

# Using Artificial Intelligence Techniques for Prediction and Estimation of Photovoltaic System Output Power

Deogratias Nurwaha



Institute of Applied Pedagogy, Department of Physics and Technology, University of Burundi, Bujumbura

\* Corresponding author email: [deogratiasnurwaha@gmail.com](mailto:deogratiasnurwaha@gmail.com)

Received: 24 January 2020 / Accepted: 05 March 2020 / Published: 06 March 2020

## Abstract

Two artificial intelligence methods, namely, support vector machines (SVM) and gene expression programming (GEP), were explored for prediction and estimation of the Photovoltaic (PV) output power. Measured values of temperature  $T$  ( $^{\circ}\text{C}$ ) and irradiance  $E$  ( $\text{kWh}/\text{m}^2$ ) were used as inputs (independent variables) and PV output power  $P$  (Kw) was used as output (dependent variable). The statistical metrics were used to assess the predictive performances of the methods. The results of the two models were estimated and compared. The results showed that the two techniques performances are better and similar. Using GEP technique, the relationships between the two parameters and output power were established. Importance of each parameter as contributor to PV output power was also investigated. The results indicated that the SVM and GEP would become the powerful tools that could help estimate the PV output power capacity reserve.

**Keywords:** artificial intelligence; PV output power; GEP; SVMs.

## 1 Introduction

Now days, the photovoltaic (PV) energy conversion plays a crucial role in supplying electrical energy worldwide. The use of (PV) presents a number of advantages such as reducing greenhouse gas emission, consuming less fossil fuel, providing higher penetration of renewable energy source, heating, lighting, etc. Also, depletion and price increase of fossil based fuels, global warming, air pollution, strict environmental laws, etc. are promoting the use of solar energy. The performance of PV systems is directly affected by various parameters such as irradiance ( $E$ ), ambient temperature ( $T$ ), etc. The main challenge of a PV operator is to predict changes of the PV output power in order to estimate the reserve capacity. More recently, few studies using artificial intelligence approaches have been used to predict and model the PV power production [1]-[4]. However, their prediction accuracy is still a controversial issue and more attention is needed in order to achieve acceptable predictable accuracy. The main objective of this study is to explore the newly artificial intelligence methods to predict the PV output power. Support Vector Machines method and Gene expression Programming will be used. The results will be compared.

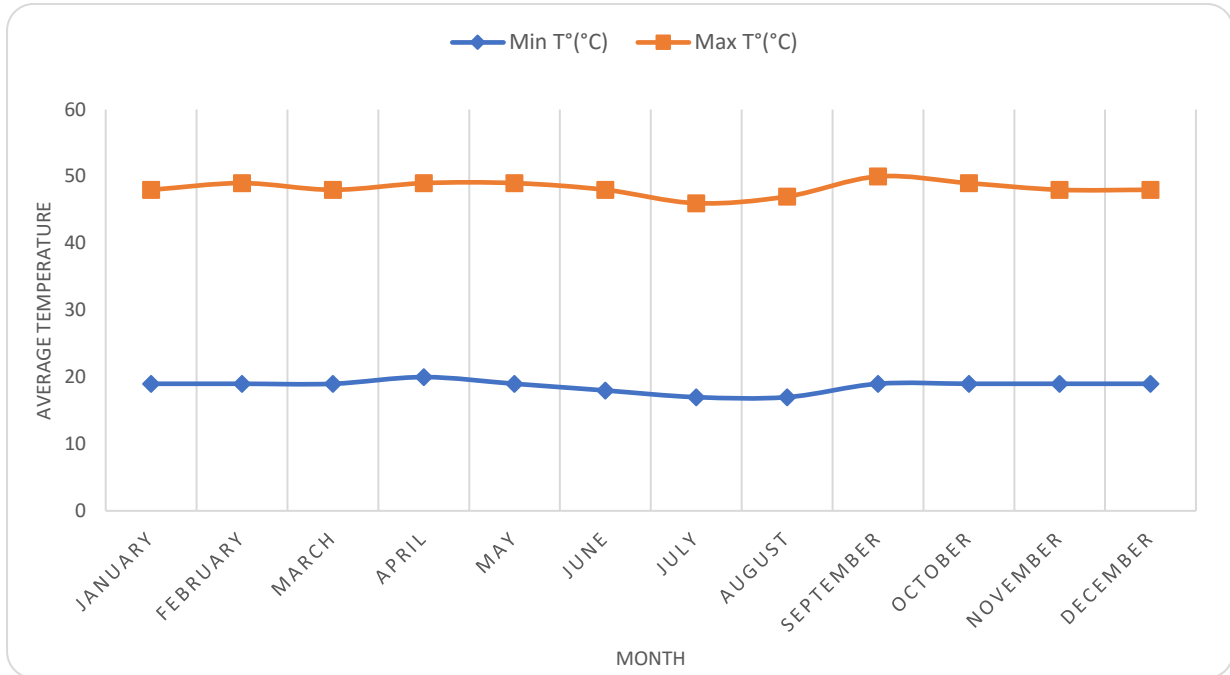
## 2 Data collection

Data used in this study were collected from the PV system of the Hospital of the University of Burundi. Figure 1 shows the configuration of the system. The PV site has latitude of  $3^{\circ}24'$  and longitude of  $29^{\circ}21'$ . The average temperature is of  $29^{\circ}\text{C}$ .



Figure 1: PV system configuration

Figure 2 illustrates the fluctuations of the temperature during the year. The annual average of solar irradiance is of 4.98kW/m<sup>2</sup> and due to its geographical position near the equator (3° S), it has not major fluctuations during the year. 1920 PV panels are installed on the area of 6300 m<sup>2</sup>. Its power capacity is of 400kW.



**Figure 2:** PV system site's temperature fluctuations during the year

Table 1 represents the specifications of the PV system. Hourly raw data including solar irradiation  $E$  ( $kWh/m^2$ ), temperature  $T$  ( $^{\circ}C$ ) and output power  $P$  (kW) were collected for a period of two months in rainy season (February and March) and two months in sunny season (July and August) 2017. Table 2 and Table 3 show the attributes of the collected data sets as well as their statistical properties.

**Table 1:** Specifications of the PV system

Number of Panels	1920
Output power: $P$ (kW):	400 kW
Open circuit voltage: $V_{OC}$ (V)	407.2
Short circuit current: $I_{CC}$ (A)	1336.8
Current at maximum power: $I_{pm}$ (A)	1221.6
Voltage at maximum power: $V_{pm}$ (V)	330.4
Conversion efficiency: $\eta$ (%)	17.2
Area (m <sup>2</sup> )	6300

**Table 2:** Statistical properties of the attributes of collected data in rainy season

Variable	Minimum	Maximum	Mean	Standard Deviation
E	1.64	7.08	4.80	1.37
T	21.47	26.42	24.24	1.25
P	140.86	550.35	373.92	100.22

**Table 3:** Statistical properties of the attributes of collected data in sunny season

Variable	Minimum	Maximum	Mean	Standard Deviation
E	2.28	6.72	5.63	0.69
T	22.22	26.66	24.95	0.81
P	142.05	469.46	368.58	51.28

### 3 Implementation of the two methods

GEP evaluation was performed using 10-fold cross-validation. The Generations required to train the model, the complexity of the model before simplification, the complexity of the model after simplification, the generations required for simplification and a number of evaluations of the fitness function were optimized on trial and error based. The values of these parameters are given in Tables 6&7. More details on the GEP can be found in [5, 6].

For the SVMs, we first apply the RBF kernel function to map the data into a different space where a hyperplane can be used to do the separation. The effectiveness of SVM depends on the selection of kernel, the kernel's parameters and the soft margin parameter  $C$ . A common choice is a Gaussian kernel, which has a single parameter  $\gamma$  [7, 8]. To find the optimal parameter we used grid and pattern search methods on  $C$  and  $\gamma$ , given in tables 4 and 5. We conducted grid search and pattern search methods using four and 10-fold cross-validation on the training and the validation data and reported the validation results. One subset is chosen for training and remaining 9 subsets are used for testing and the process is repeated until all the subsets are chosen for the testing. For implementation, DTREG software [9] was used to execute both SVM and GEP.

**Table 4: SVM parameters (rainy period data)**

Parameter	Epsilon	C	$\gamma$	P	NSV
Values	0.001	5000.08295	0.01114167	0.91279617	57

**Table 5: SVM parameters (sunny period data)**

Parameter	Epsilon	C	$\gamma$	P	NSV
Values	0.001	14999.0494	4.62078545	2.29057834	57

NSV: number of support vectors

**Table 6: GEP parameters (rainy period data)**

Parameter	GRTM	CMBS	CMAS	GRS	NEFF
Values	1.891	34	15	92	114.7

**Table 7: GEP parameters (sunny period data)**

Parameter	GRTM	CMBS	CMAS	GRS	NEFF
Values	1.456	21	15	1	110.15

**GRTM:** Generations required to train model; **CMBS:** Complexity of model before simplification; **CMAS:** Complexity of model after simplification; **GRS:** Generations required for simplification; **NEFF:** Number of evaluations of the fitness function.

### 4 Results analysis and discussion

For comparison purposes, data collected in rainy and sunny periods were used. Prediction performances were evaluated using the statistical criteria, namely, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). RMSE and MAE are the measures of the deviation between the actual and predicted values. The smaller the values of RMSE and MAE, the closer are the predicted power values to the actual power values. All these statistical criteria are defined by equations (1)-(3). Other statistical metrics usually used such as mean squared error (MSE), proportion of variance explained by model ( $R^2$ ) and correlation between actual and predicted ( $R$ ) will be considered for comparison reasons.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - P_p)^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - P_p| \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|P_i - P_p|}{|P_i|} \times 100\% \quad (3)$$

Where  $n$  is the number of pairs,  $P_i$  and  $P_p$  are the  $i$ -th actual and calculated outputs, respectively. Table 8&9 respectively show the training and validation results of the two models. Figures 3 represent the 10-fold cross validation process. The mathematical expressions, as the models to represent the PV output power as a function of irradiance and temperature, were generated and are given by equations (4) and (5) for the rainy and sunny periods, respectively:

$$PR = (59.426669 * E) + ((34.443245 - T) * T) - 122.38452 + (-3.215283) * T \quad (4)$$

$$PS = (59.751209 * E) + (-509.5) + (T * (((-0.5931767) * T) + 38.017001)) + T \quad (5)$$

**Table 8: Training results for the two models**

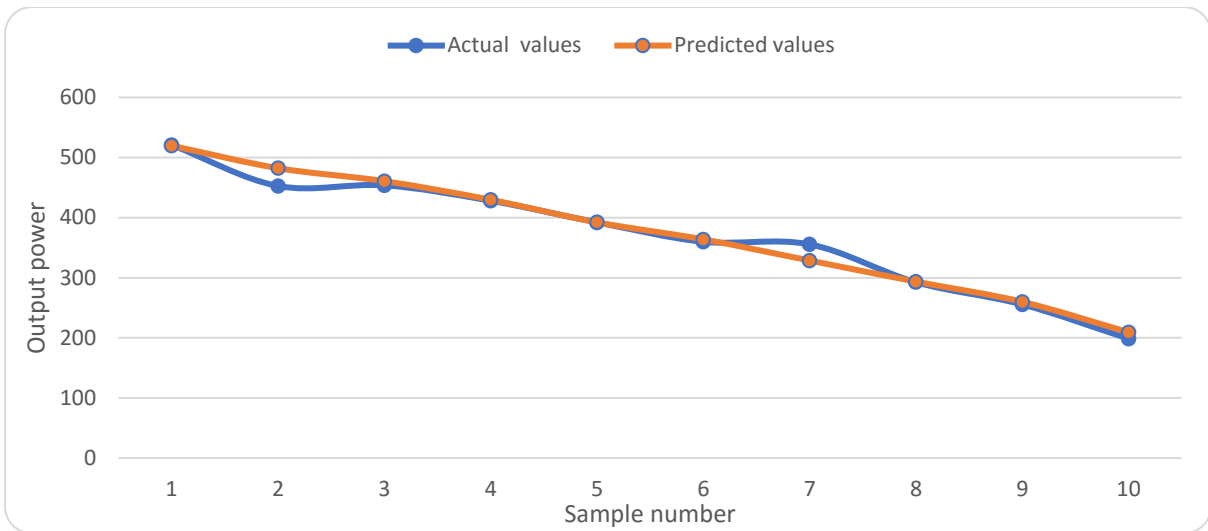
Statistical parameter	Rainy period data		Sunny period data	
	SVM	GEP	SVM	GEP
RMSE	38.68	38.34	18.36	30.27
MAE	19.35	21.00	12.82	23.86
MAPE (%)	5.46	5.89	3.43	6.95
MSE	1496.74	1470.54	337.14	916.51
R <sup>2</sup> (%)	85.09	85.36	87.18	65.15
R (%)	92.37	92.39	93.42	80.72

**Table 9: 10-fold cross validation results for the two models**

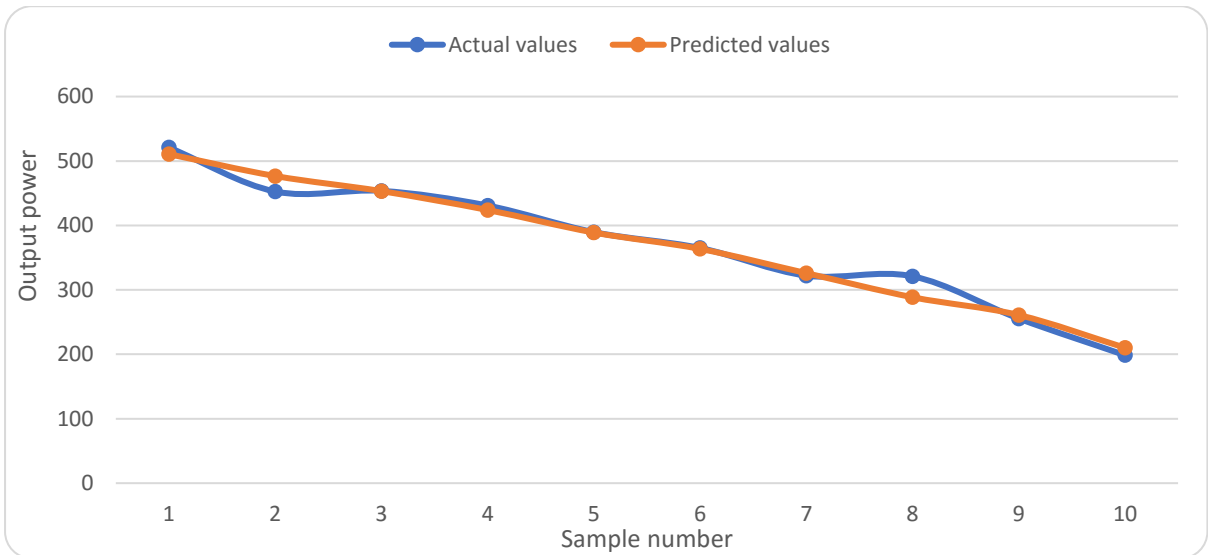
Statistical parameter	Rainy period data		Sunny period data	
	SVM	GEP	SVM	GEP
RMSE	38.88	40.49	47.91	36.51
MAE	20.38	22.42	23.74	28.16
MAPE (%)	5.77	6.24	8.89	8.27
MSE	1512.22	1640.23	2295.52	1333.20
R <sup>2</sup> (%)	84.94	83.67	54.04	49.31
R (%)	92.26	91.48	12.73	73.34

**Table 10. Computed importance of irradiance and temperature contribution on PV output power (rainy period)**

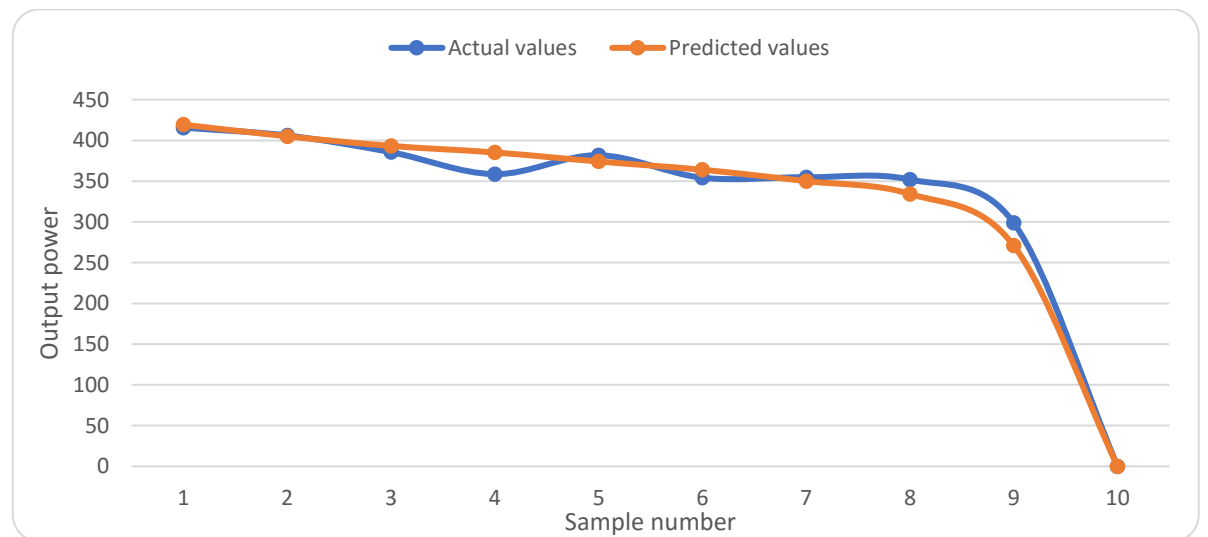
Method	SVM		GEP	
Variable	E	T	E	T
Importance (%)	100	1.97	100	2.96



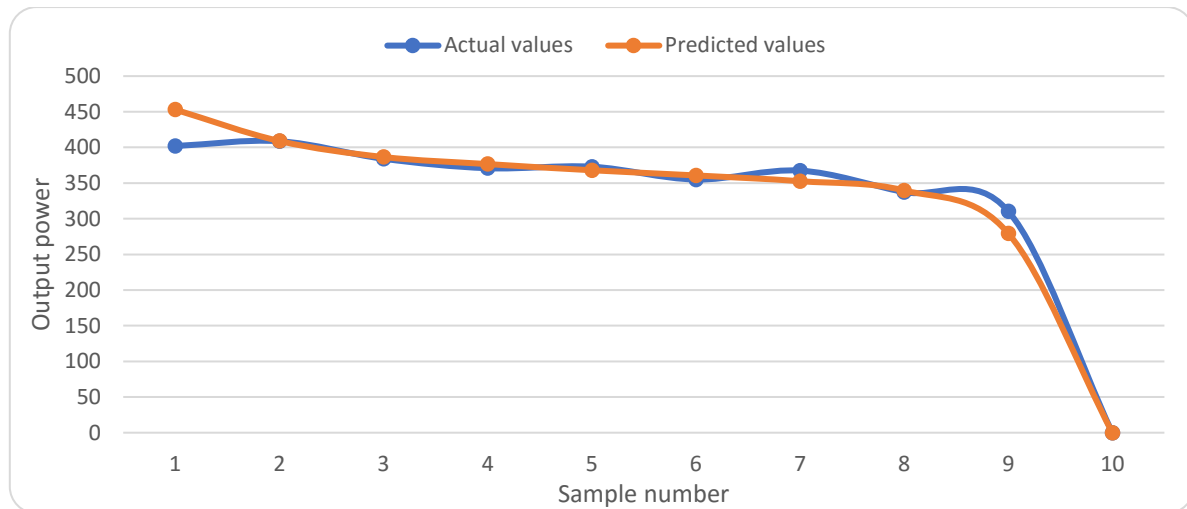
**Figure 3 (a):** 10-fold cross validation for SVM (rainy period data)



**Figure 3 (b):** 10-fold cross validation for GEP (rainy period data)



**Figure 3(c):** 10-fold cross validation for GEP (sunny period data)



**Figure 3 (d):** 10-fold cross validation for SVM (sunny period data)

For SVM, training RMSE=18.36, MAE=12.82 and MAPE=3.43 for the sunny period data set are smaller than that of the rainy period data (RMSE=38.68, MAE=19.35, MAPE= 5.46). But the reverse situation is observed for the validation results where (RMSE=38.88, MAE=20.38 MAPE=5.77) respectively for rainy season against (MSE=47.91, MAE=23.74 MAPE= 8.89) for the sunny season data set.

For GEP, training RMSE=38.34 is greater for the rainy season data than that of the sunny season data (RMSE=30.27) while MAE=21.00 and MAPE=5.46 for the rainy season data are smaller than that of sunny period data (MAE=23.86, MAPE=6.95). Validation RMSE=40.49 for the rainy period data against RMSE=36.51 for the sunny period data, MAE=22.42, MAPE=6.24 for the rainy period data against MAE=28.16, MAPE=8.27 for the sunny season data.

The two models provided better validation correlations between actual and predicted output power values ( $R=92.26$  for SVM,  $R=91.48$  for GEP) for the rainy period data and worst correlations for the sunny period data ( $R=12.73$  for SVM,  $R=73.34$  for GEP). From the above observations, important information can be retained: the two models predictive accuracies levels depend on the selected data set used. To judge the predictive performances, the validation results provided by the two models must be analyzed and compared. As it can be seen from the results, the two models provide better performances with a little difference in favour of SVM model. Computed importance of irradiance and temperature contributions on PV system output power are illustrated in Table 10. From the table, even though contribution of the temperature is non negligible, it can be seen that the irradiance is the main contributor in performing the PV output power. Also, from equations (4) & (5), it is observed that the relationship between irradiance and PV output power is linear while relationship between temperature and PV output power is nonlinear.

## 5 Conclusion

This study proposed two prediction models for PV system output power using artificial intelligence methods, namely, support vector machines (SVM) and gene expression programming (GEP). These techniques have been successfully applied to a wide range of pattern recognition problems. Predictive accuracy of each model has been evaluated using statistical criteria. The results showed that the two intelligent techniques provided good and similar accuracies. Relationships between output power and the two parameters have been illustrated. The main advantage of the explored methods is that the obtained results could allow the PV system operator to estimate the capacity of energy reserve. This advantage is not yet widely explored in controlling solar cell output energy. However, alternative artificial intelligence techniques should be explored and compared in order to assess the expected predictive accuracy level.

## 6 Competing Interests

Author declares that no potential conflict of interest exists related to this article

### How to Cite this Article:

D. Nurwaha, "Using Artificial Intelligence Techniques for Prediction and Estimation of Photovoltaic System Output Power", *J. Mod. Sim. Mater.*, vol. 3, no. 1, pp. 15-21, Mar. 2020. doi : 10.21467/jmsm.3.1.15-21

### References

- [1] R. Hossain, A. Maung Than O, A. B. M. Shawkat Ali, Hybrid Prediction Method for Solar Power Using Different Computational Intelligence Algorithms, *Smart Grid and Renewable Energy*, 2013, 4, pp.76-87.
- [2] Soteris A. Kalogirou and A.Şencan, Artificial Intelligence Techniques in Solar Energy Applications, [www.intechopen.com](http://www.intechopen.com).
- [3] Hussein A Kazem, Jabar H. Yousif, Miqdam T Chaichan, Modelling of Daily Solar Energy System Prediction using Support Vector Machine for Oman, *International Journal of Applied Engineering Research* ISSN 0973-, 2016, pp. 10166-10172.
- [4] F. H. Anuwar and A. M. Omar, Future Solar Irradiance Prediction using Least Square Support Vector Machine, *International Journal on Advanced Science Engineering Information technology*, 2016, pp.
- [5] Ferreira, C. "Gene Expression Programming: A New Adaptive Algorithm for Solving Problems" (PDF). *Complex Systems*, 2001, pp. 87–129.
- [6] Ferreira, C. (2002). "Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence". Portugal: Angra do Heroismo. ISBN 972-95890-5-4.
- [7] Cortes, Corinna; Vapnik, Vladimir N, (PDF). *Machine Learning*, 1995, pp. 273.297. doi:10.1007/BF00994018.
- [8] Ben-Hur, Asa; Horn, David; Siegelmann, Hava; and Vapnik, Vladimir N.; "Support vector clustering"; *Journal of Machine Learning Research*, 2001, pp. 125–137.
- [9] <https://dtreg.soft112.com/>

#### Publish your research article in AIJR journals-

- Online Submission and Tracking
- Peer-Reviewed
- Rapid decision
- Immediate Publication after acceptance
- Articles freely available online
- Retain full copyright of your article.

Submit your article at [journals.ajjr.in](http://journals.ajjr.in)

#### Publish your books with AIJR publisher-

- Publish with ISBN and DOI.
- Publish Thesis/Dissertation as a Book.
- Publish Monograph.
- Publish Edited Volume/ Book.
- Publish Conference Proceedings
- Retain full copyright of your books.

Submit your manuscript at [books.ajjr.org](http://books.ajjr.org)