

AN ENHANCED APPROACH FOR CROP YIELD PREDICTION SYSTEM USING LINEAR SUPPORT VECTOR MACHINE MODEL

¹K.Priyadharshini,
Dept. Computer Science and Engineering
PERI Institute of Technology
Chennai, India
priyared7@gmail.com

²R. Prabavathi,
Department of Information Technology.
Sri Sairam Institute of Technology
Chennai, India
prabavathi.it@sairamit.edu.in

³V. Brindha Devi
Department of Information Technology
Sri Sairam Institute of Technology
Chennai, India
hodit@sairamit.edu.in

⁴P.Subha
Department of Information Technology.
Sairam Institute of Technology
Chennai, India
subha.it@sairamit.edu.in

⁵S.Mohana saranya
Dept.Computer Science and Engineering
Kongu Engineering College
Erode, India
mohanasaranya.cse@kongu.edu

⁶K.Kiruthika
Dept.Computer Science & Engineering
Panimalar Engineering College
Chennai, India
kiruthikaitdept@gmail.com

Abstract- Smart Agriculture is an emerging progressing field which is used for the management of farming to increase the yield of the crops. Since India is a populated country, urge of food production also increases. This situation is one of the reasons that hindering the development of country. At present farmers get more yield for their crop, but the market price for that crop is very less. To conquer these problems, a machine learning technology is used. The prediction will assist the farmers to select whether the specific crop is suitable for certain season and crop price values. Prediction techniques like linear regression, SVM, KNN method and decision tree of machine learning is widely used in the field of agriculture. This paper proposes a novel method that would deliver suitable support vectors for a SVM classification based on auxiliary information. This optimized method is applied to a real time agricultural application situation which utilize accuracy classification in turn aid production management. The proposed SVM method gives an accuracy of 91% than the existing system. This method can be implemented in several government sectors like APMC, kissan call centre etc., by which the government and farmers can get the information of the future crop yield and the market price.

Keywords: Machine Learning, Linear Support Vector Machine (SVM), Predictive Analysis, Accuracy, Crop yield.

INTRODUCTION

With a growing global population and a changing climate, it is more important than ever to protect the world's food supply. Due to climates and economic instability, farmers

must make difficult decisions about productivity and sustainability [1]. Farmers can make the best decisions for their agricultural scenarios if they have timely and reliable information about soil, temperature, subsurface water, and rainfall [2]. This might help them increase crop output if the conditions are favourable, or it could assist them to reduce crop yield losses owing to unfavourable conditions. A few research have investigated how machine learning approaches may be used to improve agricultural yield prediction and have achieved results in a variety of climate scenarios. The prediction system equips the farmers to decide about whether the selected crop is appropriate for current season and crop price values.

Prediction can be performed by using a variety of machine learning techniques like linear regression, SVM, K-Nearest Neighbour method and decision tree algorithm. This paper proposes novel method to supply proper support vectors for a SVM classification depending on auxiliary input information. The SVM model is proposed in this work for predicting the yield of the agricultural crop. Prediction accuracy is one of the significant demands identified in the existing methodologies. To improve the prediction accuracy, this paper proposes a novel method Linear SVM classifier. The data were analysed, and conclusions were drawn about its performance in increasing crop yield production.

The rest of the paper is structured as follows. Related works are discussed in Section II, Section III discussed about proposed system design, Section IV lists out the performance evaluation and experimental results of the proposed method and comparative analysis has been done. Section V details the conclusion and next moves.

II LITERATURE SURVEY

Machine learning models are used for the study of soil parameters for predicting the yield of the crop. The authors have used moisture content of a soil, climatic factors for the predicting the right crop for producing maximum yields. The system has used SVM and RVM techniques for learning the soil parameters to produce best results [2].

To increase crop productivity, effective prediction system was developed for management of the agricultural lands in producing the best crop using potential parameters. Also, climatic smart adaption methodologies were also aided in learning models.[3].

The authors have developed prediction system to recognize the most imperative climate variable and to estimate the utilization of remote sensing data. The model blends with the climate variables in producing the solution to the farmers for choosing the right crop at right time for yield prediction. The models regulate the implication of long historical data records for yield prediction.[4].

Climatic change is one of the biggest challenges faced by the farmers. The effect of climate influencing parameters like temperature, rainfall, and humidity are measured to predict crop productivity [5]. The SVM Classifier is implemented to discriminate crop and weed. The methods used by the authors are investigated and analysed for observing the maize and weed plants. Most of the crop models are employed for rice progress prediction in the growth phases using the SVM models [3]. Soil classification was also performed by models to categorize best suitable crop using SVM [4]. The fundamental challenge in the modelling of SVM is to resolve the kernel functions, tuning the hyper parameters and the coefficient penalty. Aerial hyperspectral values used for analysis and modelling is also a challenging factor in SVM. The system was designed to analyse SVM based aerial hyperspectral values like biomass, plant height, and leaf greenness. The comparative analysis has been done with stepwise regression method. Most of the surveys done shows that the supervised learning algorithms are widely used in the prediction of the crop yield [6].

The prediction factors like minimum and average temperature, evapotranspiration and precipitation are used in SVM system in determining the climatic change for producing the accurate and efficient results [6]. The soil

parameters like code, thickness, entropy, pH of the soil and nitrogen content are used in SVM model for predicting the fertility of the soil [6].

Weather data is used as the classification feature for recommendation of the best seasonal crops. This system has been tested using MATLAB [17] The SVM classification features in this system used presumed parameters for prediction. Prediction analysis using Linear regression method was implemented based on Soil quality, soil nutrients and soil composition [6].

From the forementioned study, it is examined that the weather parameters, pH, soil nutrients, type of the crop and seed type are used as the input parameters. The input feature is then pre-processed to eliminate erroneous and noisy data. After which the feature vector is used by the model for the prediction of the crop yield. For measuring the performance of the system various metrics are used by different machine learning algorithms. A comparative analysis is done on the performance of various algorithms. Furthermore, from the related study it is observed that digital innovations in the field of agriculture with an enormous amount and variety of data is being used in producing accurate results. Hence, agriculture innovations are the current requirement for the farmers for the increasing the yield of the crops. Machine learning is in trend for providing solutions. This paper provides a technological solution for crops using the best analysis and prediction method and suggest the best crop for the farmers to sow on the land.

III SYSTEM DESIGN

The architecture of the prediction model is given in figure 1. The model consists of crop data set collection module, pre-processing module, Linear SVM model and finally crop yield prediction output module. The architectural elements of the proposed method are explained in the following section.

The proposed methodology focuses primarily on the production of crop which is based on the four components and the linear support vector machine learning method is used for the prediction of the yield. The prediction yield parameters namely, temperature, subterranean water, and rainfall are assumed to categorize whether a crop can grow in that place using SVM model.

The SVM is a supervised machine learning technique which is used to resolve binary classification problems.

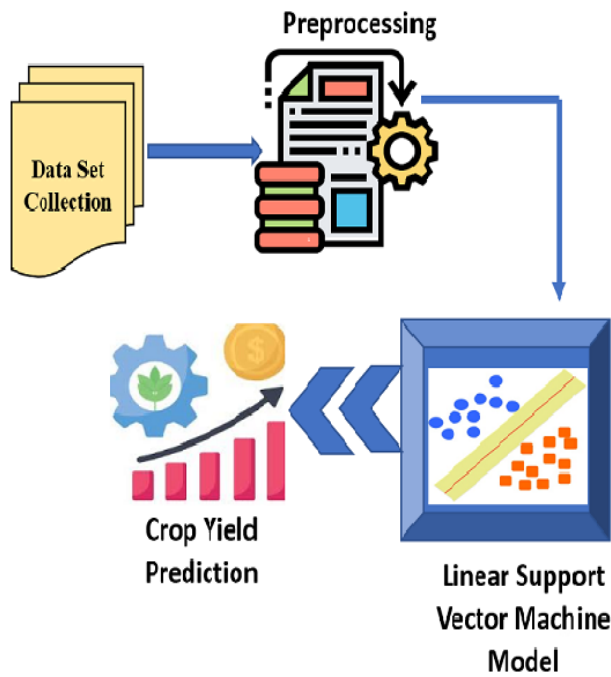


Figure 1: Prediction Model for Crop Yield

The intention of the proposed methodology is to build a hyper plane in an N-dimensional space, where N denotes the number of features in a dataset that will be used to categorise data points.

The mapping of N number of hyper planes is possible but the proposed method's primary objective is to determine the plane with the maximum Margin, or the longest distance between data points of the features. To obtain high accuracy in classification, the system must find maximal distance. The data points which are in log distance from all other points in Optimum hyper plane, produces maximum margin.

Data Set Collection

To properly train a model, efficient dataset is required. The four data parameters like nitrogen level in the soil, annual subterranean water, annual rainfall, and temperature are employed in this proposed method. The data is focused on India's various states in India. so, we gathered data from a publicly available dataset on the Indian government website, which covers all the annual values of the criteria indicated previously. The rainfall dataset was composed from the Government of India website data.gov.in. In this dataset, the target variable is "output," which has two values: 0 and 1, this indicates whether crop is suitable in that area, and vice versa. There are 12 variables and 650 observations in this dataset.

Pre-processing

The pre-processing has been done to enable the training and testing procedure by properly transforming and scaling the entire given input dataset. Pre-processing step removed outliers and scaled the features to an equivalent range.

Linear SVM Model

Crop yield prediction is treated as a binary classification problem in this proposed work. Classifier has been utilized to check if a Crop Yield is appropriate for that season and location. These two classes are referred as "CYFIT" and "CY NOT FIT".

Let's start with the most basic scenario: linearly separable training patterns. That is, a linear function is formulated as in Eq.1.

$$f(c) = W^T C + b \quad (1)$$

Such that the function returns as follows in Eq2. and Eq3.

$$f(c) > 0 \text{ for } y_i = 1 \quad (2)$$

and

$$f(c) < 0 \text{ for } y_i = 0 \quad (3)$$

or each training case c_i . In other words, the hyperplane $f(c) = W^T c + b = 0$ isolates two training samples from two dissimilar classes.

Various hyper planes are available to separate the classes for a given training set, the SVM classifier is employed depending on the hyperplane which produces maximal margin between the two classes. Alternatively, SVM seeks out the hyperplane that results in the peak isolation of decision function values for the "borderline" cases from the two classes. This hyper plane can be obtained mathematically by reducing the following cost function as in Eq.4

$$J(w) = 1/2 w^T w = 1/2 \|w\|^2 \quad (4)$$

the isolating constraints as in Eq.5.

$$W^T X + b = > +1, \text{ for } y_i = 1 \quad (5)$$

Linear SVM Crop yield Prediction method

In this proposed model crop features are used as input vector, which is pre-processed and is used by the model for the prediction. The output vector is the yield prediction. The following procedure has been employed by proposed linear SVM crop yield prediction method.

Step 1: Once Dataset has been collected, these raw data has been transformed into required input information. In certain situation that the proposed system applied while not getting enough raw data to work with model, apply duly designed format has been applied for obtaining the appropriate results.

Step 2: Categorization of the dataset is done to form two clusters as training and testing dataset. The training set consists of the highest rate of information for training the instances for the production. Approximately 75% of the dataset collected from the government website is used for training and to ensure the suggested model's performance for testing, the remaining 25% of the dataset was used.

Step 3: The proposed prediction model Linear SVM is applied on the feature dataset for producing the prediction of the right crop and yield of that crop.

Step 4: The performance of the system is measured using R^2 , RMSE (Root Mean Squared Error). The accuracy of the system is also obtained.

Step 5: Trained model is then used to perform testing on the test dataset. Then the performance metrics is once again calculated to find the efficiency of the system. The proposed model gives high accuracy and R^2 values, and low RMSE values which indicates that the proposed model is highly efficient for producing the crop yield prediction.

IV PERFORMANCE EVALUATION

Each instance is divided into two classes using a binary Classification model. The two types of classes are true and false. As a result, each case can be classified into one of four categories:

True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) are used to for representing the Contingency matrix.

In a confusion matrix, the classifications are measured for a phenomenon which is then compared to the expected classifications of a model. The accurate classifications obtained from the model are referred to as positives and negatives. The most accurate system produces true positive and true negative results only and the remaining fields is set to 0. Several model performance measures can be obtained from the confusion matrix. The most widely used metric is accuracy, which is defined as the classifier's total success rate.

Recall and Precision are the other two performance indicators to also be considered. The sensitivity is defined as the percentage of correctly categorised instances. The fraction of erroneously classified instances is referred to as specificity. Sensitivity/Recall = $TP / (TP + FN)$ can be used to calculate. The F1 Score is a metric for determining how accurate a test is. Precision and recall are considered while calculating the score. Its best and worst value are 1 and 0.

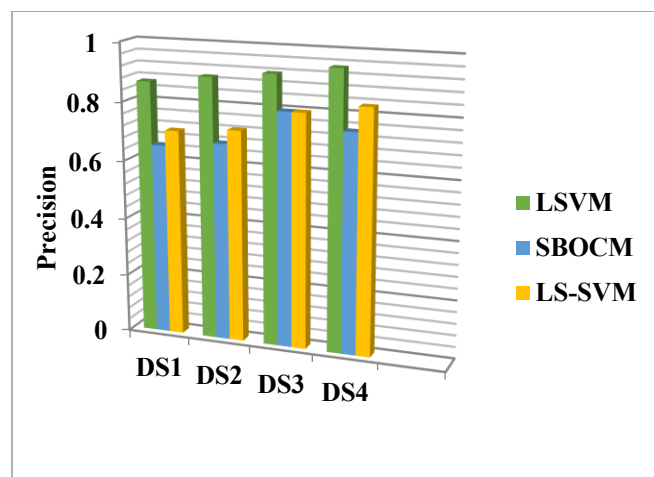


Figure 2. Precision Values Comparative Analysis

The proposed SVM model produces 78.76 percent accuracy, 68.17 percent sensitivity, and 83.97 percent specificity. The classifier's error statistics are 0.39 root mean squared error. Results show that the system equipped with proposed LSVM gives precision values minimum (0.85) and maximum (0.97) for a given datasets. Therefore, the proposed LSVM method achieves higher precisions than the SBOCM and LS-SVM as shown in Figure 2. This implies that the number of false positives is less the proposed LSVM approach which proves that it is more reliable. The proposed LSVM method produces Recall values in the range 0 to 1.2 and minimum (0.79) and, maximum (1.1) as shown in Figure 3. Correspondingly it is also proved that from Figure 2 SVM has the highest Recall value of 1.0 on compared to SBOCM (0.79). The high Recall values imply that the number of false negatives is very less in proposed LSVM approach.

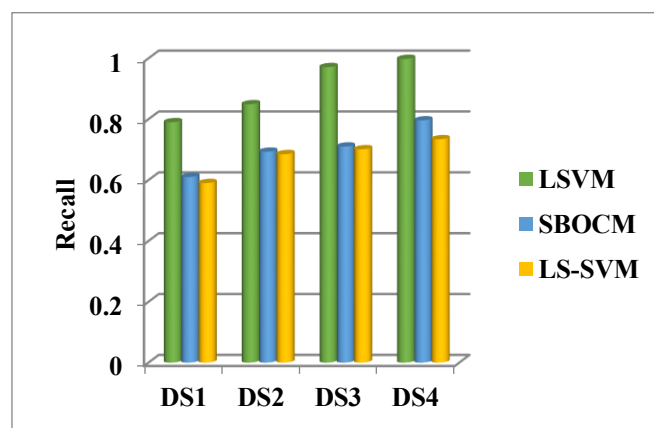


Figure 3. Recall value Comparative Analysis

Figure 4 depicts the F-Measure of proposed LSVM method, and it is compared with SBOCM and LS-SVM approach. F-Measure is a metric used to weigh precision and recall uniformly. The F Measure values is drawn out in the interval 0 to 1.2. It is apparent that proposed

LSVM has the highest F-measure value of 1.09 when compared to SBOCM and LS-SVM approach (0.81) implies that the proposed LSVM is much more efficient than other existing systems.

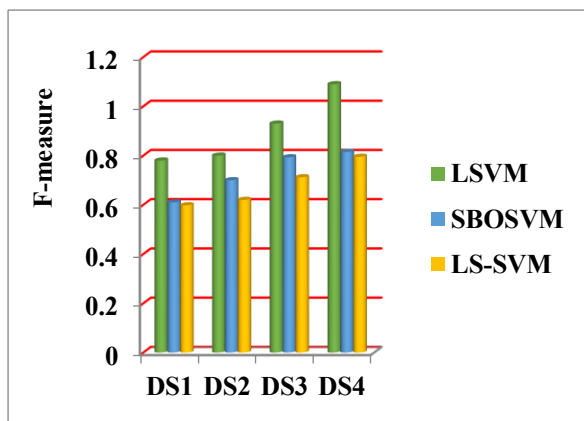


Figure 4. F- measure valuecomparative analysis

V. CONCLUSION

This paper proposed novel method which provides the appropriate support vectors for the classification of the auxiliary information using the linear support vector machine. This optimized approach is suitable for applying to any operational agricultural application scenario for producing the best yield of the crop. The proposed LSVM gives more accuracy than the existing system. The model can be implemented in several government sectors like

APMC, kissan call centre etc., by which the government and farmers can get the information of the future crop yield and the market price. In future fertilizer recommendation system will enhance the quality of the crop and maintain the fertility of the soil.

REFERENCES

- [1] M. S. S et al., "Optimized Deep Convolutional Generative Adversarial Network for Tomato Leaf Disease Identification," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2021, pp. 1301-1307, doi: 10.1109/ICECA52323.2021.9675945.
- [2] G. S. Sajja, S. S. Jha, H. Mhamdi, M. Naved, S. Ray and K. Phasinam, "An Investigation on Crop Yield Prediction Using Machine Learning," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 916-921, doi: 10.1109/ICIRCA51532.2021.9544815.
- [3] I. P. Handayani and N. Folz, "Adaptive Land Management for Climate-Smart Agriculture," 2021 IEEE International Conference on Health, Instrumentation & Measurement, and Natural Sciences (InHeNce), 2021, pp. 1-7, doi: 10.1109/InHeNce52833.2021.9537265.
- [4] A. Mohanty, "Impacts of climate change on human health and agriculture in recent years," 2021 IEEE Region 10 Symposium (TENSYP), 2021, pp. 1-4, doi: 10.1109/TENSYP52854.2021.9550876.
- [5] Su, Ying-xue, Huan Xu, and Li-jiao Yan. "Support vector machine-based open crop model (SBOCM): Case of rice production in China." *Saudi journal of biological sciences* 24.3 (2017): 537-547.
- [6] S. Sharma, G. Rathee and H. Saini, "Big Data Analytics for Crop Prediction Mode Using Optimization Technique," 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2018, pp. 760-764, doi: 10.1109/PDGC.2018.8746001.