VOL. 10, NO. 3, FEBRUARY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

UTILIZATION OF ARTIFICIAL IMMUNE SYSTEM IN PREDICTION OF PADDY PRODUCTION

A. B. M. Khidzir[1], M. A. Malek[2], Amelia Ritahani Ismail[3], Liew Juneng[4] and Ting Sie Chun[1]
[1]Department of Civil Engineering, Universiti Tenaga Nasional, Kajang, Malaysia
[2] Institute of Energy, Policy and Research (IEPRe), Universiti Tenaga Nasional, Kajang, Malaysia
[3]Department of Computer Science, Kuliyyah of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia
[4]School of Environmental and Natural Resource Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
E-Mail: ainulbmk@gmail.com

ABSTRACT

This paper proposed an Artificial Immune System (AIS) approach using the Clonal Selection Based Algorithms (CSA) to analyze the pattern recognition capability of the paddy trend, and to predict the paddy production based on climate change effects. Climate factors and paddy production are used as input parameters. High percentage of accuracy ranges from 90%-92% is obtained throughout the training, validation and testing steps of the model. The results of the study were tested using the Root Mean Square Error (RMSE), Mean Average Percentage Error (MAPE) and coefficient of determination (R^2). Based on the results of this study, it can be concluded that the CSA is a reliable tool to be used as pattern recognition and prediction of paddy production.

Keywords: Artificial Immune System (AIS), climate, Clonal Selection Based Algorithm (CSA), paddy production, prediction.

1. INTRODUCTION

Rice is consumed as a staple food by most people all over the world especially in Asia (Iqbal, 2007). Figure-1 shows the statistic of rice production by region. It shows that Asia contributes approximately 90 percent to the world's rice production (Food and Agriculture Organization of the United Nations, 2014).

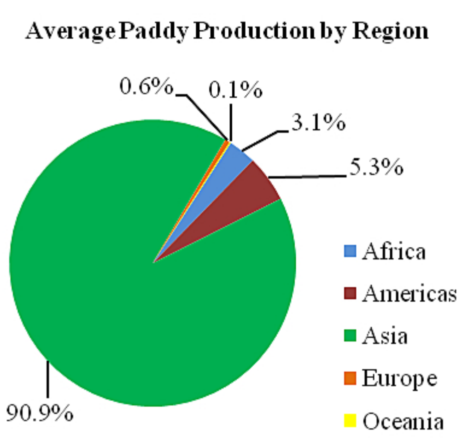


Figure-1. Average paddy production by region.

China, India, Indonesia and Myanmar are among the top rice producers in Asia. In Malaysia, cultivation of paddy focuses on self consumption. According to the Malaysia Agricultural Research and Development Institute (MARDI), local rice production in Malaysia caters only 70% to 80% of the country's needs. Thus, continuous efforts are carried out in order to increase the local rice production. This includes the introduction of hybrid rice technology which is able to increase the rice yield up to 20% (Malaysian Agricultural Research and Development Institute, 2013).

Pests, crop diseases, soils, fertilizers etc. are among the various factors that contribute towards paddy growth. Apart from these factors, researchers are currently probing further into the effects of climate change to the paddy growth. The climate change is due to the natural variability or as results of human activity (Intergovernmental Panel on Climate Change, 2007). The climate change caused extreme events such as droughts and floods. Studies have shown that climate change will physically affect the growth of paddy and ultimately reduce the rice production and its quality (Ramirez, 2010).

Paddy cultivation period can be divided into three main stages which are vegetative stage, mid season stage and late season stage. The amount of water required at each stage varies from 0mm-100mm. The adequacy amount of water in every stage is important for the survival of paddy plant. At late season stage, hot weather is expected in order to help the ripening process of paddy plant (Mohamed Azwan, Mustapha and Puasa).

AIS is an algorithm and system inspired by the biological immune system. According to Castro and Timmis (2002), "Artificial Immune System (AIS) are adaptive systems, inspired by theoretical immunology and observed immune functions, principles and models, which are applied to problem solving." The main function of a biological immune system is to defend the human body from foreign elements known as antigens. Antigens can be parasites, bacteria or viruses. These antigens will then be eliminated or neutralized by the immune system (Castro and Timmis, 2002).

The immune system works based on two defense mechanisms which are innate and adaptive immune systems. Innate immunity will attack against basic pathogens that enter the body and are present almost immediately after birth. On the other hand, adaptive immunity will attack against complex and mutated pathogens that cannot be removed by innate immunity (Castro and Timmis, 2002).

There are three simplified models of algorithms in AIS which are Clonal Selection based Algorithms (CSA), Negative Selection based Algorithms (NSA) and Immune Networks (IN) (Greensmith, Whitbrook and Aickelin, 2010).

Each of these algorithms is applied for different purposes. For example, CSA mostly used for pattern recognition and optimization. (Castro and Timmis, 2002). NSA usually applied in fault detection and computer security application while IN commonly used for clustering, classification, data analysis and data mining applications. Based on these features, CSA has been chosen to be used in this study (Al-Enezi, Abbod and Alsharhan, 2010).

The reason AIS method has been proposed in this study is because AIS is a new branch of computational intelligence family which can be used to solve complex problems. It has strong capabilities in pattern recognition, function approximation, learning and associative memory and data analysis. Therefore, AIS can be viewed as a powerful information processing and problem-solving tool (Zhong, Zhang, Gong and Li, 2007).

To date, the usage of the Artificial Immune System in crop prediction is uncommon.

The objective of this study is to develop a prediction model on paddy production with respect to climate change.

2. METHODOLOGY

2.1. Datasets

Data on paddy production for type MR 84 variety were acquired from the Muda Agricultural Development Authority (MADA), Kedah, Malaysia through the help from MARDI.

Climate data such as rainfall, wind speed, humidity, solar radiation and temperature (Subash and Mohan, 2012) were acquired from the School Of Environmental And Natural Resource Sciences, Faculty Of Science And Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia.

Both datasets used 180 monthly samples from the year 1985 to year 1999. The analysis was made within the year 1985-1999 because the paddy variety was planted within that year only. This was requested by Malaysian Agricultural Research and Development Institute (MARDI) as there might be chances this type of paddy variety will be replanted in the future. This study focuses on paddy plots within the MADA cultivation areas in Kedah, Malaysia.

2.2. Empirical method

Empirical according to Oxford English Dictionary is "based on, concerned with, or verifiable by observation or experience rather than theory or pure logic". Several empirical models used in paddy prediction are as follows:

Subash et al. (2012) applied DSSAT-CSMv4.5 model to evaluate the impact of climatic trends and variability in rice-wheat system productivity. This study focuses on Indo-Gangetic Plains of India. This model utilized genotypic coefficients, soil profile, field management and weather data as input. The daily weather data used were: rainfall, minimum and maximum temperature and sunshine hours. In this study it has been found that there is very good agreement in the results of simulation and the actual data which were compared in terms of Spearman correlation coefficient (r_s).

Perez et al. (2002) constructed CATCHCROP model to simulate crop yields response to water deficit and fertility depletion in Northern Thailand. Among the studied crops are paddy, maize, potato, soybean and groundnut. The model input parameters are daily rainfall data (five years set of data), daily infiltration rate, potential evapotranspiration, vegetative period duration, water stress coefficient, soil type, soil depth, crop coefficient. Results show are fairly satisfactory. However due to limited availability and reliability of field data, this model is not appropriate to be used at the farm plot level to assess cropping practices.

Timnisa et al. (2006) analyzed the performance of CERES-Rice models using compiling data from various studies across Asia and Australia. The models are evaluated in terms of anthesis and maturity dates, grain and biomass yield, in-season Leaf Area Index (LAI) and growth, nitrogen uptake and loss as well as methane emission. The model input parameters used in this study were anthesis and maturity dates, in-season leaf area index (LAI) and growth, grain and biomass yield, soil water and nitogen. It has been found that the yield prediction of CERES-Rice for grain and biomass is quite varied, with generally good predictions under optimal Nitrogen and water conditions. In addition, the anthesis and maturity dates for CERES-Rice is fairly well predicted. However, this model shows poor performance under low Nitrogen, water deficit and low temperatures conditions.

2.3. Clonal selection based algorithm (CSA)

CSA is the theory used to explain basic response of adaptive immune system to an antigenic stimulus. It describes how the immune cells eliminate a foreign antigen and develop the idea that only cells which are able to recognize an antigenic stimulus will proliferate and differentiate into memory cells and plasma cells (Castro and Timmis, 2002). This clonal selection idea leads to step in finding a solution of a problem through a process of cloning, mutation and selection.

Figure-2 shows the diagram on the clonal selection principle process. In the selection stage, an antigen will bind with antibody (carried by cell) that have the same shape as itself. Then, the cell is activated and started to proliferate. A clone will arise. This clone will then differentiate into memory cells and plasma cells.

ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Most of the clonal selection based algorithms are applied to optimize problems and multi-objective optimization (Timmis, Hone, Stibor and Clark, 2008). A study made by Lakshmi and Vasantharathna (2014) has proven the CSA approach is an efficient optimization algorithm to solve the wind-thermal energy systems scheduling problem.

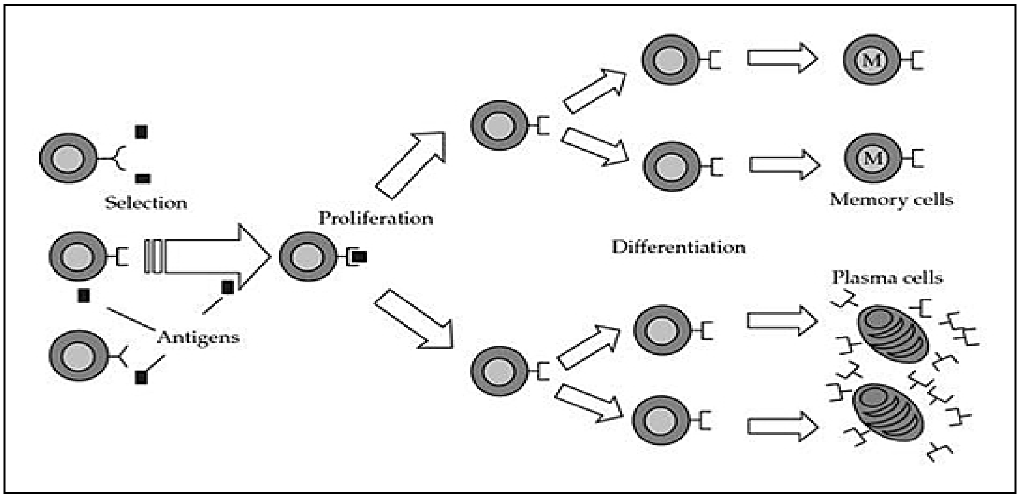


Figure-2. The clonal selection principle process. (Castro & Timmis, 2002).

CSA also been applied as the immune learning, memory and pattern recognition tools (Castro and Timmis, 2002). Chen and Zang (2009) indicated that CSA is a good immune learning algorithm and pattern recognition for the structure damage classification problem. Also a research by Cheng et al. (2012) has signified the accuracy and effectiveness of CSA in pattern recognition for primary open-angle glaucoma problem. CSA is simple and gives efficient approximation algorithm for achieving optimum solution (Nanda, 2009).

2.4. Model development

In this study, climate parameters represent a set of antigen (Ag) while paddy production represents a set of antibody (Ab). Figure-3 shows the CSA model development framework. Development of the proposed model is divided into four main steps:

- **Step 1: Data training**

k-fold cross-validation type is applied. The k-fold cross-validation type was chosen because it has the advantage where all datasets are eventually used for training, validation and testing (Price, Ramsden, Hope, Friston and Seghier, 2013). Cross-validation is performed by dividing the data into 4-folds equal size of subsamples. These subsamples were rearranged at 10 random sequences and each of the four different subsamples was run at 10 iterations. Figure-4 illustrates an example of cross-validation performed at two random sequences. It is found that iteration more than 10 times will produce approximately similar results at 10 iterations.

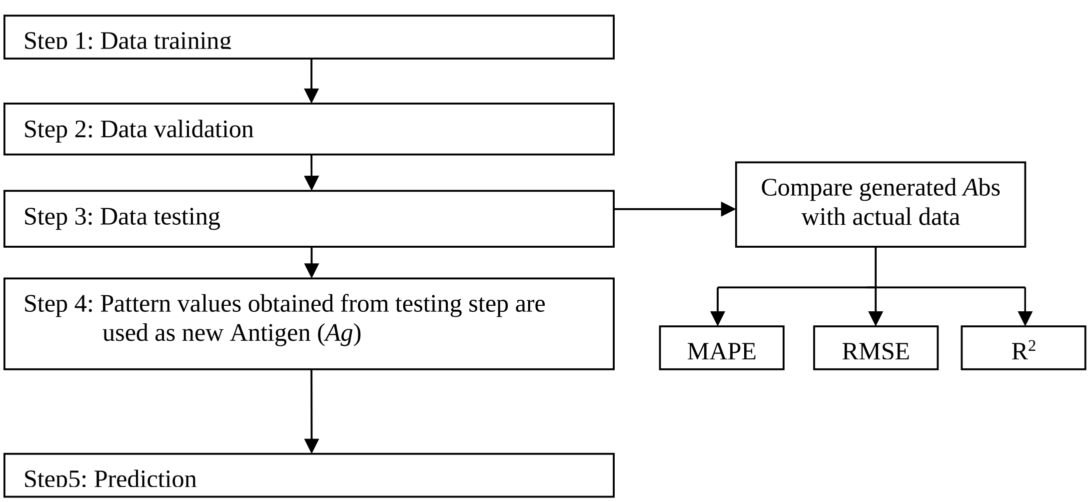
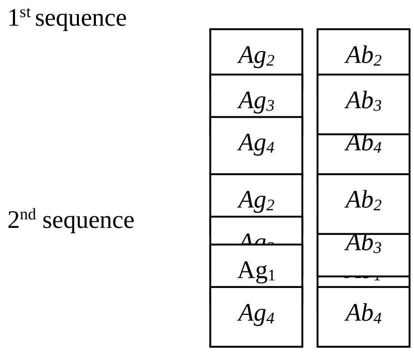


Figure-3. CSA framework.

VOL. 10, NO. 3, FEBRUARY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

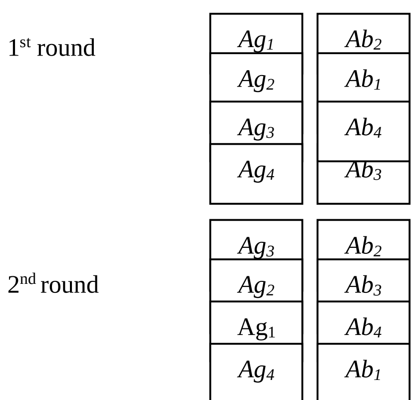


Ag = climate parameters data
Ab = paddy production data

Figure-4. Example of cross-validation for two sequences.

- **Step 2: Validation**

In this step, Ag and Ab data are cross with each other for 10 rounds. Each round runs at 10 times for every data cross. Figure-5 illustrates the validation step.



Ag = climate factors data
Ab = paddy production data

Figure-5. Example of validation step for two rounds.

- **Step 3: Data testing**

In this step, the pattern values are obtained. The equation used to find pattern values is as follows:

$$\text{pattern value} = \frac{[\text{shortest}+\text{rand}\ (1,1)/\text{no.of sa}}{[(v+(x-0.8)*z]} \qquad (1)$$

where
v = dimension number
x = value of x is obtained through trial. The final value of x is chosen when the matrix dimension of pattern value and input data are agreeable to each other. Example, if the matrix dimension of the input data is 12x1, therefore the final pattern value must also be 12x1 matrix dimension.
z = number of parameter

The pattern values obtained are used as the current Ag (climate parameters). The final testing step is performed by comparing the generated antibodies (Abs) with actual data.

- **Step 4: Prediction**

Final pattern values obtained at testing step are used as the new Ag values for future prediction on paddy yield.

3. RESULTS AND DISCUSSIONS

The proposed CSA prediction model was developed using self-written coding via MATLAB R2012a. Results obtained from the training, validation, testing and prediction steps are presented and discussed as follows:

3.1. Training

Table-1 shows results of the 1st fold of data at 10 runs. Results shows an average value of 92.04075. The same procedure was repeated four times for different fold as presented in Table-2. The values of average accuracy obtained from cross-validation of 10 sequences are exhibited in Table-3. The final result of cross-validation is obtained by averaging the average percentage of accuracy for 10 runs. In this study, the value obtained is 91.80%.

Table-1. Results of 1st fold cross-validation at 10 runs.

No. of run	Percentage of accuracy
1	92.0407
2	92.0406
3	92.0409
4	92.0408
5	92.0407
6	92.0407
7	92.0408
8	92.0406
9	92.0406
10	92.0411
Average	92.04075

Table-2. Average value of cross-validation of 1st sequence for four fold.

No. of fold	Average of 10 runs
1	92.04075
2	92.35736
3	90.63722
4	92.18127
Average	91.80415

Table-3. Cross-validation for 10 sequences at four folds of 10 runs.

No. of sequence	Average percentage accuracy for 10 runs.
1	91.80415
2	91.80414
3	91.80411

4	91.80415
5	91.80412
6	91.80413
7	91.80415
8	91.80411
9	91.80411
10	91.80412
Average	91.80413

3.2. Validation

Results in Table-4 represents the compatibility between Ag, Ab and the proposed model. The final result obtained in validation step for 10 rounds of cross data at 10 times run is 91.81115%.

Table-4. Results for 10th round of validation step run at 10 times.

No. of round	Percentage of accuracy
1	92.26938
2	92.111
3	91.40916
4	91.33889
5	92.19905
6	91.49725
7	91.49724
8	92.11099
9	91.40923
10	92.26937
Average	91.81115

3.3. Testing

The pattern values obtained are used as pattern recognition for testing and prediction steps. Figure-6 shows the graph of generated Abs (paddy production) against the actual data. The graph shows very small difference in value between the generated Abs and actual data.

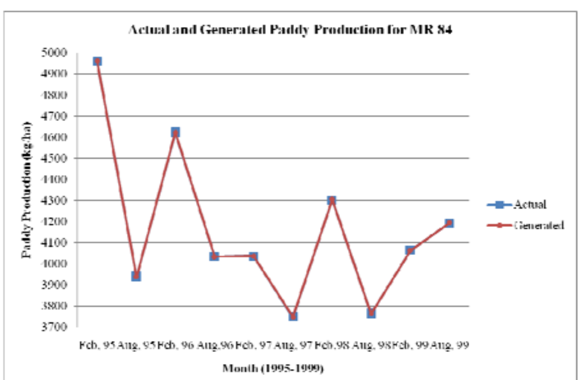


Figure-6. Graph of generated Abs and actual paddy production.

These results were further evaluated using statistical tests such as the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2) as shown in Figure-7. The RMSE is found to be 0.63, MAPE is 0.0103% and R^2 is 1.

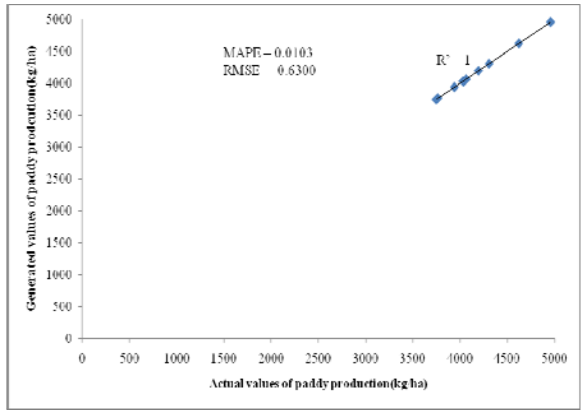


Figure-7. Statistical tests used for model accuracy.

3.4. Prediction

In this study, the prediction of future paddy production type MR 84 variety is carried out for five years, 2000-2004. In Malaysia, paddy is produced twice a year, which are in the month of February and August. Based on the prediction model developed in this study, Figure-8 shows the prediction values of paddy production for year 2000 till the year 2004. Results from the graph shows that high paddy production are mostly found in February. This is because weather during this cultivation period suits the needs of the paddy growth. Since the results of RMSE, MAPE and R^2 in testing stage are found to be between 60%-90% accuracy, therefore it can be concluded that the prediction performed from this proposed model has high accuracy.

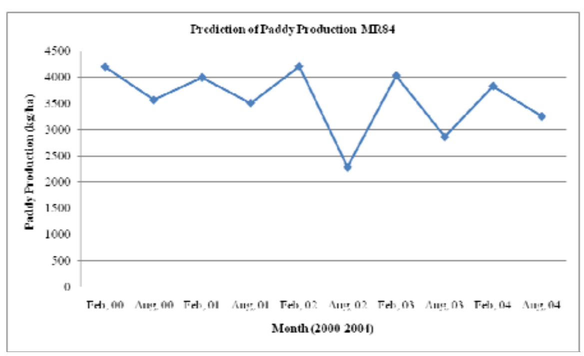


Figure-8. Graph of predicted paddy production produced twice per year in 5 years duration.

CONCLUSIONS

Results obtained exhibited a high accuracy between the cloned Abs and the actual data as evaluated by the RMSE, MAPE and R^2 tests. This study has achieved its goal to explore the capability of the proposed CSA in prediction of paddy production for type MR 84 variety. Throughout this work, it is proven that CSA is capable to cater with chaotic time series data such as climate data and data on paddy yield. The data on paddy yield used represent the behaviour of such crops cultivated in Malaysia. Prediction was performed for the year 2000-2004 for paddy type MR 84 variety, since the planting of MR 84 variety paddy has stopped in year 2000. Future prediction will be performed for paddy type MR 219 variety, since MR 219 paddy type was planted from the year 2000 onwards.

ACKNOWLEDGEMENTS

Authors thank Dato' Dr. Mohamad Zabawi Abd Ghani and Miss Shaidatul Azdawiyah Abd Talib from Malaysian Agricultural Research and Development Institute (MARDI) for their assistance.

REFERENCES

Al-Enezi, J.R., Abbod, M.F. and Alsharhan, S. (2010). Artificial Immune Systems -Models, Algorithms and Applications. International Journal of Research and Reviews in Applied Sciences, 3(2), pp.118-131.

Castro, L.N. and Timmis, J. (2002). Artificial Immune Systems: A New Computational Intelligence Approach. Canterbury: Springer.

Chen, B. and Zang, C. (2009). Artificial immune pattern recognition for structure damage classification. Computers and Structures, 87(21-22), pp.1394–1407.

Cheng, L., Ding, Y., Hao, K. and Hu, Y. (2012). An ensemble kernel classifier with immune clonal selection algorithm for automatic discriminant of primary open-angle glaucoma. Neurocomputing, 83, pp.1-11.
Food and Agriculture Organization of the United Nations. (2014).

Greensmith, J., Whitbrook, A. and Aickelin, U. (2010). Artificial Immune Systems.
Intergovernmental Panel on Climate Change Glossary. (2007).

Iqbal, S. (2007). Higher Education Commission Pakistan.
Lakshmi, K. and Vasantharathna, S. (2014). Gencos wind–thermal scheduling problem using Artificial Immune System algorithm. Electrical Power and Energy Systems , 54, pp.112-122.

Malaysian Agricultural Research and Development Institute. (2013).

Mohamed Azwan, M.Z., Mustapha, S. and Puasa, Z. (2010). Determination of Water Requirement in a Paddy Field at Seberang Perak Rice Cultivation Area. Journal of the Institution of Engineers, Malaysia, 71(4), pp.32-41.

Nanda, S.J. (2009). Artificial Immune Systems: Principle, Algorithms and Applications. Rourkela, Odisha, India.

Perez, P., Ardlie, N., Kuneepong, P., Dietrich, C. and Merritt, W.S. (2002). CATCHCROP: modeling crop yield and water demand for integrated catchment assessment in Northern Thailand. Environmental Modeling & Software, 17(3), pp.251-259.

Price, C., Ramsden, S., Hope, T., Friston, K. and Seghier, M. (2013). Predicting IQ change from brain structure: A cross-validation study. Developmental Cognitive Neuroscience , pp.172-184.

Ramirez, A. (2010). The Impact of Climate Change on Rice Production. Austin.

Subash, N. and Mohan, H. R. (2012). Evaluation of the impact of climatic trends and variability in rice–wheat system productivity using cropping system model DSSAT over the Indo-Gangetic Plains of India. Agricultural and Forest Meteorology, 164, pp.71-81.

Timmis, J., Hone, A., Stibor, T. and Clark, E. (2008). Theoretical Advances in Artificial Immune Systems. Theoretical Computer Science, 403(1), pp.11-32.

Timsina, J. and Humphreys, E. (2006). Performance of CERES-Rice and CERES-Wheat models in rice–wheat systems: A review. Agricultural Systems, 90(1–3), pp.5-31.

Zhong, Y., Zhang, L., Gong, J. and Li, P. (2007). A Supervised Artificial Immune Classifier for Remote-Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing, 45(12), pp.3957-3966.