The Seybold REPORT

ISSN 1533-9211

DOI 10.5281/zenodo.6622423

AN IOT BASED INTELLIGENT DECISION SUPPORT SYSTEM FOR SMART FARMING USING MACHINE LEARNING TECHNIQUES

M. SURESH,

Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India. Email: ssuresh.mg@gmail.com

Dr S. MANJU PRIYA

Professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India. Email: smanjupr@gmail.com

ABSTRACT

Farmers and agricultural executives rely on crop production and drought predictions to aid agriculture-affected regions all over the world. The prediction of Drought is crucial for early warning and reducing the effects of drought on agricultural yield. Drought forecasting research aims to enhance predictability skills and deepens the understanding of the physical mechanics of drought by combining all available predictability sources. In this paper, an intelligent decision support system for smart farming is proposed along with wrapper-based feature selection and support vector machine classifier is used for the productivity of crop and the prediction of the drought. The results are evaluated using three real-time datasets and compared with five existing prediction models namely K-Nearest Neighbour, Naïve Bayes, Back Propagation Neural Network and Wrapper based PART. The productivity of the crop's sorghum, jowar and sugarcane are used for the prediction and the Experiment outcome reveals that the proposed wrapper-based feature selection and wrapper-based classification is best suitable for drought prediction and productivity of crops.

Keywords: Machine Learning, Agriculture, Drought Forecasting, IoT in Agriculture, Modernization of Agriculture, Intelligent Farming.

1. INTRODUCTION

The worldwide dry zone covers nearly a third of the land surface. According to data, meteorological catastrophes accounted for roughly 85% of all losses caused by various natural disasters, with drought accounting for around 50% of total losses caused by meteorological disasters. Droughts frequently cause disasters in the occurrence area's social economy [1]. For example, from the late 1960s to the early 1970s, Africa suffered a severe drought that lasted for 6 years, and the deep wells in the area of 300square kilometres were dry and waterless. India’s arid zone covers an area of about100square kilometres, accounting for 1/3 of the country's land area. The annual national economic loss caused by drought is about 40 billion US dollars [2]. Most of the areas in Asia and Africa are a typical arid and semi-arid climate zone, with frequent droughts for every spring. With climate change, droughts in southern part of the country have also got from time to time and are more serious in recent years.

The Seybold REPORT
ISSN 1533-9211
DOI 10.5281/zenodo.6622423

The cause of drought has regional characteristics. Previous studies believe that the formation of drought can be attributed to the sinking branch of the subtropical Hadley circulation; the sinking movement of the leeward slope of the mountains (the rain shadow area on the leeward slope); and the source of water vapour away from the ocean. Less; abnormal atmospheric circulation, resulting in weak and few precipitation weather systems, etc. [3]. The study also pointed out that sea temperature will have an impact on uneven precipitation distribution [4]; Charney [5] proposed a soil moisture-albedo-precipitation positive feedback desert drought maintenance theoretical mechanism. Webb et al. [6] used atmospheric circulation models (General Circulation Mode, (GCM) to study the relationship between surface characteristics, soil moisture profile and soil water holding capacity, and pointed out the influence of different underlying surface characteristics on soil water holding capacity and soil moisture. Wang Chenghai et al. [7] proposed a work based on observational data from the National Center for Environmental Prediction, he evaluated the yearly fluctuation features of sensible heat and latent heat in a typical dry location in northern China, and found that the sensible heat fluxed in spring was rather substantial (NCEP). The features of sensible and latent heat flux in the dry area of Northwest China are revealed by reanalyzing the data. The annual average latent heat flow in the Loess Plateau region is bigger than the sensible heat flux as a result of global warming, and the latent heat flux has been dropping over the past 50 years while the sensible heat flux has been growing. The decrease in latent heat and increase in sensible heat mean that drought has a tendency to aggravate. Therefore, drought is an abnormality in the positive and negative feedback loops between precipitation, soil moisture, evapotranspiration, and ecosystems. Therefore, drought monitoring and prediction have great difficulties. Generally, drought monitoring and prediction are carried out by integrating multiple elements and multiple processes to construct drought indicators .

According to different academic viewpoints and different emphasis on drought investigation, droughts are usually divided into meteorological droughts, hydrological droughts, agricultural droughts, and other economic and social droughts [6-11]. Although different disciplines have different criteria for determining drought events, the basic understanding of the characteristics of drought is the same, that is, drought is regional, persistent and harmful[12-17].

The existing drought index can be divided into 5 categories:

Meteorological Drought Index;
Hydrological Drought Index;
Agricultural Drought Index;
Social Economic Drought Index;
Drought Index based on Remote Sensing.

Although the various drought indices mentioned above have their own advantages, they all involve in a variety of empirical parameters in the calculation process, and their determination is more complicated and highly dependent on regions. How to simplify the index without losing accuracy in actual business is crucial to the detection and prediction of drought. The main contributions of this work are,

The Seybold REPORT
ISSN 1533-9211
DOI 10.5281/zenodo.6622423

1. Need to build a smart agricultural system based on IoT that is efficient and use wrapper-based machine learning algorithms.
2. Need to design a farmer and farming staff interaction and decision support system which is based on cultivated crop and rainfall.
3. To predict the rainfall forecast that is drought prediction and crop productivity prediction for seasonal decision support to the farmers.

The rest of this work is structured as follows: The recommended system for agriculture field and drought prediction in IoT-based smart farming is described in Section 2. The findings analysis is presented in part 3, and the study is concluded in section 4 with future scope and referenced sources.

2. SMART FARMING USING IOT BASED DECISION SUPPORT SYSTEM

Establishing the linear correlation between the predictive factor and the predictive object is a common method of climate prediction. On this basis, the weather factor should be used to estimate the long-term weather conditions in the future. However, there is a complex and highly non-linear relationship between the predicted object and the cause, among which drought is a long-lasting and wide-range climatic abnormal event. Drought prediction is usually based on the estimation of future temperature, precipitation, etc., and then inferring the drought, the accuracy is greatly affected. To improve the accuracy, a large number of samples are required, and meteorological observations only have decades of data and it cannot meet its requirements. The prediction model established by the SVM method is to find the best solution for existing samples. This model recognition technology provides the possibility for drought prediction directly from the climate factors. This research intends to use the Support Vector Machine (SVM) model wrapper based recognition technology to predict the drought and productivity of the crops Sorghum, Jowar and Sugarcane in the country, and use relevant data to test the model, in order to provide an effective way of predicting the drought and predicting the productivity of the crops Sorghum, Jowar and Sugarcane. The proposed IoT based smart farming is achieved in 3 phases. As a first phase data are collected from the IoT sensors and farmers then pre-processed for further processing in phase two. Feature selection is performed as the third phase. In this step wrapper-based feature selection is performed for feature selection and SVM classifier is used for the prediction. As next phase, selected best features are feed further for the drought and productivity of the crop prediction. In this phase, dashboards are provided for the end users and farming staffs for the interaction and data interpretation analysis. The IoT based smart farming is shown in Figure 1 and wrapper-based feature selection and classification is shown in Figure 2.

ISSN 1533-9211
DOI 10.5281/zenodo.6622423

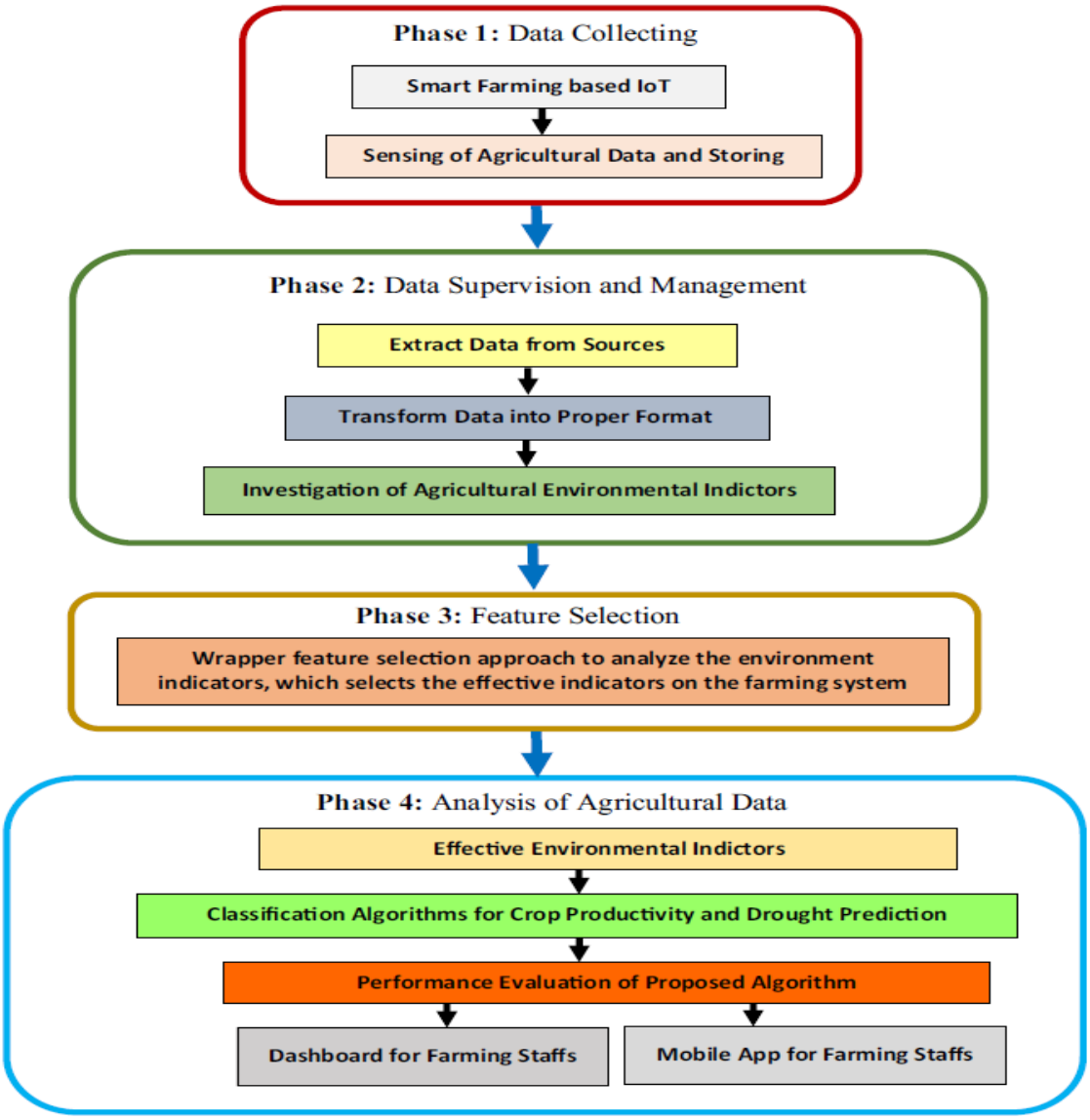

Fig 1. Smart Farming using IoT based Decision Support System [3]

2.1 INTELLIGENT DECISION SUPPORT SYSTEM

Machine learning is currently one of the most popular topics in computer theory and application research. Research is done to find rules for the development of observation data (samples) and applying these rules to predict future data for effective decision support. So far, there is no commonly accepted theoretical framework for machine learning. Machine learning based on traditional statistics requires a preliminary understanding of the distribution

ISSN 1533-9211
DOI 10.5281/zenodo.6622423

of samples. The function of training samples is to estimate the parameters of the distribution, this type of machine learning studies the asymptotic theory when the count of observations approaches to infinity, however in real world problems, the number of observations is often limited, so some theoretically excellent learning methods may not perform as expected in practical applications. Vapnik and others are committed to statistics. The study of learning theory Statistical Learning Theory or SLT, which establishes a new theoretical system and statistical reasoning rules for small sample statistics. It is taken into account not just because of the criteria for asymptotic performance, but also how to get the best outcomes with little information. The Support Vector Machine (SVM) is a statistical learning theory-based intelligent learning approach. The main idea is to build an ideal hyper plane in sample space or feature space that maximises the distance between the hyper plane and various sorts of sample sets in order to reach the highest level of generality. SVM offers novel approaches to nonlinear problem solving. It is one of the most successful approaches to nonlinear classification and regression. It presents a "framework" in place of the standard "empirical risk minimization principle" for evaluating and testing the model. The concept of risk minimization induction considers both the fitting error and the generalisation ability, therefore it considers both accuracy of the training sample and also the model's promotion ability. Many weather forecasting and forecasting issues have strong non-linear aspects. This method is expected to be widely used in weather forecasting research and business.

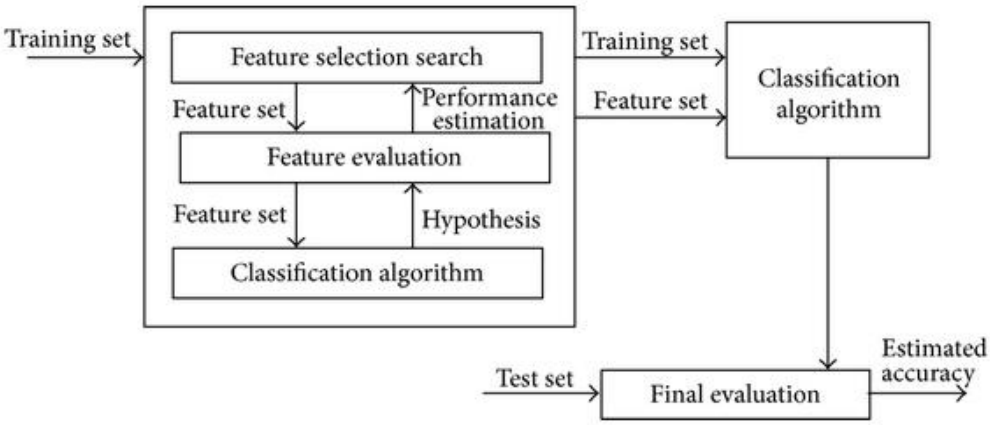

Fig2. Flowchart of Wrapper-based Approach to Feature Selection.

2.2 WRAPPER BASED SUPPORT VECTOR MACHINE (WSVM)

There are three phases in proposed wrapper-based prediction (Shown in Figure 3), they are,

1. Data Preparation Phase (DPP): The agricultural data acquired from the Sensor of smart farming system are pre-processed and managed in this stage before it can be used in the subsequent phases.

ISSN 1533-9211
DOI 10.5281/zenodo.6622423

2. Feature Selection Phase (FSP): The wrapper feature selection approach is used in this step to examine environmental index, which selects the most effective indicators for the IoT-based agricultural system.
3. Prediction Phase (PP): In this phase, the SVM algorithm is utilised to build a classifier for predicting drought and agricultural production, and the suggested prediction technique is then tested and compared to the existing ones.

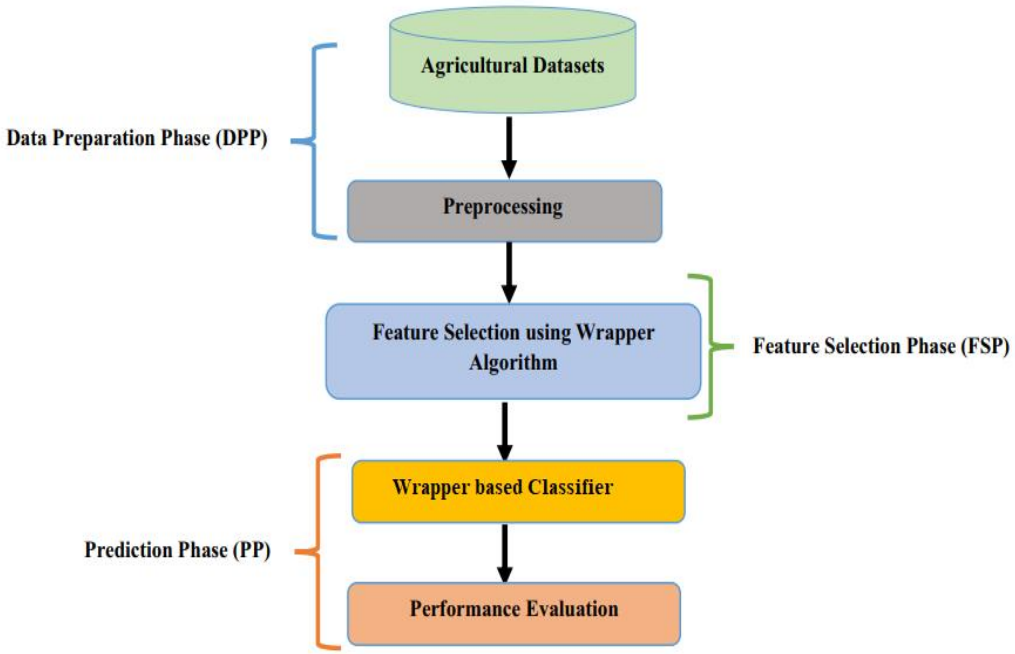

Fig 3. Proposed Prediction Model [3]

For a few datasets, several classification algorithms were used to forecast drought and agricultural yield. Wrapper approaches use a greedy search algorithm that examines all potential feature subset combinations and picks the accurate subset for a given machine learning algorithm. As a result, the Wrapper based feature selection approach chooses the most constructive environment indicators for the SVM classification algorithm, which is a faster classification algorithm than the others.

Drought is one of the main meteorological disasters in India. Its occurrence is very slow, the symptoms before the occurrence of the disaster are unclear, and it is difficult to judge the beginning and end of the drought and the extent of its development. In order to quantify the severity of the drought, a comprehensive index of meteorological drought is used to quantify the drought grading in the autumn season. The Weather and Drought Comprehensive Index (CI) is a kind of drought evaluation index based on the Standardized Precipitation Index (SPI) and relative Humidity Index (MI). The index has a more comprehensive consideration factor,

The Seybold REPORT
ISSN 1533-9211
DOI 10.5281/zenodo.6622423

taking into account the weather conditions in the previous period, and has a better comparison of time and space. The CI index value of the autumn season in country from 2000 to 2011 represents the severity of the drought in the autumn season over the years. The classification of its autumn historical drought information is based on the fact that the CI index reaches the drought level or higher and is recorded as '1', and the remaining records are recorded as '0', thus realizing the forecasting target of the historical seasonal drought (Kharif and Rabi).

2.3 CLASSIFICATION MODEL

Considering that there are N model samples in this experiment, cross-validation is used to establish the optimal model, that is, a part of the total sample data is randomly choosen as the training set, where as the other samples are used as the test set, and the training set is used for modelling. Then use the test set to test and adjust the model parameter until the optimal parameter is found. Multiple random samples are drawn to establish a model, the stability of these models is compared, and the optimal model is selected. Multiple extractions of data are equivalent to multiple cross-modelling. The more times it is performed, it is equivalent to the number of cross-modelling. In this test, 90% of the samples were randomly choosen as the training set and 10% of the samples were used as the test set when the model was built. A total of N/10 random samples were selected for model building

The process of selecting the parameters for the optimal model is also called as the cross-validation process. At present, the optimal method of parameter optimization has not been studied. Continuous test parameters is used in this work to construct the model, TS scores the established model and also TS scores the training samples through comparison. With the TS score of the test sample, the model with the largest TS score is selected as the optimal model. The ideal model should be between the fitting accuracy and the promotion ability, and the determination of related parameters is also a complex optimization problem. The promotion ability of SVM is related to the total sample size N, the penalty coefficient C, and the set of candidate functions. The model parameters are adjusted cyclically, mainly adjusting the penalty coefficient C and the G in the radial basis function. The adjustment process is mainly divided into two steps, the first step uses a larger step length to alter the parameters in order to find the optimal model's parameters, and the second step selects a smaller step near the best model parameter to optimize the first step. TS score of the correct sample that produced the best model was 98%. The model established by SVM is determined by a small number of support vectors. Therefore, only the key samples are concerned and the redundant samples are not considered to provide a basis for the analysis of the factors.

3. EXPERIMENTAL RESULTS

The attribute of the crops such as District, year, month, rainfall, Average Rainfall, Temperature, average Temperature, Pressure and Average Pressure are collected from the sensors and farmers and framed three real time datasets, namely, Sorghum, Jowar, Sugarcane

ISSN 1533-9211

DOI 10.5281/zenodo.6622423

and Drought framed from India water portal [19].The experiments are carried out using MATLAB 2016 tool under Windows 10 Environment. The proposed model compared with five existing prediction models namely K-Nearest Neighbour, Naïve Bayes, Back Propagation Neural Networks, Wrapper based PART to identify its performance efficiency.

The repeated 10-fold cross-validation approach was used to examine the generalisation of the prediction models developed by the preceding algorithms. First, the entire dataset was partitioned into 10 sections at random. The system was then trained using nine-tenths of the data before being evaluated with the remaining tenth to determine its predicted ability. This operation was carried out nine times more. A different tenth of the sample was utilised as testing sample each time, and a different nine-tenths was used as training sample. Finally, by repeating the same standard 10-fold cross-validation 100 times with varied data splits, the average estimate over all runs was calculated. The 10-fold cross-validation testing is used to analyze the performance of all the models with and without feature selection. For the comparison study, the performance metrics Accuracy, Precision, Recall, and F-Score values are produced.

True Positive (TP) (i.e., positive instance correctly classified), False Positive (FP) (i.e., negative instance classified as positive), True Negative (TN) (i.e., negative instance correctly classified), and False Negative (FN) (i.e., incorrectly classified negative instance) are the four possible prediction outcomes for a classification problem (i.e., positive instance classified as negative). The four values are the foundation for a number of additional well-known and widely used performance criteria for classifier assessment. The formula for calculating Overall Accuracy (OA) is OA = (|T P| + |T N|)/N, where N is the total number of occurrences in a dataset. While overall accuracy allows for simpler model performance comparisons, it is not always regarded as a credible performance measure. Precision =TP/(TP+FP), Recall=TP/(TP+FN), and F=2*((Recall*Precision)/(Recall + Precision)) are all evaluated in addition.

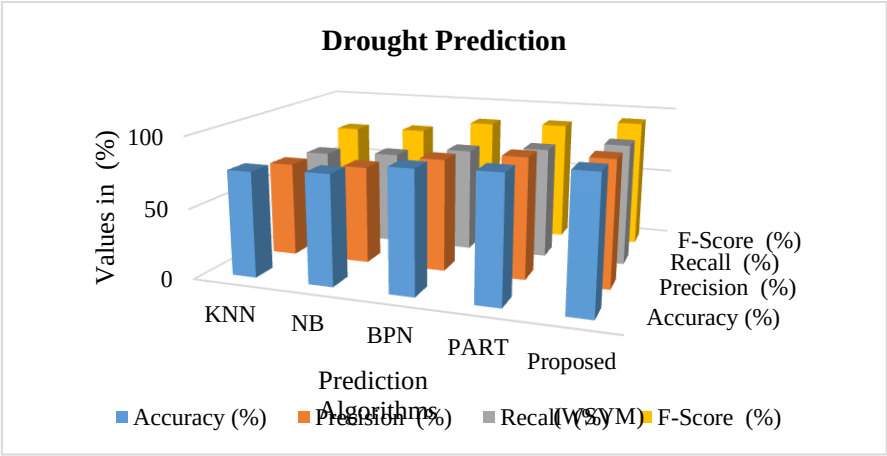

Fig 4. Drought Prediction

ISSN 1533-9211
DOI 10.5281/zenodo.6622423

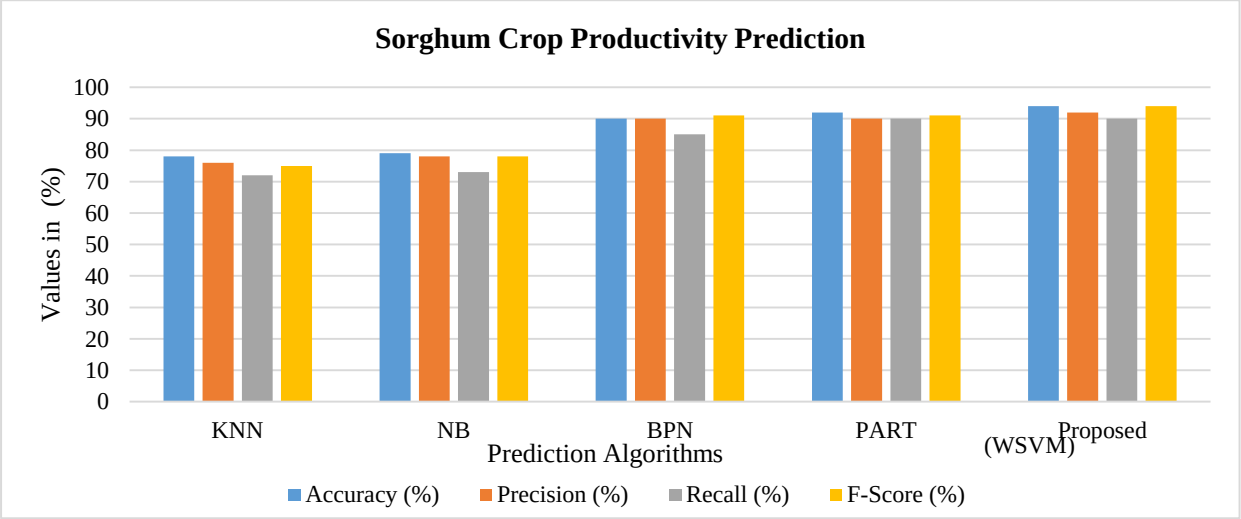

Fig 5. Sorghum Crop Productivity Prediction

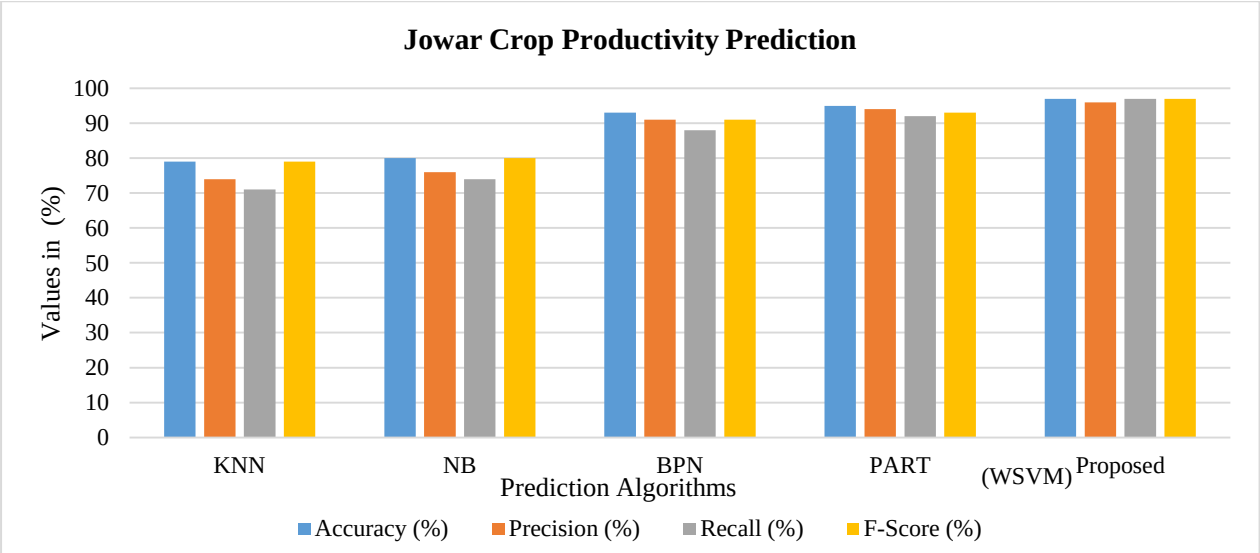

Fig 6. Jowar Crop Productivity Prediction

ISSN 1533-9211

DOI 10.5281/zenodo.6622423

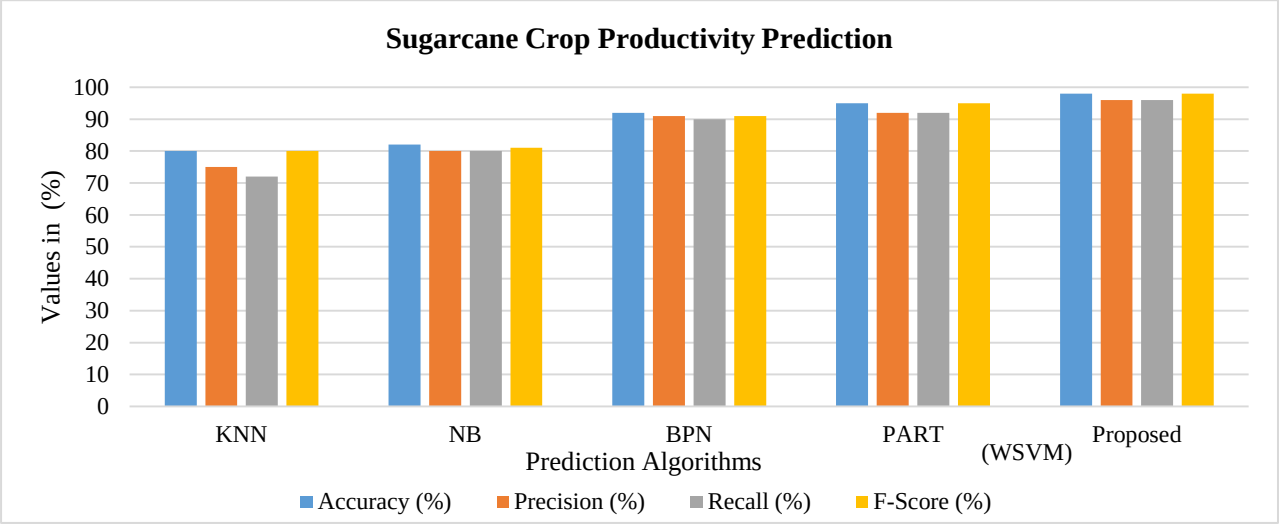

Fig 7. Sugarcane Crop Productivity Prediction

From the Figures 4, 5, 6 and 7 it is revealed that the proposed wrapper-based classifier achieves highest accuracy when compared to the other classifiers. When comparing recall values, the proposed prediction model gives the highest values than the other single classifiers. The proposed classifier enhances the prediction accuracy as well as feature selection. In wrapper-based feature subset selection, the redundant features are eliminated to avoid the misclassification and reduced the prediction time.

Sl No	Performance Measures	KNN	NB	BPN	PART	WSVM (proposed)
1	Accuracy	75	78	80	82	90
2	Precision	68	70	78	80	88
3	Recall	65	68	75	78	85
4	F-Score	75	70	80	82	90

Table. 1 Drought Prediction

The following are the outcome of this research,

1. An efficient IoT based decision support system for smart farming is developed for drought and crop productivity prediction.
2. Whether forecast is alerted earlier to decide to choose the right crop for the right season to take right action to get higher yield.
3. Prediction of crop productivity will support farmers to choose the right crop for the right season (Kharif and Rabi) if rainfall increase or decrease.

The Seybold REPORT
ISSN 1533-9211
DOI 10.5281/zenodo.6622423

4. CONCLUSION

Short-term climate predictions are currently mainly based on statistical methods (similarity analysis, stepwise regression, and principal component analysis), etc. These analyses are based on linear correlation, emphasizing the role of some obvious influencing factors, and the climate system itself is a complex non-linear system are consistent with the theory of SVM pattern recognition to deal with nonlinear problems. SVM uses the introduction of the radial kernel theorem to transform the non-linear problem into a linear problem in the high-dimensional feature space. The algorithm is simple and can be widely used in the field of climate prediction. The forecast of drought is generally first to predict the future climate elements by conventional methods. (Temperature, Precipitation, etc.), and then the climatic element value is used to judge the drought or humid state, and the wrapper SVM is carried out by direct inference, and a small number of support vectors are used to describe the non-linear dependence between factors and objects. In this paper, Wrapper based SVM drought prediction model is established, the drought indicators are used to classify the prediction objects of the drought training samples, the predictors closely related to the drought are selected, and the cross-validation method is used to select the optimal model parameters, and carry out effectiveness test. The results show that the established drought prediction model can directly predict the drought in the both kharif and rabi season, and has a high prediction accuracy rate, F score values and the model has a good extension.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study is carried out as a part of Research of own interest of authors. There is no funding involved in this research. The authors declare that they have no conflicts of interest. This article does not contain any studies involving animals performed by any of the authors. This article does not contain any studies involving human participants performed by any of the authors.

REFERENCES

1. Qian, Z. A., Wu, T., Song, M., Ma, X., Cai, Y., &Liang, X. (2001). Arid disaster and advances in arid climate researches over Northwest China. Advances in Earth Science, 16(1), 28-38.
2. Yu, M., Li, Q., Hayes, M. J., Svoboda, M. D., & Heim, R. R. (2014). Are droughts becoming more frequent or severe in China based on the standardized precipitation evapotranspiration index: 1951–2010?. International Journal of Climatology, 34(3), 545-558.
3. Rezk, N. G., Hemdan, E. E. D., Attia, A. F., El-Sayed, A., & El-Rashidy, M. A. (2021). An efficient iot based smart farming system using machine learning algorithms. Multimedia Tools and Applications, 80(1), 773-797.
4. Wang, C., Wang, Z., & Guo, Y. (2012). Application and Verification of Drought Index in Meteorology Drought Monitoring and Prediction. Advances in Earth Science, 27(9), 957-968.
5. CharneyJG. Dynamics of Desert Sand Droughtinthesahel[J]. Quarterly Journal of the Royal Meteorological Society, 1975, 101(428):193-202.
6. Webb, R. S., Rosenzweig, C. E., & Levine, E. R. (1993). Specifying land surface characteristics in general circulation models: Soil profile data set and derived water-holding capacities. Global Biogeochemical Cycles, 7(1), 97-108.

The Seybold REPORT
ISSN 1533-9211
DOI 10.5281/zenodo.6622423

7. Wang, C. H., &Wang, L. D. (2010). Characteristics of Sensible and Latent Heat Fluxes in Semi-Arid Region of Northwest China and Change Trend in Recent 50 Years. Plateau Meteorology, 4.
8. Wang Chenghai, Wang Shigong, Yang Debao, and Dong Anxiang. (2002). Study on the characteristics of spring precipitation in northwest China and Pacific sea temperature. Journal of Tropical Meteorology, 18(4), 374-382.
9. Suresh M, Manjupriya S, (2021) Internet of Things: Views, Implementations and Challenges in Deployment, IT in Industry, 72 - 76
10. Bai Qingmei, Tian Wenshou, Feng Zhaodong, Wang Chenghai, & Jin Liya. (2010). Correlative characteristics of precipitation and oceanic warm pool climate in arid/semi-arid regions of Asia. Glaciology and Geocryology, 32(2), 295-308.
11. Zhang Shuangjie, Wang Baoling, Liu Dexiang, and Cai Zhonglan. (1998) Study on drought and flood indicators in northwest China. Highland Weather , 17(4), 381-389.
12. Yuan Wenping, Zhou Guangsheng. (2004). A comparative analysis of the application of standardized precipitation index and Z index in China. Journal of Plant Ecology , 28(4), 523.
13. Wang Jinsong, Guo Jiangyong, Zhou Yuewu, and Yang Lanfang. (2007). Progress and prospect of drought indicator research. Arid Zone Geography , 30(1), 60-65.
14. Yuan Wenping, Zhou Guangsheng. (2004). The theoretical analysis and research outlook of drought indicators. Earth Science Advances , 19(6), 982-991.
15. Zhang Qiang, Zhang Liang, Cui Xiancheng, and Zeng Jian. (2011). The development of drought monitoring and evaluation technology and its scientific challenges. Earth Science Advances, 26(7), 763-778.
16. M. Suresh, S Manju Priya. (2020) Internet of Things in Agriculture: an overview of Concepts and the Challenges in its Implementation, Test Engineering and Management, 16995 – 16999.
17. Huang Ronghui, Chen Yilong, Zhou Lian Tong, zhang Qingyun. (2003) Research on the relationship between major climate disasters in China and the east Asian climate system.
18. Wang Chenghai, . . . Hui Xiaoying. (2005). The change of desertification and grassland boundary in China is seen by the vegetation index 0.12. Chinese Desert, 25(1), 88-92.
19. Zhou Yaqing, Song Yan, Xiao Zibu, Zhang Hongying. (2021). The response of global sea temperature to solar flux anomalies and their effect on precipitation. Journal of Geophysics, 64(5), 1518-1530.
20. Vapnik, Vladimir N. (1995). Support-vector networks (PDF). Machine Learning. 20 (3): 273–297. CiteSeerX 10.1.1.15.9362. doi:10.1007/BF00994018. S2CID 206787478.
21. https://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-data-list%20rain%20gauge%20stations-India-meteorological-department/district wise monthly rainfalldata list%20rain%20gauge%20stations-India-meteorological-department.