasets are considered for training to supervised learning algorithms with linear and logistic regressions by considering short term forecasting and predict immediate next day by using last 28 days data.

Subject Classification: *62Pxx : Application of Statistics.*

Keywords: *Geographical, Forecasting, Probability, Linear regression, Logistic regression, Sustainable development, Comparative analysis.*

This work is with respect to solar energy forecasting comparative analysis by using linear and logistic regression in Indian context.

* *E-mail:* **siddharth.wadehra20eph@iimranchi.ac.in; (Corresponding Author)**
wadehra.siddharth@gmail.com
† *E-mail:* **jagadeesh.varanasi20eph@iimranchi.ac.in;**
varanasi.jagadeesh@gmail.com
§ *E-mail:* **ambuj@iimranchi.ac.in**

1. Introduction

In this present scenario the necessity of affordable clean energy is increasing and the source for producing this energy is can be achieved from non-conventional resources. Because, these non-conventional sources are becoming extinct we can see a shift towards use of these resources for the generating the power.

Fossil fuel reserves are diminishing due to deforesting and rapid urban planning with taking any sustainable measures and this leads to environmental emissions and further cause's renewable energy, climate change, holds an growing share of the energy mix. Solar , wind, wave, and hydro energy are dependent on the extremely unpredictable weather patterns, so their increased penetration would result in major variations in the power pumped into the grid that needs to be regulated. For the smooth integration of large amounts of solar , wind, wave, and hydropower into the grid as well as the smooth integration of large amounts of solar , wind, wave, and hydropower into the grid, accurate, high-quality forecasts of renewable power generation are therefore important

There is an rising global demand for oil, but also for a reduction in environmental stress. Recently, large investments have been made in renewable energy research worldwide. The growing demand for environmentally-friendly energy makes the growth and production of renewable energy a sound investment for the future.

Solar power is the form of an energy derived from the sun that is non-conventional sources. The sun's energy is transferred to the energy's thermal form. Solar energy is spread over a broad geographical area and the cost of this energy is constant with low maintenance costs for the long-term in the future. A detailed study is therefore very important for specific model suitability for a specific location and climatic conditions of different time periods.

It is explaining about the logistic regression in machine learning and applying the statistical applications and its interpretation in machine learning [7].

Renewable energy datasets are available in this ninja for geographical locations parameters i.e weather, energy sources and climate [8].

It explains about the United Nations sustainable development framework for 17 objectives and clean energy and responsible production and consumption objectives are applicable for sustainable energy implementation [9][10].

This comprises of logistic regression machine learning steps and exercise [11].

In this article, the author explained about solving big scale non linear Continuous equations and proposed the hybrid conjugate methods that fulfil the descent property for global convergence [13, 15].

2. Literature

Environmental emission control is due to release of harmful gasses NOx into atmosphere becoming a major concern and an alternative source of energy generation has become feasible solution to fulfill the energy demand. Renewable energy (Solar and wind energy) is free and abundant in nature [1]. The utilization of renewable sources increasing rapidly to fulfill the global energy demand. Photo-voltaic (PV) for electricity generation is the fastest growing energy technology since 2002, experiencing an average annual increase of 48% [2].

The authors discussed power forecasting techniques for power plants works on non-conventional sources and based on statistical techniques for time series analysis and artificial neural networks and deep learning techniques for minimizing the total operational costs of the energy sources [3].

The authors mentioned about a day head accurate power forecasting based on the non-conventional energy data accumulated on the utility grids. Predicting alternative resource like solar and wind on their time horizons requires better prediction of weather prediction variables. It is difficult to extract the data from different relevant sources and different models with different time horizons [4]

This article describes the hierarchical model methodology and obtained the forecast for PV plant by using machine learning algorithms and compared with SVR and artificial neural network techniques and also calculated the different types of forecasting errors by using statistical methods [5].

This article is about forecasting for wind firms extend over a region and the authors mentioned about aggregate prediction method based on the search similarities among present wind speed predictions for set of locations by using its past historical data [6].

The data from a PV plant comprises of hourly data on solar power generation, temperature, and diffuse and direct irradiance over a long year[8]. These work renewable sources forecasting objectives are based on the sustainable development goals framework [9].

This paper is about predict the forecasting accuracy and probability for expected solar output generation by using the logistic regression machine learning algorithm with reference to Indian geographical locations PV data for short-term/ a day head forecasting [12].

3. Methodology and Experimentation:

We have used the logistic regression binary classification algorithm for predicting the short term forecast for 29th day by using previous four weeks data and assess and compare the forecast for six locations for all the seasons and presented the comparative and graphical analysis for all Indian six geographical locations with different latitude and climatic conditions

Logistic Regression: It is considered as one of the machine learning type supervised algorithm. It will take binary values either 0 or 1. It will be useful for binary classification problems and here it can be more or less applicable for power release function. If we keep the threshold limit value as 0.5 then if the probability exceeds 0.5 then the plant can release power based on the demand else it cannot release power. Linear regression does not have any decision boundary values but logistic regression lies between 0 and 1.

Logistic regression uses an equation as the representation, very much like linear regression Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). A key difference from linear regression is that the output value being expressed in binary values which takes 0 and 1 and lies between this interval rather than a numeric value and follows sigmoid curve [7].

We have used Origin 7.0, data analysis software, Origin Lab Pro (2017) for the graphical and data analysis purpose and Octave m interface of Matlab in order to determine the forecast accuracy.

Our methodology is based on applying all the methods feasible for the specified/particular location by using the past 28 days solar energy data from the PV plants of various Indian geographical locations and predicting the demand for 29th day for short term hourly basis forecast. We have applied the statistical modelling and determined the errors for the metrics below mentioned and the forecast assessment for the following and choose the supervised machine learning algorithm for the solar output prediction probability. This process can be also automated with highest accuracy for the specified location.

Most of the cases the existing literature discuss about the one particular method applied to different locations with different context by applying Artificial Neural Networks and provides some comparative analysis by using Machine Learning/Artificial Intelligence techniques and not specified exactly applying prediction of probability and applying all the methods at a time for different Indian geographical locations with North West East and South locations [14].

Initially PV plant data collected and cleaned manually and separated into specific data with labels for input and output variables. The above data was plotted to choose the variables the machine learning algorithm is the called logistic regression. The relationship between the variable power output-irradiance and other variables are carefully studied and calculated the gradient as per the slope of line separation.

Model is developed and fit with data accordingly after the data collection. Model trained and determined the forecast accuracy and finally test the optimized algorithm.

3.1 *Experimentation*

Work requirements:

In this work, solar PV plant data collected from the below mentioned website.

1. The following parameters considered for short term forecasting analysis for solar power such as PV plant hourly data comprising irradiance, plant power output, and local temperature and other weather conditions. The solar power PV plant data is collected from the below given website which provides information from satellite [8].

The above required data is collected from the following website: *https://www.renewables.ninja* **[8]**

In this case the data for a total of six locations which include Bangalore, Kanyakumari, Mumbai Chennai, Delhi, Kolkata in order to study the topology of India and applicability of various methods in various terrains.

4. Discussion and Analysis of Results

This algorithm program is used to check the accuracies for the method which include simple average method, moving average method, moving weight method, exponential smoothing method, linear trend regression.

The accuracy checks for the forecasting include mean absolute deviation, bias, running sum average method, tracking signal which help us understand the level of forecasting i.e. over forecasting and under forecasting and help us improve the final code model to attain high level of precision in forecasting for a short term using time series analysis.

If the forecast is more than the actual value, i.e. high positive deviation from the actual value then it is called over forecast else it is called as under forecast.

In order to get a better accuracy of logistical regression-trained data algorithms, the data cleaned and segregated to be fed for prediction must be previously optimized for highly accurate range.

The accuracy of the train is different for certain seasonal dates, i.e. 100 percent accuracy can be achieved for a few dates, whereas about 95 percent accuracy is the limit that can be archived for a Table 1. But, for this logistic regression algorithm, the minimum predicted precision for any date is 90 percent. Since the method chosen is Analysis of statistical data, the essence of the probabilistic prediction is and accurate.

The plots below show that there is a linear Figure-1, unlike cyclic fashion, between the two variables, i.e. power output and irradiance for each day shown in the previous variable. To feed the prediction algorithm, these variables can then be used.

The prediction accuracy:

Table 1

(date: 21-12)/Winter:

S. No	Location	Latitude	Prediction accuracy	Day hours	Irradiance
1	Chennai	13.0683	99.72	11.2	1.398
2	Kanyakumari	8.0857	100	11.53	1.531
3	Bangalore	12.9457	99.72	11.24	1.705
4	Mumbai	18.8543	100	10.89	1.622
5	Delhi	28.6585	99.5842	10.16	0.661
6	Kolkata	22.5126	91.47	10.61	1.178

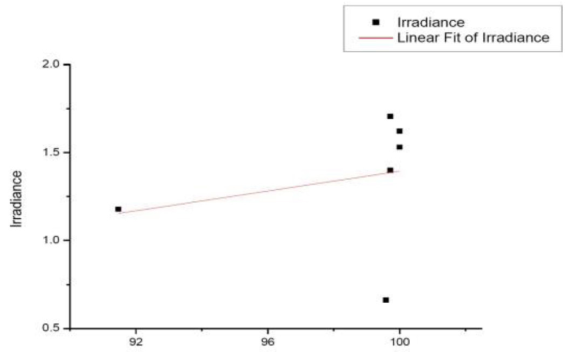


Figure 1
Graph for first week of winter Irradiance vs output

Table 2
(date: 22-05)/Summer

S.No	Location	Latitude	Prediction accuracy	Day hours	Irradiance
1	Chennai	13.0683	100	13.18	0.811
2	Kanyakumari	8.0857	100	12.69	0.571
3	Bangalore	12.9457	100	12.65	1.117
4	Mumbai	18.8543	99.792	13.57	1.147
5	Delhi	28.6585	99.794	12.95	0.662
6	Kolkata	22.5126	99.794	12.40	0.660

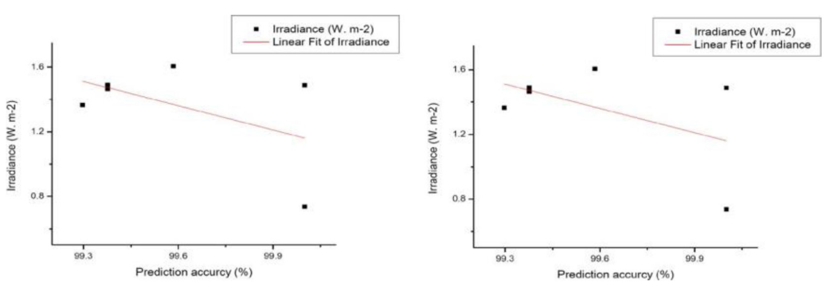


Figure 2
Graph for first week of Spring Irradiance vs output

Table 3

(date: 15-08)/Rainy

S.No	Location	Latitude	Prediction accuracy	Day hours	Irradiance
1	Chennai	13.0683	99.7525	12.78	0.371
2	Kanyakumari	8.0857	100	12.45	0.377
3	Bangalore	12.9457	100	12.43	0.421
4	Mumbai	18.8543	99.7921	13.04	1.155
5	Delhi	28.6585	100	12.63	0.626
6	Kolkata	22.5126	100	12.27	0.356

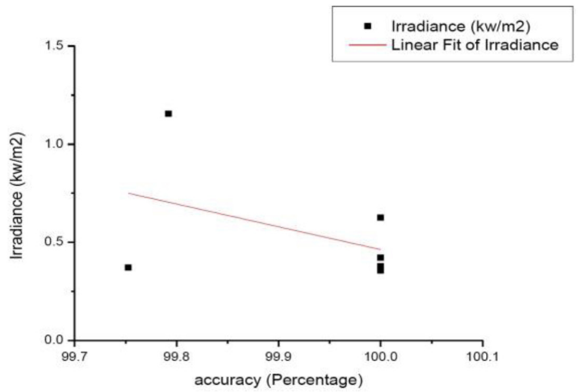


Figure 3

Graph for first week of Summer Irradiance vs output

Table 4

(date: 25-02)/Spring

S.No	Location	Latitude	Prediction accuracy	Day hours	Irradiance (kw/m^2)
1	Chennai	13.0683	99.2972	11.45	1.365
2	Kanyakumari	8.0857	99.5842	11.68	1.605
3	Bangalore	12.9457	99.3762	11.70	1.489
4	Mumbai	18.8543	99.3762	11.27	1.464
5	Delhi	28.6585	100	11.56	0.736
6	Kolkata	22.5126	100	11.81	1.488

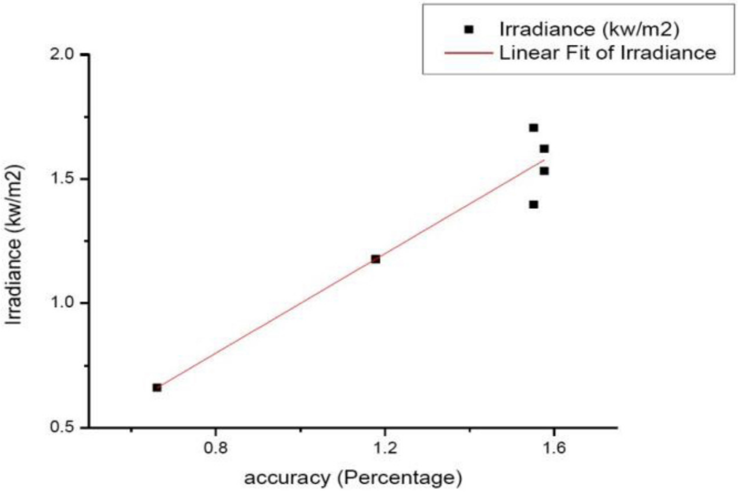


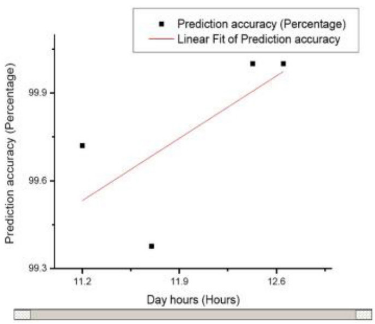
Figure 4

Irradiance vs. output plot for the month of August/Rainy season

Table 5

Bangalore : Latitude :12.9457

S.No	Season	Predictionaccuracy	Irradiance	Day hours
1	Winter	99.72	1.398	11.2
2	Spring	99.3762	1.489	11.70
3	Summer	100	1.117	12.65
4	Rainy	100	0.421	12.43



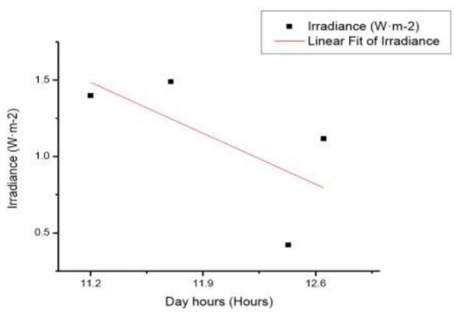
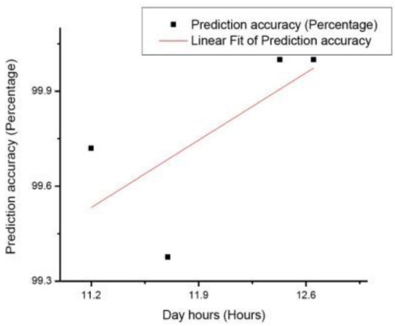


Table 6

Chennai : Latitude : 13.0683

S. No	Season	Prediction accuracy	Irradiance	Day hours
1	Winter	99.72	1.398	11.2
2	Spring	99.2972	1.365	11.45
3	Summer	100	0.811	13.18
4	Rainy	99.7525	0.371	12.78



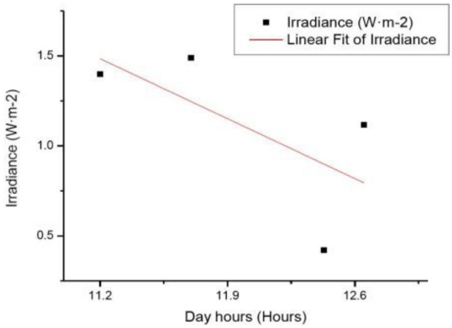
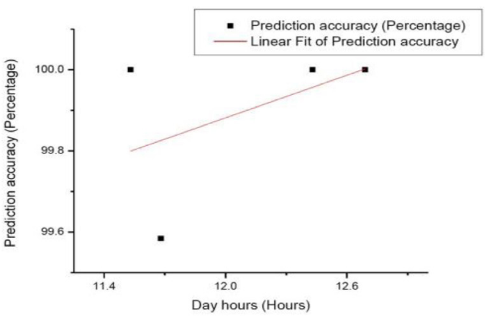


Table 7

Kanyakumari : Latitude : 8.0857

S. No	Season	Prediction accuracy	Irradiance	Day hours
1	Winter	100	1.531	11.53
2	Spring	99.5842	1.605	11.68
3	Summer	100	0.571	12.69
4	Rainy	100	0.377	12.43



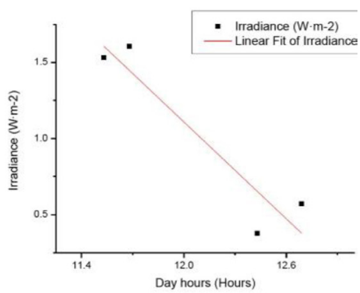
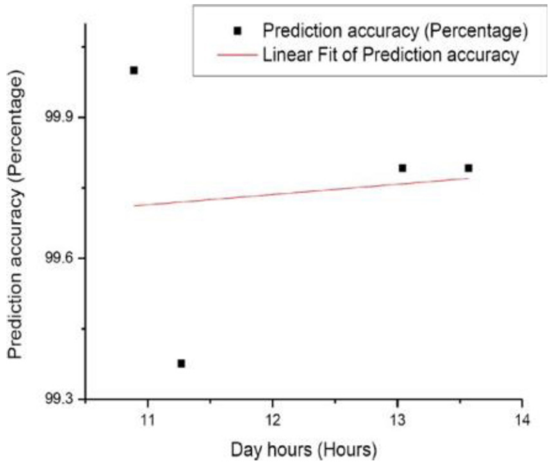


Table 8

Mumbai : Latitude : 13.0683

S. No	Season	Prediction accuracy	Irradiance	Day hours
1	Winter	100	1.622	10.89
2	Spring	99.3762	1.489	11.27
3	Summer	99.792	1.464	13.57
4	Rainy	99.7921	1.155	13.04



5. Conclusion

In this article, we have analyzed the forecasting accuracy for different Indian geographical locations and presented the graphical comparisons for Table 5, Table 6, Table 7 and Table 8, Delhi and Kolkata locations throughout the seasons.

We have calculated the likelihood of power released output in this article and performed comparative analysis for prediction accuracy throughout all the seasons in different geographical Indian locations with different latitude and longitudinal axis for different Indian metro cities and the power output and temperature plots were collected during the years.

We have applied the supervised machine learning algorithm for predicting the probability of solar power output by using logistic regression. This process can be also automated with highest accuracy for the specified location. Multinomial logistic regression is applicable in order determine seasonal forecasting accuracy for the Table 4, Table 2, Table 1 and Table 3 seasons and binomial logistic regression suits for checking the power release probability.

Unlike linear fashion, the above plots demonstrate that there is a non-linear relationship between the two variables, i.e. months and forecasting errors. for every month as shown in the figure.

Based on the above analysis the conclusion can be drawn as follows Bangalore

1. Irradiance is Figure 2
2. Irradiance is Figure 4
3. Prediction accuracy is high in winter
4. Day light hours is high in Summer Chennai

1. Irradiance is Figure 1
2. Irradiance is Figure 4
3. Prediction accuracy is high in Rainy
4. Day light hours is high in Summer Kanyakumari :

1. Irradiance is high in spring
2. Irradiance is low in Rainy
3. Prediction accuracy is up to maximum 100% in all the seasons except spring
4. Day light hours is high in Summer Mumbai:

1. Irradiance is high in winter
2. Irradiance is low in Rainy
3. Prediction accuracy is up to maximum 100% in winter and little above in other seasons
4. Day light hours is high in Figure 3 Delhi:

1. Irradiance is high in spring
2. Prediction accuracy is up to maximum 100% in spring and rainy seasons
3. Day light hours is high in Summer Kolkata:

1. Irradiance is high in spring
2. Irradiance is low in Rainy
3. Prediction accuracy is up to maximum 100% in spring and rainy seasons
4. Day light hours is high in Summer

References

[1] Harendra Kumar Yadav , Yash Pal , M.M. Tripathi "Photovoltaic Power Forecasting Methods in Smart Power Grid" IEEE INDICON 2015 1570186525.

[2] Hossein Sangrody , Morteza Sarailoo, Ning Zhou, Nhu Tran, Mahdi Motalleb, Elham Foruzan "Weather forecasting error in solar energy forecasting" *IET Renew. Power Gener.*, Vol. 11 Iss. 10, pp. 1274-1280 (2017).

[3] Gensler, A., Henze, J., Sick, B., &Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEn- coder and LSTM Neural Networks. In 2016 IEEE international conference on systems, man, and cybernetics (SMC) (pp. 002858-002865). IEEE (2016, October).

[4] Haupt, S. E., &Kosovic, B. Big data and machine learning for applied weather forecasts: Forecasting solar power for utility operations. In 2015 IEEE Symposium Series on Computational Intelligence (pp. 496-501). IEEE (2015, December).

[5] Khosravi, A., Koury, R. N. N., Machado, L., &Pabon, J. J. G. Prediction of hourly solar radiation in Abu Musa Island using ma- chine learning algorithms. *Journal of Cleaner Production*, 176, 63-75 (2018).

[4] Lauret, P., Voyant, C., Soubdhan, T., David, M., & Poggi, P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Solar Energy , 112, 446-457 (2015).

[5] Li, Z., Rahman, S. M., Vega, R., & Dong, B. A hierarchical ap- proach using machine learning methods in solar photovoltaic energy production forecasting Energies forecasting. *9*(1), 55 (2016).

[6] M. G. Lobo and I. S´anchez, "Regional wind power forecasting based on smoothing techniques, with application to the Spanish peninsular system," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1990–1997 (Nov. 2012).

[7] https://machinelearningmastery.com/logistic-regression-for-machine-learning/

[8] https:// www.renewables.ninja

[9] https://www.undp.org/content/undp/en/home/sustainable-development-goals.html

[10] https://sustainabledevelopment2015.org/sdg-conceptual-framework

[11] Programming Exercise 2: Logistic Regression-Machine Learning Course by Andrew Ng.

[12] V. Jagadeesh, K. Venkata Subbaiah & Jyothi Varanasi, Forecasting the probability of solar power output using logistic regression algorithm, Taylor& Francis Group, *Journal of Statistics and Management Systems,* Vol. 23 (1),2020,1-16

[13] Voyant, C., Notton, G., Kalogirou, S., Nivet, M. L., Paoli, C., Motte, F., &Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105, 569-582 (2017).

[14] Sharma, N., Sharma, P., Irwin, D., &Shenoy, P. Pre- dicting solar generation from weather forecasts using machine learn- ing. In 2011 IEEE international conference on smart grid communications (SmartGridComm) (pp. 528-533). IEEE (2011, October).

[15] Mompati Koorapetse & P. Kaelo (2019) An efficient hybrid conjugate gradient-based projection method for convex Mompati Koorapetse & P. Kaelo. An efficient hybrid conjugate gradient-based projec- tion method for convex constrained nonlinear monotone equations, *Journal of Interdisciplinary Mathematics,* 22:6, 1031-1050 (2019).

Received May, 2020