OPEN ACCESS

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-6, March 2021

Crop Yield Prediction using Granular SVM

G. Sudha Sadasivam

Abstract: Agriculture is the backbone of the Indian economy. Farming is a major source of income for many people in developing countries. Prediction of yield of crops is desirable as it can predict the income and minimise losses for the farmersunder unfavorable conditions. But predicting crop yield is a challenging task in developing countries like India. Conventionally, crop yield prediction is done using farmer's expertise. The sustainability and productivity of a crop growing area are dependent on suitable climatic, soil, and biological conditions. So, data mining techniques based on neural networks, Neuro-Fuzzy Inference Systems, Fuzzy Logic, SMO, and Multi Linear Regression can be used for prediction. Previous work has performed yield prediction based on crop models considering only some of the environmental factors. This work uses a Support Vector Machine (SVM) to predict the crop yield under different environmental conditions that include soil, climate, and biological factors. Applying granular computing enables dividing the problem space into a sequence of subtasks. So, the hyperplane construction of SVM can be parallelized by splitting the problem space. Testing can also be parallelized. The main advantage is that linear SVM can be used to handle higher dimension space. Time complexity is reduced. Prediction using granular SVM can be parallelized using appropriate techniques like MapReduce/GPGPU. IoT-based agriculture increases crop yield by accurate prediction, automation, remote monitoring, and reducing wastage of resources. IoT-based monitoring systems can be used by farmers, researchers, and government officials to analyze crop environments and statistical information to predict crop yield. This paper proposes an IoT-based system to predict crop yield based on climatic, soil, and biological factors using parallelized granular support vector machines.

Keywords: Yield Prediction, SMO, Granular Support Vector Machines, MapReduce, GPGPU, IoT, Automation, Remote monitoring.

I. INTRODUCTION

Agriculture is the spine of the Indian economy. 69% of the Indian population has agriculture as their main occupation or side business. Predicting crop yield is a challenging and desirable task for decision-makers in developing countries like India. Such predictions can minimize losses under unfavorable conditions and maximize gain under favorable conditions. Farmer's expertise plays a major role in predicting crop yield. The Internet of things (IoT) enables users to organize, obtain and consume information by connecting real-world objects. IoT enables various applications in the digital agriculture domain. IoT-based agriculture can predict crop yield. With the advent of IoT, wireless sensor networks and the web play an important role in the digital agriculture domain. A wireless sensor network is a network composed of a set of nodes integrating the functions of acquiring, processing, communicating. Once deployed, the nodes cooperate autonomously to collect and transmit data to a base station for monitoring and controlling a device. Precision agriculture can be defined as the art and science of using technology to improve crop production. This paper uses a basic IoT system to sense soil parameters, humidity, temperature, and sunshine. The sensed information is used as test data for prediction in granular SVM (GSVM) to predict crop yield. A lot of research has been carried out in applying data mining techniques for crop yield prediction.

Manuscript received on February 20, 2021.
Revised Manuscript received on February 24, 2021.
Manuscript published on March 30, 2021.

* Correspondence Author
Dr. G Sudha Sadasivam*, Department of CSE, PSG College of Technology, Coimbatore, India, Email: gs.cse@psgtech.ac.in

II. LITERATURE SURVEY

IoT frameworks and platforms are still immature for agriculture, but there is a trend now to apply IoT in the agricultural sector. Duan Yan-e et al [1] proposed an IoT application to predict fertilizer application to crops using WSN. Xiangyu Hu et al. [2] developed an IoT system for the efficient use of water resources. Agri-IoT [3] analyses and processes data coming from WSN exploiting the semantic aspects. Using the Bluetooth model for communication, has its own limitations like limited range and device accommodation. Use of IoT in agriculture is mentioned by an author in a paper [4]. However, it shows a lack of interoperability which is necessary to handle large agricultural fields. Suryadevara et al. have used concepts of pervasive computing, data aggregation, etc to monitor the environmental factors using Zigbee [5]. An increase in the number of sensors is suggested by the author to improve the accuracy of the data collected. However, it might raise the issue of more power consumption as more nodes have been deployed. Real-time information about crops can be provided to farmers. Concepts of IoT, cloud computing, mobile computing [6], and Phenonet [7], a network of smart wireless sensor nodes that shares information as well as a central system can be used in smart agriculture. Yet both the papers however do not provide any interpretation of the data even though a large amount of useful data is generated. Ramesh et al. [8] had proposed a K-means technique for clustering regions based on rainfall and multiple linear regression for predicting the rainfall. They considered the East Godavari district of Andhra Pradesh data set, which includes the parameters like year, rainfall, area of sowing, and production. The advantage of this paper is proper rainfall prediction and the disadvantage of this paper is the computational cost of the K-means algorithm. Manjula et al. [9] had proposed the K-means technique for soil classification, K-nearest neighbors for simulating climate variables, and support vector machines for analyzing the different possible changes of the climate scenario. They considered the soil and climate datasets, the soil dataset includes parameters like (N, P, PH, Mg, S, Fe, Zn, Mn, Cu) and the climate dataset includes parameters like rainfall, humidity, minimum temperature, maximum temperature. The advantage of this paper is exact crop yield prediction and the disadvantage of this paper is less number of data attributes. Monali Paul et al. [10] had proposed a K-nearest neighbor technique for soil classification. They considered the soil dataset, which includes parameters like N, P, PH, Mg, S, Fe, Zn, Mn. The advantage of this paper it helps farmers to decide the land for sowing that may result in better crop production and the disadvantage of this paper is the smaller dataset used. Raine A. A et al. [11] had proposed the K-means technique for detect atmosphere pollution and soil classification and support vector machines for weather classification. They considered weather and soil datasets. The weather dataset includes parameters like humidity, temperature, precipitation, air pressure, and the soil dataset includes parameters like(N, P, PH, Mg, S, Fe, Zn, Mn, Cu). The advantage of this paper is used to solve complex agriculture problems using K-means and support vector machines techniques and the disadvantage of this paper is K-means did not produce good results when compared to support vector machines. Niketa Gandhi et al. [12] had proposed support vector machines for creating functions from a set of labeled data. They considered climate & production datasets. The climate dataset includes parameters like rainfall, humidity, minimum temperature, maximum temperature, and the production dataset include crop production and yield. The advantage of this paper is that the SMO classifier provides good accuracy and sensitivity and the disadvantage is that the support vector machine provides lower accuracy when compared to SMO classifier. Rajeshekhar Borat et al. [13] had proposed a K-nearest neighbor for classifying various crops and linear regression for analyzing rainfall, humidity. They considered the climate dataset, it includes parameters like rainfall, temperature, humidity. This approach shows the annual yield for the user and the disadvantage is less number of attributes considered. B. Vishnu Vardhan et al. [14] had proposed density-based clustering techniques for clustering rainfall. They considered the East Godavari district of Andhra Pradesh dataset, which includes parameters like year, area of sowing, rainfall, production. This work considers a large volume of data but produces an approximate prediction. Narayanan Balakrishnan et al. [15] had proposed Support vector machines, Adaptive support vector machines, and Adaptive naïve Bayes for classification. They considered climate and crop production datasets. The Climate dataset includes parameters like rainfall, temperature, humidity, and the crop production dataset includes parameters like crop production and yield. Naushina Farheen et al. [16] had proposed a K-nearest neighbor to cluster and make predictions. They considered the weather dataset, it includes parameters like humidity, temperature, precipitation. The advantage is prediction is good but considers only one technique. A.T.M Shakil Ahamed et al. [17] had proposed a K-means technique for grouping the climate conditions, linear regression technique for determining the relationship between the dependent and independent variables, and a K-nearest neighbor technique for comparing given tests examples with similar training examples. Both climate and biological factors were considered. Veenadhari et al. [18] had proposed C4.5 algorithms to find out the most influencing climatic parameters on the crop yield of selected crops. They considered the climate dataset, it includes rainfall, minimum and maximum temperature, humidity. It addresses the problem in land grading but requires an internet connection. Raju Prasad Paswan et al. [19] had proposed linear regression techniques for predict crop yield based on the good quality of clusters and neural network techniques for rainfall forecasting. Aakunuri Manjula et al. [20] had proposed a spatial data mining technique with weather datasets to predict crop yield and a multi-linear principal component analysis technique for the feature reduction phase. Ashwani Kumar et al. [21] had proposed an Agro algorithm to get quality and improved crop yields. They considered the soil dataset, which includes parameters like (N, P, PH, Mg, S, Fe, Zn, Mn, Ca, Cu). Niketa Gandhi et al. [23] have proposed Bayesian networks techniques for predict rice crop yield for Maharashtra state. They considered climate and crop production datasets. Only Kharif season was considered to predict the crop yield. Haedong Lee et al. [24] had proposed Kernel smoothing techniques for predicting the apple yield using weather datasets. Ramesh et al. [25] have proposed multiple linear regression to model the linear relationship between a dependent variable and one or more independent variables. Density-Based Clustering Technique is used to group the parameters. They considered the East Godavari district of Andhra Pradesh. Aditya Shastry et al. [26] had proposed the Fuzzy logic technique ion and ANFIS technique for predicting the wheat yield by configuring its parameters. They used the Multiple linear regression techniques to model training data using linear predictor functions and the unknown coefficients are estimated. They considered Biomass, Extractable soil Water, Solar Radiation, rain datasets. The advantage of this paper is ANFIS model gives better accuracy than MLR and FL models with a lower RMSE value and the disadvantage of this paper is improper datasets are used. V.R.Thakare et al. [27] had proposed the Fuzzy logic technique for predicting the maximum yield from crops. They considered climate and soil datasets. Yuchun Tang Bo et al. [22] had proposed Granular support vector machines for protein homology prediction and Granular support vector machines-association rule for modeling granular support vector machines by building information granules. Granular Support Vector Machines divide the problem into multiple smaller problems, so errors were reduced and prediction accuracy increases. As the yield varies among cities, granular support vector machines have been proposed to predict yield.

III. SYSTEM ARCHITECTURE

A basic IoT system using an Arduino microcontroller (Aurdino Mega 2560) connects field sensors (fig. 1). The various steps include

A. **Data Collection**: Data regarding sunshine, moisture and soil parameters are collected using various sensors. DH22 Temperature/humidity sensor is used. An electrochemical sensor system is used for soil nutrient determination. The electrochemical sensor consists of two electrodes which respond to targeted ion and transforms the reactions to detectable electrical signals. An A/D converter is used to convert the analog values to digital before providing input to the microcontroller. Ion Selective Field Effect Transistor (ISFET) can be chemically modulated and the measured voltage is related to the concentration of the targeted ion. The ability to integrate multi -ISFETs on one chip tends to detect not only N, P, K but also other nutrients such as ammonium (NH4), manganese, cobalt, sulphur, iron, calcium, and others.

B. **WiFi module:** The microcontroller retrieves data from the sensors and transmits information through the WiFi module to the system for analysis. The ESP8266 WiFi module is a self contained SOC with an integrated TCP/IP protocol stack that facilitates the microcontroller to access the WiFi network.

C. **Cloud analysis:** The real time data is transferred to a trained GSVM classifier that can predict the crop yield.

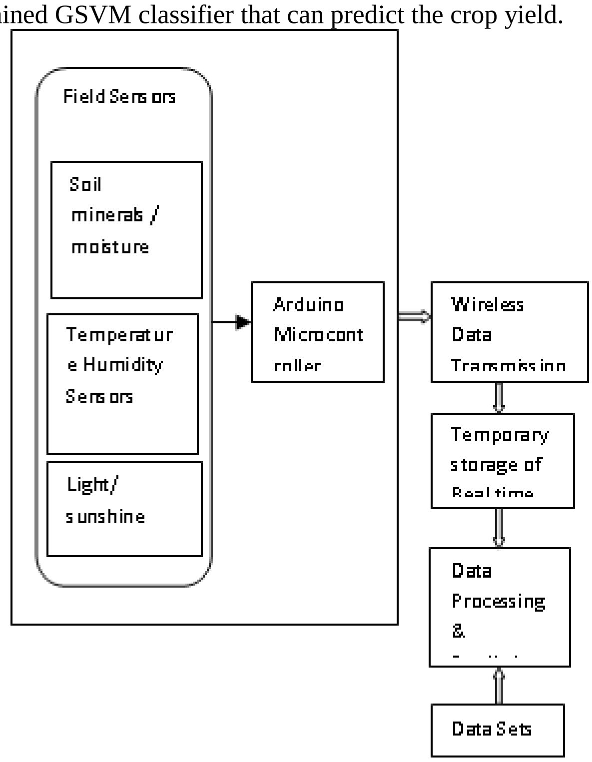

Fig.1 Overall System architecture

IV. GRANULAR SUPPORT VECTOR MACHINE (GSVM)

A Support Vector Machine (SVM) is a supervised learning method that can be used for classification and regression analysis. The standard SVM takes a set of input data and predicts its class membership. This makes the SVM a no probabilistic binary linear classifier. More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin). In general, the larger the margin the lower the generalization error of the classifier

For the given training set $T=\{(x_1,y_1),\ldots\ldots,(x_i,y_i)\} \in (X \times Y)^i$, where $x_i \in X \equiv R$ and $y_i \in Y \equiv \{-1,1\}$ $(i=1,2,\ldots.,l)$, a real-valued function g(x) on $X \equiv R^n$ was sort as the decision function f(x).

$$f(n)=\mathrm{sgn}(g(x)) \qquad (1)$$

Consider ξ_i as slack variable and C as penalty coefficient. The linearly separable training set min $1/2\ w^2 + C\sum_{i=1}^{l}\xi_i$ was obtained with the constraint in SVM

$$y_i((w.\ x_i)+b)+\xi_i \geq 1, i=1,\ldots,l \qquad (2)$$

The hyperplane is then constructed. The decision function is then obtained as

$$f(x)=\mathrm{sgn}((w^*\ .\ x)+b^*) \qquad (3)$$

However, SVM modeling is computationally expensive, quadratic to the number of samples and linear to the number of features. Granular SVM is more scalable.

Granular computing splits the problem into subspaces called information granules. The problem is then solved in each information granule. Being knowledge-oriented, prior knowledge is used to improve the classification. Granular computing involves,

- *Granular split*: to split a huge problem space into a sequence of granules.
- *Granule shrink:* that defines the granule size for a particular problem.

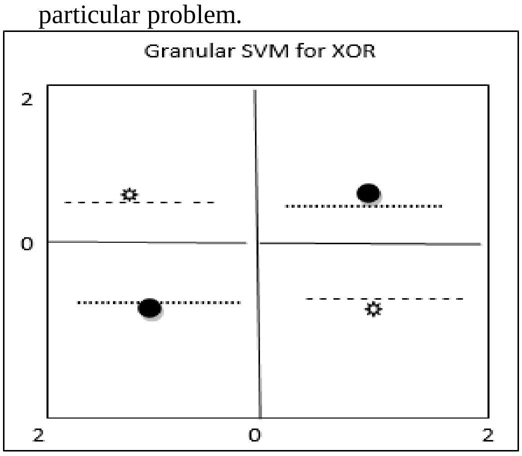

Figure 2: GSVM for XOR

Granular Support Vector Machines (GSVM) aims at integrating granular computing into statistical learning. GSVM extracts a sequence of information granules with granule split and/or granule shrink, and then builds SVMs on some of these granules. Prediction of GSVM is effective as it considers the local significance and global correlation among data subsets. GSVM speeds up the classification process by eliminating redundant data locally. It can convert a linear non-separable problem to a totally linear separable one. GSVM provides effective classification as it grasps inherent data distribution by a tradeoff between the local significance of a subset of data and global correlation among different subsets of data. For example, the XOR problem is not linearly separable. This is to be converted to higher dimensionality for linear separability.

Retrieval Number: 100.1/ijrte.F5417039621
DOI:10.35940/ijrte.F5417.039621
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

Instead, Granular computing can be used to separate the space into two granules (subspaces at x=0) and then construct an SVM for each granule as shown in figure 2.

The steps of GSVM modeling are as follows:

a) *Granulation*: Decision Trees, Association Rules, sampling, bagging, boosting and clustering algorithms, can be used to split the original feature space into a sequence of subspaces.

b) *Classification*: Multiple information granules created in the first step are classified by classifiers. Each information granule has an SVM classifier.

c) Aggregation: This information is then aggregated as data fusion, decision fusion, knowledge fusion, or hybrid information fusion.

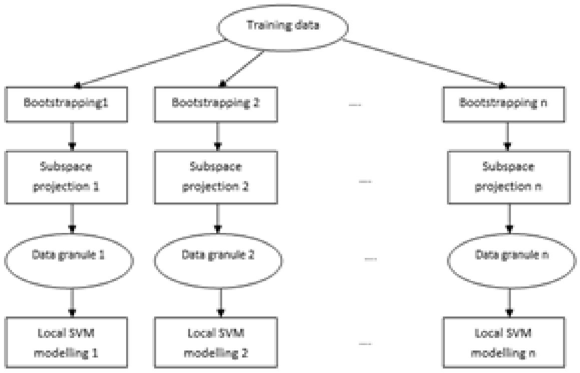

Fig 3: Training Phase of GVSM

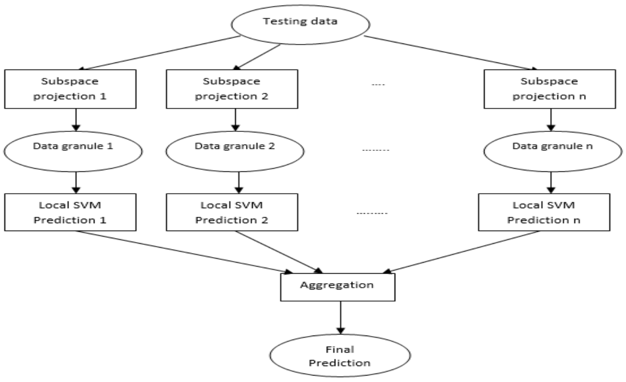

Fig 4: Testing Phase of GVSM

During the training phase (figure 3), GSVM extracts subsets of samples from the original training dataset. Each training subset forms a granule. Data from the granule is used to train an SVM. During testing (figure 4), the data is tested with each local SVM. Results of the SVMs are aggregated by the Bayesian Sum Rule to obtain the final decision. This is done by summing the probability estimates of each local SVM for final prediction. This approach is easily parallelizable.

V. RESULTS

The hardware we used is a PC with a P4-2.8MHz CPU and 256M memory. The software used is SVM Classifier Matlab Toolbox which implements a Mat lab interface to LIBSVM. The linear, polynomial, and radial basis kernel functions were used for SVM training, after which fivefold cross-validation was used for modeling and testing, respectively. For linear SVM, regulation parameter $C \equiv 1$; for RBF SVM, kernel parameter $\gamma \equiv 1$ and regulation parameter $C \equiv 1$.Parameters considered include PH, soil nutrients like nitrogen, phosphate, potassium, organic carbon, magnesium, sulphur, manganese, copper, iron, zinc; electrical conductivity, temperature, rainfall, humidity, sunshine, season, crop name, production.

Table 1: Performance comparison of GSVM with other SVM models

Crop Yield	Accuracy				Time(seconds)			
	Linear SVM	***Poly SVM***	***RBF SVM***	***GSVM***	***Linear SVM***	***Poly SVM***	***RBF SVM***	***GSVM***
Rice	0.96	0.93	0.97	0.98	419	430	461	3
Wheat	0.82	0.8	0.75	0.85	405	417	448	2
Maize	0.83	0.85	0.8	0.84	407	425	456	2

GSVM shows good accuracy along with time reduction. Future Extension is to parallelize the operations using Map Reduce programming.

VI. CONCLUSION

This paper proposes an IoT-based system (Arduino microcontroller - Aurdino Mega 2560) that retrieves data from the sensors and transmits information through the WiFi module to the system for analysis. Crop yield is predicted based on climatic, soil, and biological factors using granular support vector machines. To overcome the complexity in the dataset, the problem space is split into granules and GSVM was implemented. The overall accuracy and time for prediction has been improved using granular SVM. Experimental results demonstrate that the accuracy of GSVM is as good as a linear SVM, but time complexity is reduced.

REFERENCES

1. Duan Yan-e, "Design of Intelligent Agriculture Management Information System Based on Io", Fourth International Conference on Intelligent Computation Technology and Automation 2011, vol. 1 pp.1045 – 1049.
2. X. Hu and S. Qian, "IoT application system with crop growth models in facility agriculture", *6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)*, Seogwipo, Korea (South), 2011, pp. 129-133.
3. Andreas Kamilaris, Feng Gao, Francesc X. Prenafeta-Boldú and Muhammad Intizar Ali., "Agri-IoT: A Semantic Framework for Internet of Things-enabled Smart Farming Applications", IEEE World Forum on the Internet of Things (WF-IoT), Reston, VA, USA, December 2016, pp. 442-447
4. Junyan Ma, Xingshe Zhaou, Shining Li, Zhigang Li, "Connecting Agriculture to the Internet of Things through Sensor Networks", IEEE International Conference, 2011, pp. 184-187
5. Sean Dieter Tebje Kelly, Nagender Kumar Suryadevara and Subhas Chandra Mukhopadhyay, "Towards the Implementation of IoT for Environmental Condition Monitoring in Homes", IEEE Sensors Journal,vol.13, no.10, October 2013, pp.3846-3853,
6. Hemlata Channe, Sukhesh Kothari, Dipali kadam,"Multidisciplinary model for Smart Agriculture using IoT, Sensors, Cloud-Computing, Mobile Computing & Big data Analysis", Int.J.Computer Technology & Applications,Vol 6 (3), 2018, pp. 374-382
7. Prem Prakash Jayaraman, Doug Palmer, Arkady Zaslavsky, Dimitrios Georgakopoulos, "Do-it-Yourself Digital Agriculture Applications with Semantically Enhanced IoT Platform", IEEE Tenth International Conference (ISSNIP) Singapore,7-9 April 2015, pp. 1-6.
8. D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol.2, Issue.9, September 2013, pp. 233-240.
9. E.Manjula, S.Djodiltachoumy," Analysis of Data Mining Techniques for Agriculture Data", International Journal of Computer Science and Engineering Communications, Vol.4, Issue.2, 2016, pp.1311-1313.

10. D Ramesh, B Vishnu Vardhan, "Analysis Of Crop Yield Prediction Using Data Mining Techniques", International Journal of Research in Engineering and Technology, vol.04,Issue:01, Jan 2015, pp.470-474.
11. Raorane A.A, Kulkarni R.V, "Data Mining: An effective tool for yield estimation in the agricultural sector", International Journal of Emerging Trends &Technology in Computer Science(IJETTCS), vol.1, Issue 2, August 2012, pp. 220-225.
12. Rajshekhar Borat, Rahul Ombale, Sagar Ahire, Manoj Dhawade, P.S. Kulkarni, "Data Mining Technique to Predict Annual Yield for Major Crops", International Journal for Scientific Research & Development| vol. 4, Issue 03, 2016, pp.1835-1837.
13. B.VishnuVardhan, D. Ramesh "Density-Based Clustering Technique on Crop Yield Prediction", International Journal of Electronics and Electrical Engineering Vol. 2, No. 1, March 2014, pp.246-250.
14. N.Gandhi, Leisa J. Armstrong and OwaizPetkar, Amiya Kumar Tripathy, "Rice Crop Yield Prediction in India using Support Vector Machines", 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016, pp.1-5.
15. Yuchun Tang, Bo Jin, Yan-Qing Zhang, "Granular Support Vector machines with association rules mining for protein homology prediction", Artificial Intelligence in Medicine, vol. 35, Feb.2005, pp. 121-134.
16. Monali Paul, Santosh K. Vishwakarma, Ashok Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach", International Conference on Computational Intelligence and Communication Networks, 2015, pp. 766-771
17. Narayanan Balakrishnan and Dr.Govindarajan Muthukumarasamy, "Crop Production-Ensemble Machine Learning Model for Prediction", International Journal of Computer Science and Software Engineering, Volume 5, Issue 7, July 2016, pp.32-39.
18. Naushina Farheen, R. V. Argiddi, "Annual Crop Yield Prediction and Recommend Planting of different crops by Using Data Mining Techniques", International Journal of Innovative Research in Computer and Communication Engineering., vol. 4, Issue 10, October 2015, pp. 1-6.
19. A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman, "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Bangladesh", IEEE SNPD 2015, June 2015, pp.1-8.
20. S.Veenadhari**,** Dr. Bharat Misra, Dr. CD Singh, "Machine Learning Approaches for Forecasting Crop Yield based on Climatic Parameters", International Conference on Computer Communication and Informatics (ICCCI -2014), Jan. 03 – 05, 2014.
21. Raju Prasad Paswan, Shahin Ara Begum, "Regression and Neural Networks Model for Prediction of Crop Production", International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September 2013, pp.98-108.
22. Aakunuri Manjula1 and Dr. G.Narsimha, "Crop Yield Prediction with Aid of Optimal neural network in spatial data mining: new approaches", International Journal of Information & Computation Technology, vol. 6, no.1, 2016, pp.25-33..
23. Ashwani Kumar Kushwala, SwetaBhattachrya, "Crop yield prediction using Agro Algorithm in Hadoop", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 5, No2, April 2015, pp.271-274.
24. N.Gandhi, Leisa J. Armstrong and Owaiz Petkar, Amiya Kumar Tripathy, "Predicting Rice Crop Yield Using Bayesian Networks", Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 2016, pp.795-799.
25. Haedong Lee, Aekyung Moon, "Development of Yield Prediction System Based on Real-time Agricultural Meteorological Information", ICACT2014, 2014, pp. 1302-1305.
26. Aditya Shastry, Sanjay H A and Madhura Hegde, "A Parameter Based ANSIF Model for Crop Yield Prediction", IEEE International Advance Computing Conference (IACC), 2015, pp.253-257
27. V.R.Thakare, H.M.Baradkar , "Fuzzy System for Maximum Yield from Crops", International Journal of Applied Information Systems, 2013, pp. 4-9

AUTHORS PROFILE

Dr G Sudha Sadasivam works as a professor and heads the department of CSE at PSG College of Technology, Coimbatore. She has published 5 books and 30 papers in reputed journals. She has done established a laboratory on Big Data Analytics sponsored by Yahoo and Nokia. She has carried out government sponsored research projects under AICTE, UGC and DST in area of Distributed Systems, Big data analytics and assistive technologies. Her area of research work includes Big Data Analytics and Distributed systems.