

# FOOD INSECURITY IN AFRICA: AGRICULTURAL DIVERSIFICATION AS A PANACEA



1st

Proceedings of

## International Conference of Agriculture and Agricultural Technology {ICAAT 2019}

**VENUE:** Federal University of Technology, Minna  
School of Agriculture and Agricultural Technology

**DATE:** 23rd - 26th April, 2019



### EDITED BY

- ❖ Dr. C. E. Chinma
- ❖ Mr. A. Ibrahim
- ❖ Prof. J. N. Nmadu
- ❖ Dr. O. J. Ajayi

- ❖ Prof. S. O. E. Sadiku
- ❖ Prof. R. S. Olaleye
- ❖ Prof. A. S. Gana
- ❖ Dr. M. A. Ojo

- ❖ Prof. A. Aremu
- ❖ Dr. A. J. Jirgi
- ❖ Dr. B. A. Alimi
- ❖ Dr. S. O. Olayimika

## ARTIFICIAL IMMUNE SYSTEM (AIS) ALGORITHMS FOR CROPS CLASSIFICATION USING PRINCIPAL COMPONENTS ANALYSIS

\*<sup>1</sup> Akinwande, O.T and <sup>2</sup> Abdullahi, M.B

Department of Computer Science, Federal University of Technology, Minna, Niger State, Nigeria

\*Email of Corresponding Author: [oladayoakinwande@yahoo.com](mailto:oladayoakinwande@yahoo.com) GSM: 07039146008

### ABSTRACT

*There have been tremendous increase in crop production data which can be used to characterize and predict models in data mining for agriculture. Recently, researchers have shown a lot of interest in applying biologically inspired systems for solving classification and recognition problems. Several solutions have been proposed using Artificial Immune System (AIS), Ant Colony Optimization and so forth in classification problems as another machine learning technique. The field of agriculture is not left behind in the use of machine learning technique for crop and soil classification but few research has been carried out in using AIS as a machine learning technique for crop edibility and disease classification. In this paper, we propose an Artificial Immune System (AIS) solution using AIRS, Clonal and Immunos algorithms with PCA for crop edibility and crop disease classification. The proposed solution is tested on two crop dataset (Mushroom and Soybeans dataset). The results show significant improvement of the proposed solution over other techniques in most of the cases. Accuracy, true positives and false positives were used as performance measures. The proposed model can be used to enhance crop productivity.*

**KEYWORDS:** Expert System, Artificial Immune System, Feature Extraction, Principal Components Analysis

---

### INTRODUCTION

The vast amount of data in the agricultural sector has made the application of machine learning technique in agriculture to be on the rise. Accurate prediction of crop production, classification of crop type and diseases detection in crops can be better analyzed with the use of machine learning techniques which are brought together by computer science and statistics.

Machine learning involves adaptive mechanism that enable computers to learn from experience, learn by example and learn by analogy. In machine learning techniques, knowledge is giving to the machine to learn some certain features in order to classify or make predictions that can be used for decisions. Supervised and unsupervised are two types of machine learning techniques (Subhdra et al ,2016). Over time, learning capabilities can improve the performance of an intelligent system. Neural networks, Bayesian networks and AIS are one of the popular approaches to machine learning.

In order to verify the effectiveness of different AIS classifiers algorithms on the combination of principal component analysis (PCA) as a preprocessing method for the crops production, we used mushroom and Soybeans dataset to make relevant experiments step-by-step. Firstly, we build the experiment evaluation environment with major steps: environment setup, data preprocessing, choosing the data mining software. Secondly, we select the three

of most popular AIS classifier algorithms. An overview of how specific values of these algorithms were identified as well as their classification performance is given. Finally, we come up with the performance comparison between the 3 selected classifiers with mushroom dataset and soybean dataset. The outcome of this AIS based mining model can be utilized for decision support in improving agricultural crops productivity.

The paper is organized as follows. The section 2 discussed some research work in the use of machine learning for crop classification. In section 3, the AIS algorithms are explained. Section 4 gives the description of the experiments and the methods used. The results obtained are discussed in section 5. Conclusion and future work is given in section 6.

### METHODOLOGY

As Classification problems is important in computing so it is in agriculture. For this work, we use Soybeans and Mushroom dataset. All testing are done on a Personal Computer (PC) Windows 8.1 Machine (Intel Pentium 2.4GHZ, 6 GB RAM). The software used for the evaluation was WEKA (Waikato Environment for Knowledge Analysis) and WekaClassalgos, an open source machine Learning workbench for artificial immune systems algorithms.

WEKA has an extensive collection of pre-processing methods and Machine learning algorithms implemented in Java as classes with an optional graphical user interface. WEKA Version 3.7.8 and WekaClassAlgos 1.8 were used in this study.

### A. Artificial Immune System

The AIRS algorithm was one of the first AIS technique that is designed to tackle classification and recognition problems. The Artificial Immune Recognition System belongs to the field of Artificial Immune Systems, and more broadly to the field of Computational Intelligence. It was extended early to the canonical version called the Artificial Immune Recognition System 2 (AIRS2) and provides the basis for extensions such as the Parallel Artificial Immune Recognition System. It is related to other Artificial Immune System algorithms such as the Dendritic Cell Algorithm, the Clonal Selection Algorithm, and the Negative Selection Algorithm (Jason, 2012). The information processing objective of the technique is to prepare a set of real-valued vectors to classify patterns. The function of the AIRS algorithm is to prepare a pool of recognition or memory cells (data exemplars) which are representative of the training data the model is exposed to, and is suitable for classifying unseen data. In AIRS, clonal expansion and affinity maturation are used to encourage the generation of potential memory cells which are later used for classification. Hypothetically, AIRS has four stages to learning which are initialization, memory cell identification, resource competition and finally; refinement of established memory cells. The original AIRS1 algorithm uses a user defined *mutate rate* parameter to determine the degree to mutate a produced clone, and simply replaced attribute values with randomly generated values within the attributes normalised range. AIRS2 introduced the concept of somatic hyper mutation where the amount of mutation a clone receives is proportional to its affinity to the antigen in question (Jason,2012 ). Another important difference between AIRS1 and AIRS2 is the manner in which clones are mutated.

### 3.2 Clonal Selection Algorithm

The clonal selection algorithm (CLONALG) is actuated from the clonal selection theory. It is applied to optimization and pattern recognition problem. The clonal selection theory credited to Burnet was

proposed to account for the behavior and capabilities of antibodies in the acquired immune system (). Inspired itself by the principles of Darwinian natural selection theory of evolution, the theory proposes that antigens select-for lymphocytes (both B and T-cells).(Jason,2012) The information processing principles of the clonal selection theory describe a general learning strategy. This strategy involves a population of adaptive information units (each representing a problem-solution or component) subjected to a competitive processes for selection, which together with the resultant duplication and variation ultimately improves the adaptive fit of the information units to their environment.

### 3.3 Immunos

IMMUNOS1 algorithm is an artificial immune system based algorithm which assumes no data reduction, thus the clone population prepared is maintained and is used to classify unknown data instances. The artificial immune network algorithms includes the base version and the extension for optimization problems called the optimization artificial immune network algorithm (Jason, 2005). The Immunos-2 implementation is same as Immunos-1, the only difference is it seeks to provide some form of basic generalization via data reduction, and thus a closer representation to the original Immunos-81 proposal (Andrew and Jon, 2002 ).

### A. Data Set

Dataset was selected from the (UCI) machine learning repository (<https://archive.ics.uci.edu/>), specific to agricultural crops, the soybean [25] and mushroom [26] datasets. The soybean dataset contains 35 categorical features (See Table 5) and a class while the mushroom datasets contains 22 nominal features and a class (See Table 6). The mushroom data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the agaricus and lepiota family. It has information for 8124 mushroom transactions in which 4208 (51.8%) are edible and 3916 (48.2%) are poisonous. In Soybean data set, the values for attributes are encoded numerically with the first value encoded as "0" and the second as "1".

### B. Data Preprocessing

Pre-processing is carried out in order to remove noise or useless data that will not contribute to the accuracy of the classifier. Irrelevant or redundant feature can



result to high computational cost, high memory usage and reduced performance on the accuracy of the classifier.

### Normalization using the Min-Max Scaling

A suitable normalization technique is necessary to reduce the domination of features with higher values over features with lesser values so that the detection model would not be biased towards features having higher values ( You et al 2006). In this work, normalization is carried out to transform values of all features in to a common specific range by using the Min-Max scaling technique. The formulae for the normalization is giving below

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

### C. Feature Extraction Procedure

Principal Component Analysis (PCA) was used in this paper. PCA analyzes the relationships among multivariable, seeks the principal components denoted as a linear combination, and explains the entire changes with several components. The aim of PCA is to make the effective explanations through dimension reduction using linear equations (Geraldin & Bobby,2014). Although p components are required to reproduce the total system variability, often much of this variability can be accounted for by a small number, k, of the principal components. If so, there is almost as much information in the k components as there is in the original p variables. The k principal components can then replace the initial p variables, and the original data set, consisting of n measurements on p variables, is reduced to one consisting of n measurements on k principal components. The most common definition of PCA, is that, for a set of observed vectors  $\{v_i\}; i \in \{1, \dots, N\}$ , the q principal axes  $\{w_j\}; j \in \{1, \dots, q\}$  are those orthonormal axes onto which the retained variance under projection is maximal. It can be shown that the vectors  $w_j$  are given by the q dominant eigenvectors (i.e. those with largest associated eigenvalues) of the covariance matrix  $\Sigma$ .

## RESULTS AND DISCUSSION

This section presents the classification performance of AIRS1, ClonalG and Immnos1 on the dataset used. 10-fold cross validation is applied on both dataset. 10-fold cross validation process divides the dataset

into 10 part, nine parts will be used as training data and one part is used as data for testing. The results were identified for the algorithms using, TP, FP, accuracy, precision and recall. Table 1 shows the TP, FP, Accuracy, precision and recall of the Soybean without the combination of PCA. Table 2 shows Accuracy, Time taken, Precision, Recall, F-Measure and RMSE of the Soyabean dataset with the combination of PCA. Table 3 shows Accuracy, Time taken, Precision, Recall, F-Measure and RMSE of the Mushroom dataset with no combination of PCA. Table 4 shows Accuracy, Time taken, Precision, Recall, F-Measure and RMSE of the Mushroom dataset with the combination of PCA. The accuracy performance of AIRS1, Immunos1 and ClonalG has been compared to each other and this is shown in Figure 1.

AIRS1 performed best among all the other AIS algorithms in the two datasets. AIRS2 without PCA for the two dataset didn't a yield an optimal solution in a reasonable time . We compared our results with proven classifiers that have performed well on intrusion detection system. We choose ZeroR and J48 classifiers to compare with these AIS algorithms.

## CONCLUSION AND RECOMMENDATIONS

This paper provides a general baseline comparison for three popular AIS algorithms, namely AIRS1, ClonalG and Immunos1 and Immunos2 on crop dataset. Effectiveness of Mushroom edibility and Soybeans diseases detection using AIS algorithms was comparatively evaluated and the results were presented. Experimental results suggested that AIRS1 are effective and powerful for mushroom edibility and Soyabean disease detection achieved higher classification accuracies. Future work can include other AIS algorithms such as Dendritic cell algorithms (DCA) and Negative Selection Algorithms to test their performance in classifying crops for decision making.

Table 1: Soybeans without PCA

Attack Types	TP	FP	Accuracy	Time Taken	Precision	Recall	F-Measure	RMSE
CSCA	0.87	0.025	79.94	3.65	0.753	0.799	0.768	0.1455
ClonalG	0.552	0.049	55.19	0.43	0.547	0.552	0.528	0.201
AIRS1	0.813	0.021	81.25	1.23	0.756	0.813	0.778	0.141
AIRS2								
Immunos1	0.79	0.02	79.79	0	0.77	0.798	0.768	0.0213
Immunos2	0.32	0.105	31.77	0	0.32	0.32	0.187	0.268

Table 2: Soybeans with PCA

Attack Types	TP	FP	Accuracy	Time Taken	Precision	Recall	F-Measure	RMSE
CSCA	0.87	0.073	86.96	2.38	0.883	0.87	0.869	0.014
ClonalG	0.54	0.057	54.02	0.38	0.6	0.54	0.523	0.013
AIRS1	0.864	0.018	86.38	2.13	0.87	0.864	0.861	0.1197
AIRS2	0.837	0.02	83.74	0.5	0.842	0.861	0.836	0.02
Immunos1	0.861	0.01	86.09	0.06	0.902	0.483	0.867	0.143
Immunos2	0.483	0.08	48.31	0.03	0.394	0.816	0.365	0.2312

Table 3: Mushroom without PCA

Attack Types	TP	FP	Accuracy	Time Taken	Precision	Recall	F-Measure	RMSE
CSCA	0.846	0.162	84.61	494.64	0.861	0.846	0.844	0.3923
ClonalG	0.77	0.244	76.76	2.17	0.795	0.77	0.76	0.481
AIRS1	0.9	0.103	90	29.64	0.909	0.9	0.899	0.3161
AIRS2								
Immunos1	0.782	0.232	78.1	0.01	0.826	0.782	0.772	0.467
Immunos2	0.518	0.518	51.79	0.02	0.268	0.518	0.353	0.6943

Table 4: Mushroom with PCA

Attack Types	TP	FP	Accuracy	Time Taken	Precision	Recall	F-Measure	RMSE
CSCA	0.997	0.003	99.74	1795.03	0.997	0.997	0.997	0.0508

<b>ClonalG</b>	0.842	0.163	84.24	6.86	0.85	0.84	0.841	0.39
<b>AIRS1</b>	0.97	0.027	97.36	64.11	0.97	0.97	0.97	0.213
<b>AIRS2</b>	0.996	0.004	99.64	135.63	0.996	0.996	0.996	0.019
<b>Immunos1</b>	0.91	0.09	91.4	0.06	0.924	0.914	0.913	0.174
<b>Immunos2</b>	0.89	0.11	89.18	0.09	0.91	0.89	0.89	0.3289

Table 4 shows Accuracy, Time taken, Precision, Recall, F-Measure and RMSE of the Mushroom dataset with the combination of PCA.

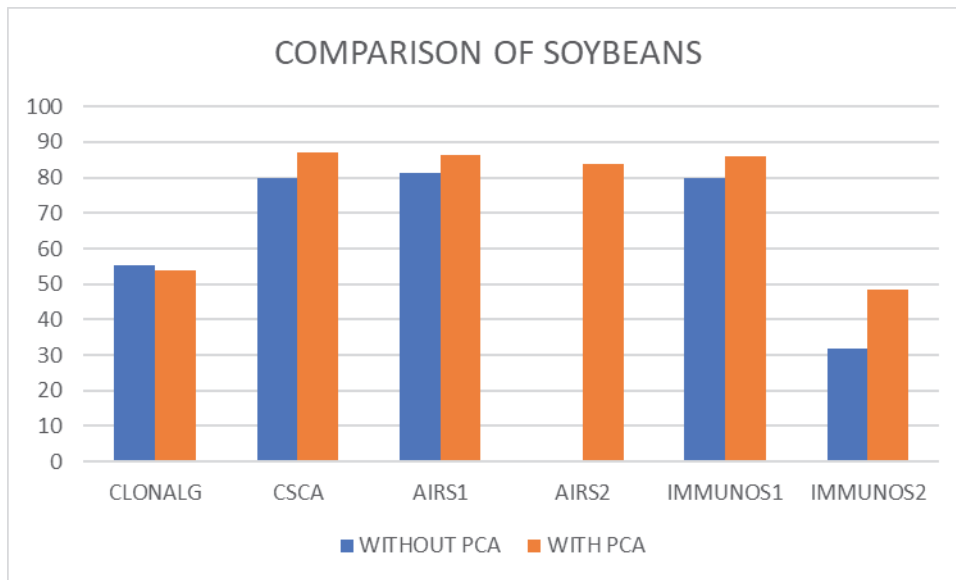


Fig. 1 Accuracy Classification for SoyBean with or without PCA

**Table 5 Mushroom Dataset**

Features and their Attributes
1. cap-shape: bell=b,conical=c,convex=x,flat=f, knobbed=k,sunken=s
2. cap-surface: fibrous=f,grooves=g,scaly=y,smooth=s
3. cap-color: brown=n,buff=b,cinnamon=c,gray=g,green=r, pink=p,purple=u,red=e,white=w,yellow=y

4. bruises?: bruises=t,no=f
5. odor: almond=a,anise=l,creosote=c,fishy=y,foul=f, musty=m,none=n,pungent=p,spicy=s
6. gill-attachment: attached=a,descending=d,free=f,notched=n
7. gill-spacing: close=c,crowded=w,distant=d
8. gill-size: broad=b,narrow=n
9. gill-color: black=k,brown=n,buff=b,chocolate=h,gray=g, green=r,orange=o,pink=p,purple=u,red=e, white=w,yellow=y
10. stalk-shape: enlarging=e,tapering=t
11. stalk-root: bulbous=b,club=c,cup=u,equal=e, rhizomorphs=z,rooted=r,missing=?
12. stalk-surface-above-ring: fibrous=f,scaly=y,silky=k,smooth=s
13. stalk-surface-below-ring: fibrous=f,scaly=y,silky=k,smooth=s
14. stalk-color-above-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
15. stalk-color-below-ring: brown=n,buff=b,cinnamon=c,gray=g,orange=o, pink=p,red=e,white=w,yellow=y
16. veil-type: partial=p,universal=u
17. veil-color: brown=n,orange=o,white=w,yellow=y
18. ring-number: none=n,one=o,two=t
19. ring-type: cobwebby=c,evanescent=e,flaring=f,large=l, none=n,pendant=p,sheathing=s,zone=z
20. spore-print-color: black=k,brown=n,buff=b,chocolate=h,green=r, orange=o,purple=u,white=w,yellow=y
21. population: abundant=a,clustered=c,numerous=n, scattered=s,several=v,solitary=y
22. habitat: grasses=g,leaves=l,meadows=m,paths=p, urban=u,waste=w,woods=d
<b>23. Class : edible or poisonous</b>

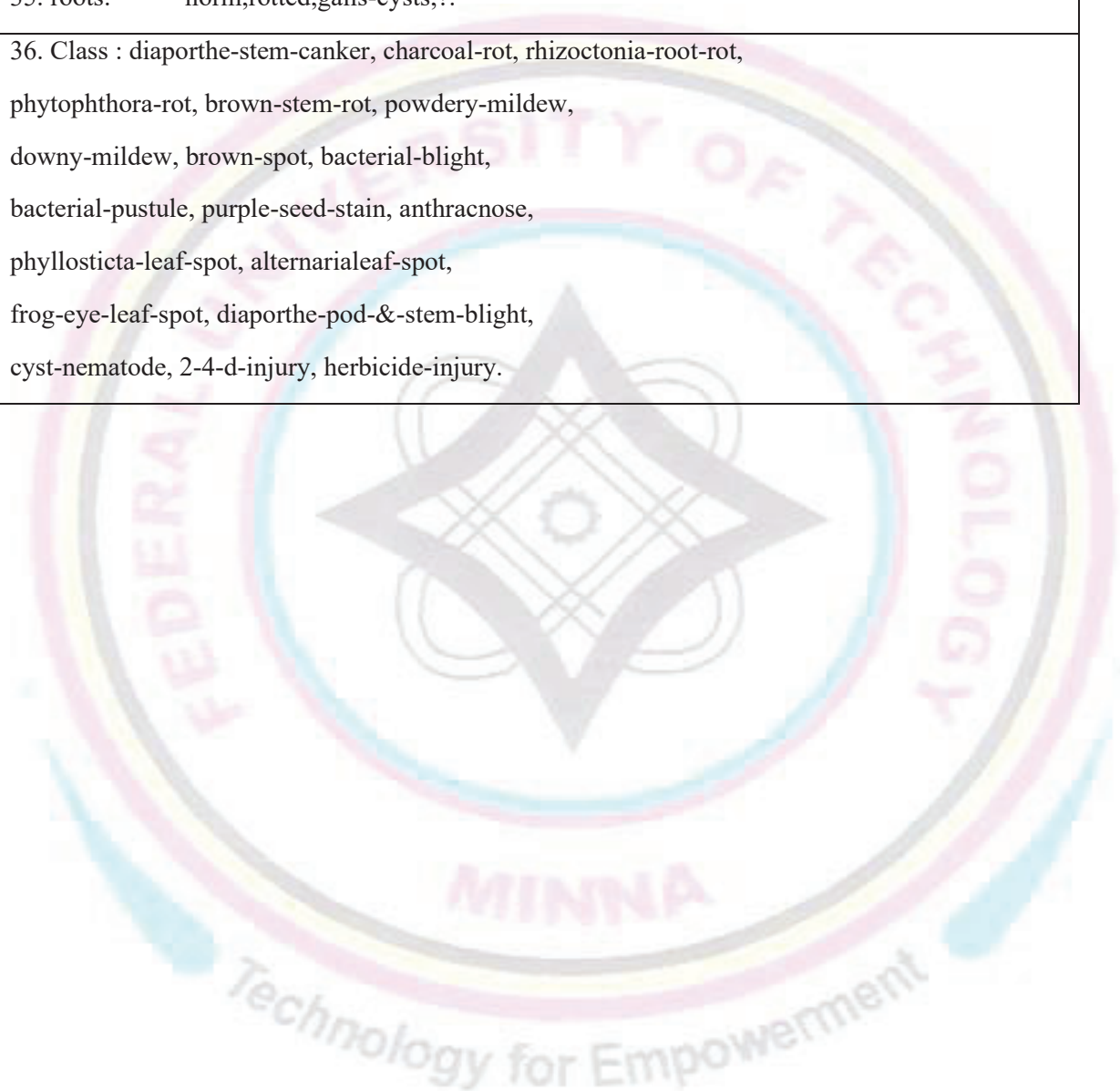
**Table 6 Soybean Data set**

Features and their Attributes	
1. date:	april,may,june,july,august,465ctober465r,465ctober,?.
2. plant-stand:	normal,lt-normal,?.
3. precip:	lt-norm,norm,gt-norm,?.

4. temp:	lt-norm,norm,gt-norm,?.
5. hail:	yes,no,?.
6. crop-hist:	diff-1st-year,same-1st-yr,same-1st-two-yrs, same-1st-sev-yrs,?.
7. area-damaged:	scattered,low-areas,upper-areas,whole-field,?.
8. severity:	minor,pot-severe,severe,?.
9. seed-tmt:	none,fungicide,other,?.
10. germination:	90-100%,80-89%,lt-80%,?.
11. plant-growth:	norm,abnorm,?.
12. leaves:	norm,abnorm.
13. leafspots-halo:	absent,yellow-halos,no-yellow-halos,?.
14. leafspots-marg:	w-s-marg,no-w-s-marg,dna,?.
15. leafspot-size:	lt-1/8,gt-1/8,dna,?.
16. leaf-shread:	absent,present,?.
17. leaf-malf:	absent,present,?.
18. leaf-mild:	absent,upper-surf,lower-surf,?.
19. stem:	norm,abnorm,?.
20. lodging:	yes,no,?.
21. stem-cankers:	absent,below-soil,above-soil,above-sec-nde,?.
22. canker-lesion:	dna,brown,dk-brown-blk,tan,?.
23. fruiting-bodies:	absent,present,?.
24. external decay:	absent,firm-and-dry,watery,?.
25. mycelium:	absent,present,?.
26. int-discolor:	none,brown,black,?.
27. sclerotia:	absent,present,?.
28. fruit-pods:	norm,diseased,few-present,dna,?.
29. fruit spots:	absent,colored,brown-w/blk-specks,distort,dna,?.
30. seed:	norm,abnorm,?.



31. mold-growth:	absent,present,?.
32. seed-discolor:	absent,present,?.
33. seed-size:	norm,lt-norm,?.
34. shriveling:	absent,present,?.
35. roots:	norm,rotted,galls-cysts,?.
36. Class :	diaporthe-stem-canker, charcoal-rot, rhizoctonia-root-rot, phytophthora-rot, brown-stem-rot, powdery-mildew, downy-mildew, brown-spot, bacterial-blight, bacterial-pustule, purple-seed-stain, anthracnose, phyllosticta-leaf-spot, alternarialeaf-spot, frog-eye-leaf-spot, diaporthe-pod-&-stem-blight, cyst-nematode, 2-4-d-injury, herbicide-injury.



## REFERENCES

- Andrew Watkins & Jon Timmis (2002). Artificial Immune Recognition System (AIRS): Revisions and Refinements
- R. Porkodi (2012). Comparison of Filter Based feature Selection algorithms: An overview
- Jason Brownlee. Clonal Selection theory and ClonalG: The Clonal Selection Classification Algorithm, Technical Report No, 2-02, January (2005)
- Jason Brownlee. 'Clever Algorithms Nature-Inspired Programming Recipes', Revision 2, 16 June 2012 ISBN: 978-1-4467-8506-5.
- Jason Brownlee (2005). 'Artificial Immune Recognition System (AIRS) A Review and Analysis', January 2005, Technical Report No. 1-02.
- Soybean (Large) Data Set. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Soybean+%28Large%29>
- Mushroom Data Set. (Online). Available: <https://archive.ics.uci.edu/ml/datasets/Mushroom>
- Sanghmitra Dash, Rabindra Kishore Mishra, Rama Krusha Das and Manisha Panda (2016). Comparison of Classifier Outputs using AIS and Non-AIS Based Data Mining Algorithms
- Felix T.S Chan, Anuj Prakash, R.K Tibrewal and M.K. Tiwari (2013). Clonal Selection Approach for Network Intrusion Detection
- Subhadra Mishra, Debahuti Mishra and Gour Hari Santra (2016). Applications of Machine Learning Techniques in Agricultural Crop Production: A Review Paper
- Ayodele Lasisi, Rozaida Ghazali, Tutut Herawan, Fola Lasisi, Mustafa Mat Deris(2015). Knowledge Extraction of Agricultural Data using Artificial Immune System
- Jihen Amara, Bassem Bouaziz, Alsayed Algergawy(2017). A deep learning-based Approach for Banana Leaf Diseases Classification
- Agung Wibowo, Yuri Rahayu, Andi Riyanto, Taufik Hidayatulloh(2018). Classification Algorithm for Edible Mushroom identification
- S.Hari Ganesh, Pritty Cindrella, Joy Christy(2015). A review on Classification Techniques over Agricultural Data
- Geraldin B. Cruz, Bobby D. Gerardo and Bartolome T. Tanguilig(2014). Agricultural Crops Classification Models Based on PCA-GA Implementation in Data Mining
- Surabhi Chouhan, Divakar Singh, Anju Singh (2016). A Survey and Analysis of Various Agricultural Classification Techniques
- Geraldin B. Cruz(2015). Comparative Study of Data Mining Classification Techniques over Soybean Disease by Implementing PCA-GA