ISSN (online) : 2277-5390
ISSN (Print): 2277-5382

A survey on crop yield prediction models

K. Menaka[1] N.Yuvaraj[2]

[1]*M.E. (Computer Science and Engineering),* [2]*Department ofComputer Science and Engineering*
KPR Institute of Engineering and Technology, Arasur, Coimbatore, India
k.menaka.06@gmail.com yuvaraj.n@kpriet.ac.in

Abstract

Objectives: To analysis various models to improve the prediction of crop yield production.
Methods: In this paper, there are different methods has been analyzed to predict the crop yield. The methods such as artificial neural network, Adaptive Neuro-Fuzzy inference System, Fuzzy Logic and Multi Linear Regression are analyzed to know the best methods for crop yield prediction. The prediction of crop yield varied by internal factors and external factors of crop an environment. The internal factors such as pesticides, water level, spacing and fertilizers and the external factors such as temperature, humidity, and rainfall. There are various models were developed to predict the crop yield prediction. This paper provides detailed information about the different models for crop yield prediction.
Findings: In this paper various models for crop yield prediction are compared through their parameters such as Root Mean Square Error (RMSE), R^2,correlation coefficient and Mean Square Error (MSE) to prove Adaptive NeuroFuzzy Inference System (ANFIS)prediction model is better than other techniques.
Application/improvements: The findings of this work prove that the Adaptive Neuro Fuzzy Inference System (ANFIS) prediction model provides better result than other approaches.
***Keywords*:** Crop yield prediction, Adaptive neurofuzzy inference system, Data mining, agriculture.

1. Introduction

Agriculture is one of the main sectors to be impacted by different sources like climatic changes, soil attributes, seasonal changes etc., Crop yield prediction [1] is based on various kinds of data collected and extracted by using data mining techniques [2] such as machine learning techniques [3] different sources which are useful for growth of the crop. It is an art of forecasting crop and the quantity of yield in advance i.e., before the harvest actually takes place. Predicting the crop yield can be extremely useful for farmers. If they have an idea of the amount of yield they can expect, they can contract their crop prior to harvest, often securing a more competitive price than if they were to wait until after harvest. The involvement of experts in prediction of crop yield leads to issues like lack of knowledge about natural events, negation of personal perception and fatigue etc. such issues can be to overcome by using the models and decision tools for crop yield prediction. Likewise, industry can benefit from yield predictions by better planning the logistics of their business.

2. Crop yield prediction models

Attribute selection:

The dependable attributes can be difficult to find. Several methods of predicting and modeling crop yields have been used in the past with varying success. Famer has to face the different problems due to various factors which affect the planning made by him in advance. These factors do not have the fixed type of impact, it varies time to time, year to year depends on the situation, climatic nature, increase in costs of various constraints under uncertain environment, ambiguity and vagueness. Fuzzy logic [4] modeling provides the formulation of mathematical modeling to find the interface results in uncertain situations. Statistical models often don't take into account characteristics of the plants, the weather, or the soil attributes limiting their usefulness. Some models are based on information from just a single year or location. When a model is developed using single location or year data, it will have limited practical applications, therefore variability from multiple environments must be included.

Crop yield prediction:

The crop yield prediction comprises of mostly all essential parameters that are needed for the better yield of crop. This enhances the classification results of the crop yield. All the essential parameters are considered as inputs to the model-ANFIS. In general, one of the difficulties faced in the prediction process is that most of the essential parameters that are necessary to consider for the accurate prediction are not consider. It reduces the efficiency of the predicted results which in turn leads to lack of proper forecasting of the crop yield. It is also more complex to predict the optimized number of input parameters that are to be considered in the prediction process.

ANFIS models:

ANFIS model is one of the efficient ways which is used for prediction, by imposing most of the essential parameters as inputs, it improves the accuracy of prediction results which has the property of learning by artificial neural network [5].ANFIS presents some linearity with respect to some of its parameters, hence it increases the overhead of computation process without increasing the efficiency. The ANFIS failed to optimize the fuzzy rules in ANFIS which degrades the performance of prediction. Inputs that are to be consider and selected depends on the heuristics. It also decreases the performance by degrading the efficiency of the prediction process.

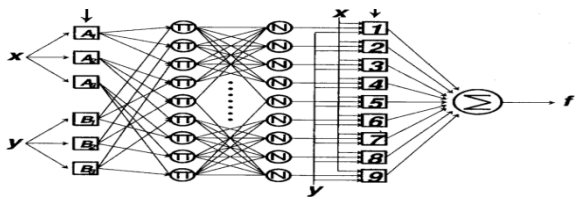
Fuzzy inference sysytem:

The implementation of Fuzzy Inference System (FIS) [6] as a technique for predicting the onset of rainy season based on the Southern Oscillation Index (SOI) data. Fuzzy set memberships and the set of rules are designed by investigating the two sets of data (via visualization and clustering). The prediction system is verified by using the actual data from the district. This explains an implementation of a Fuzzy Inference System for the prediction of rainy season onset. A fuzzy system was developed based on the knowledge gained from data exploration in a certain district to cover a larger range of area. Also, the computation of the output from the input data is based on a logical knowledge backed by observation data. Accordingly, it is expected to have a better accuracy result than the previous works.

Neuro fuzzy technique:

The model explains the multi objective linear programming problem [7] by optimization technique and ANFIS model on neuron fuzzy technique for prediction of crop yield. In this the author explains, Fuzzy logic is utilized for the purpose of effective feature extraction and classified the crop yield data with the help of extracted features using ANFIS model. The ANFIS classified the data based on the fuzzy sets created in every input variable. It analysis various optimization methods and it is proved that OR optimization solution gave better results than the other optimization solution in fixed situation. The author explains deals with MOLPP by optimization technique and ANFIS model on Neuro-Fuzzy technique. Fuzzy Logic is used for effective feature extraction and Adaptive Neuro-Fuzzy Inference System (ANFIS) is considered for the classifier model. The performance of the ANFIS model is evaluated in terms of training performance and classification accuracies and the results confirmed that the proposed ANFIS model has potential in decision making of multi criteria linear objective function from the past experiences. It provides improved performance than the other classified method of optimizations. ANFIS model gives the solution in vagueness whereas other optimization methods of OR gives solution in fixed situation. The developed ANFIS is of very simple structure and contains auto generated rules and thus can be easily implemented. Figure 1 shows ANFIS structure (for input 1 to 4). ANFIS1 is having 4 inputs and only one output.

Figure 1. ANFIS structure (for input 1 to 4)



ANN and MLR:

This model utilized Artificial Neural Network (ANN) and Multi Linear Regression Model (MLR)[8] for seed yield prediction. Artificial neural network (ANN) is an intelligence model and it imitates the procedure of the human brain works. In other words, the ANN has certain performance characteristics like biological neural networks of the human.

A typical ANN consists of a number of simple processing elements called neurons or nodes. These neurons are organized into groups termed layers. In this the author explains, the multi-layer perceptron (MLP) with back propagation algorithm was used. In this method, the characters of plants were measured such as the capsule number per plant, flowering time of 100%, the 1000-seedweight, the plant height and the seed number per capsule. ANN model is an intelligent model and it process like as human brain. This model contains number of nods and it forms layer like input layer and output layer. Such neurons were connected to each other to share the information. Like as ANN and MLR are similar to each other. But MLP constructed output layer based on the linear function.

Statistical modeling:

The model [9] utilised multi linear regression model of weather based yield forecasting for oil palm. The stepwise MLR technique gets input as monthly oil palm yields it act as dependent variable utilizing agro meteorological variables in cumulated time-lag period proceeding to harvest as the independent variables. The main intention of the author is concerned with formulation of appropriate monthly yield forecasting model more than 6 months ahead by describing quantitative relationship between time-lag meteorological variables and FFB yield of the young mature oil palm for the first six harvesting years. In addition, modeling with limited meteorological data is assessed as a quick approach for planters when lacking weather measuring instruments. Table 1 shows the Model with significant predictor variables developed for monthly yield forecast (Y) using a stepwise regression approach.

Table 1. Model with significant predictor variables

Period	Multiple regression equation	R2
2005-2011	Y = -1.256 + 0.023 X1 + 0.007 X2 + 0.006 X3	0.68

ANFIS- Intelligent System:

The model presented adaptive neuro-fuzzy inference system [10] for the prediction of wheat yield. This system utilized energy input values to produce output based on fuzzy sets. Then Artificial Neural Networks model was created for the prediction of wheat yield and both the models were compared. The ANFIS system is the combination of both fuzzy system and ANN. Table 2 tabulates the contribution of input variables to the output for wheat production.

Table 2. Contribution of Input variables

Sensitivity	yield
N	0.09
P2O5	0.10
K2O	0.03
FYM	0.00
Labor	0.06
Diesel	0.06
Electricity	0.96
Seeds	0.01
Biocides	0.01
Water	0.11
Machinery	0.87

The creation of fuzzy rules was used with suitable membership function to predict the wheat yield more effectively. However, ANFIS presents some linearity with respect to some of its parameters. The results illustrated that when the number of inputs for each ANFIS network decreased and simultaneously, the total number of ANFIS networks increased the better results was obtained. The best architecture included five networks at the first stage, two networks at the second stage and one network at final stage.

Data envelopment analysis methodology:

This model explains the introduction of a crop yield prediction model for wetland paddy farms in Malaysia. This model utilized linear production function of Cobb-Douglas equation. This function is used to optimised energy input data was developed by data envelopment analysis (DEA) and benchmarking technology. This model was processed on six energy inputs such as human labour, paddy seeds, fertilizer, chemicals, and fuel. An optimum yield predictive

model was developed by employing linear production function of Cobb-Douglas equation to the optimized energy input data generated through DEA and benchmarking methodology.

The energy inputs considered in the model are human, fuel, machinery, fertilizer, chemicals and paddy seeds. The model is of the form:

$$\ln Y=A+C_1\ln X_1+C_2\ln X_2+C_3\ln X_3+C_4\ln X_4+C_5\ln X_5+C_6\ln X_6 \text{ equ(1)}$$

Where Y = predicted optimum paddy yield (kg/ha), A = intercept (constant), X1, X2, X3, X4, X5 and X6 are respectively the human, fuel, machinery, fertilizer, chemical and seed energy (MJ/ha) and Cs are the model's estimated coefficients.

The model has coefficient of determination (R2) of 0.91, therefore, it could serve as a useful tool for performance appraisal to the paddy farmers in their use of farm inputs. By enabling them to make comparison between actual yield they obtained and the yield they should have by using the inputs optimally.

Adaptive neuro-fuzzy modeling:

The model utilized dynamics of neural networks in crop yield prediction. The main objective of the work is to predict the accurate crop yield based on Adaptive Neuro-Fuzzy Inference System (ANFIS)[11]. The ANFIS system gets different inputs like radiation, temperature, yield, vapor pressure deficit (VPD) and CO_2 for the crop yield prediction and it produce single output is yield. The fuzzy sets were created for input variables. ANFIS has only one output node, the yield. One of the difficult issues in predicting yield is that remote sensing data do not go long back in time. Therefore any predicting effort is forced to use a very restricted number of past years in order to construct a model to forecast future values. The system is trained by leaving one year out and using all the other data. Then evaluate the deviation of this estimate compared to the yield of the year that is left out.The graphs of prediction accuracy are estimated based on the ANN (MLP) and ANFIS models. These include the RMSE values of training error and the testing error of both the models that are considered in Table 3.

Table 3. Accuracy results comparing with MLP

	ANFIS	**ANN (MLP)**
Learning epochs	3	1000
Training error (RMSE)	0.093	0.116
Testing error (RMSE)	0.089	0.118

FL, ANFIS and MLR:

The model utilizes Fuzzy logic (FL), Adaptive Neuro Fuzzy Inference System (ANFIS) and Multiple Linear Regression (MLR)[12] for the purpose of prediction of wheat yield production. In this method get an input as Radiation, biomass, rain and extractable soil water (esw). This data were pre processed by removing missing values, redundant, outliers and inconsistent data in the dataset. The FL, ANFIS and MLR were used to predict the wheat prediction. The ANFIS model was applied for predicting the wheat yield by configuring its parameters. ANFIS uses hybrid training algorithm which is a combination of back propagation and least mean square methods. Its main purpose is to minimize the error due to approximation. When fuzzy inference file (FIS) is trained, number of epochs are used which are used to minimize the error. More the number of epochs lesser will be the error. In ANFIS, epochs are used to adjust the weights in order to minimize the error. From the analysis it is understood that ANFIS model is used to capture non linear relationship between the data variables than FL and MLR.

Fuzzy integrators with ANFIS model:

The model [13] introduces the optimization technique called genetic algorithm in interval type 1 and type 2 fuzzy integrators in ensembles of ANFIS models with the aim of reducing the prediction error of crop yield production. The optimization algorithm of genetic algorithm utilized to predict the chaotic time series Mackey-Glass. It gives better function parameters of Fuzzy Inference System in each integrator. This technique involves five sections first is ensemble of ANFIS, next is training and validation, next one is generate output of every ANFIS in ensemble. Then in the fourth section combines all results of ANFIS. In the final section predict the ensemble of ANFIS learning is

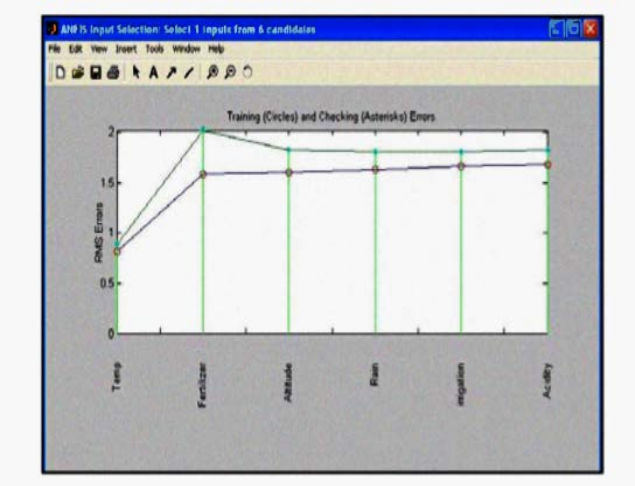
ISSN (online) : 2277-5390
ISSN (Print): 2277-5382

determined. However, this technique is failed to optimize the fuzzy rules in ANFIS it degrade the performance of crop yield prediction. The author reported the optimization of interval type-2 and type-l fuzzy integrators in ensembles of ANFIS models with GAs were used to predict chaotic Mackey-Glass time series, where the results for each ANFIS were evaluated by the method of the root mean square error (RMSE). For the integration of the results of the Ensembles of ANFIS we used the following integration methods: interval type-2and type-l FIS of kind Mamdani and integrator by interval type- 2 FIS optimized. The selection of the time series for the simulations was based on the fact that these time series, which allows comparing results with other approaches such as neural networks and linear regression.

Adaptive neuro fuzzy inference system (anfis):

The model introduces Adaptive Nero Fuzzy Inference System for Jatropha seed yield prediction[14]. There are two main components in Jatropha seed yield prediction. One is the determination of external parameters such as weather, pesticides affect the yield and another one is the determination of internal characteristics that affects the yield of Jatropha plants. These two components were determined by the ANFIS model with less number of input parameters and it produces single output called yield of Jatropha plants. An intensive study has been carried out to identify the design methodology of the research work. The development of Fuzzy Inference System characterized by a large number of input variables (more than five or six).This is very difficult especially in knowledge engineering in order to specify the real input variables, the relative relations, such as the resultant complexity of the rules and database in the system. After detailed investigation with the formers and agriculture experts with the statistical consideration, the six major attributes were considered as design attributes in developing a model. Figure 2 shows the two attribute selection form inputs.

Figure 2. Two attribute selection from inputs

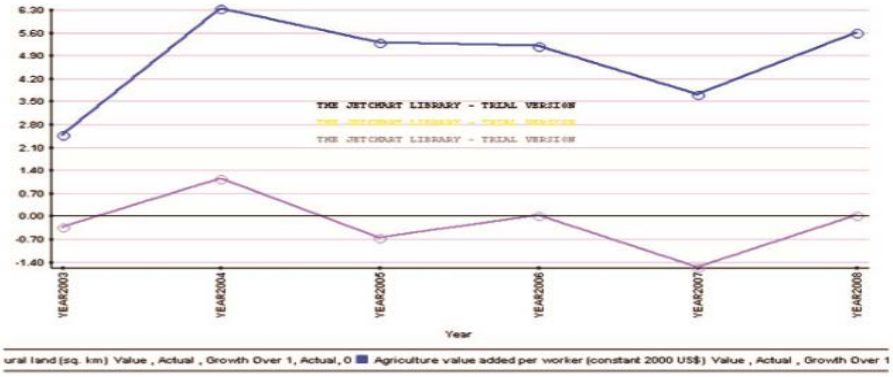


Data analysis methods:

The model [15] utilizes conventional tools or data mining techniques to examine agriculture productivity. In this paper, initially multi dimensional model of data is developed. After that, applied statistical analysis on multi dimensional model and data mining techniques were applied on correlated data to found useful information in agriculture data. In this method, various data collecting agencies were used such as Department of Economics and Statistics, World Bank, IMF, and lot of private agencies like ORG, AC-Nielsen. Figure 3 shows the growth report of agricultural land.

ISSN (online) : 2277-5390
ISSN (Print): 2277-5382

Figure 3. Growth report of AL (Agricultural land (sq. km)) and AGVA (agriculture value added per worker (constant 2000 US$)) in years (2003–2008).



3. Comparison of anfis models in crop yield prediction

Table 4 shows the comparison of ANFIS models in crop yield prediction

Table 4. Comparison of ANFIS models in crop yield prediction

Ref.no	Models	Parameters used	Issues
1	[6] Fuzzy Inference System	The amount of rainfall, the length of the rainy season, the onset of the rainy season.	Accuracy is less.
2	[7]ANFIS	Labour wages (Labour_Cost), Machinary used during farm cultivation (Machinary_Cost), Fertilizer, manure, insecticides and Seed_Cost affects the budget of farming.	More complex to predict the optimized number of input parameters.
3	[8] Artificial neural network	Flowering Time of 10%, Flowering Time of 100%, Seed maturity, Plant height, Capsule Number per plan, 1000-seed weight, Seed Number per capsule, Yield of sesame.	The ANN and MLP methods don't achieve better accuracy
4	[10]ANFIS- intelligent system	Machinery, human labor, diesel fuel, pesticides, chemi cal fertilizers, farmyardmanure (FYM),electricity, water for irrigation and seeds.	ANFIS presents linearity with respect to some of its parameters.
5	[11]Data Envelopment Analysis Methodology	Farm inputs (human labour, machinery, fuel, seeds, fertilizer and pesticides) involving six field operations (tillage, seeding, fertilizing, spraying, harvesting and slashing operations) from the 40 farm lots	It does not give efficient result in all type of forms.
6	[11]Adaptive Neuro-Fuzzy Modeling	Temperature, CO_2, vapor pressure deficit (VPD), yield, and radiation.	Future values were constructed based on the very restricted number of input values. The ANFIS system is vulnerable to dimensionality of data.
7	[12]A Parameter based ANFIS Model	Biomass, esw, Rain and Radiation	In FL,ANFIS and MLR models four parameters such as biomass, esw, rain and radiation are only considered.
8	[13] Fuzzy Integrators with ANFIS Models	A randomly generated population of N number of inputs.	ANFIS Mamdani technique failed to optimize the fuzzy rules in ANFIS it degrade the performance of crop yield prediction.
9	[14] Adaptive Neuro Fuzzy Inference System (ANFIS)	Irrigation, temperature, rainfall, altitude, acidity, fertilizers, land size, soil property	The fuzzy inference system method input selection is depends on the heuristics.

4. Conclusion

The crop yield prediction can be done with various models but Adaptive Neuro Fuzzy Inference System (ANFIS) model handles the non-linear relationship between the data than the other models for the prediction of crop yield when the input attributes are increased by adding all the essential parameters needed for the growth of the crop. This process involves enhancing the essential inputs and adding them to the ANFIS model by increasing the input intake and feeding all the essential inputs into the model. Thus the processing of all those attributes will increase the classification accuracy many times than the normal prediction models.

5. References

[1] Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua, Wang Juyun. Prediction of crop yield using big data. *8th International Symposium onComputational Intelligence and Design (ISCID).*2015;1, 255-260.

[2] Monali Paul, Santosh K. Vishwakarma, Ashok Verma. Analysis of soil behaviour and prediction of crop yield using data mining approach. *Computational Intelligence and Communication Networks (CICN).* 2015; 766-771.

[3] Subhadra Mishra, Debahuti Mishra, GourHariSantra.Applications of machine learning techniques in agricultural crop production: a review paper.*Indian Journal of Science and Technology*.2016, 9(38), 1-14.

[4] Aliyu Muazu, Azmi Yahya, W.I.W. Ishak, S. Khairunniza-Bejo. Yield prediction modeling using data envelopment analysis methodology for direct seeding, wetland paddy cultivation. *Agriculture and Agricultural Science Procedia*. 2014; 2, 181-190.

[5] Kadir, MuhdKhairulzaman Abdul, MohdZakiAyob, NadarajMiniappan. Wheat yield prediction: Artificial neural network based approach. *4th International Conference onEngineering Technology and Technopreneuship (ICE2T).* 2014; 161-165.

[6] Agus Buono. An implementation of fuzzy inference system for onset prediction based on Southern Oscillation Index for increasing the resilience of rice production against climate variability. *Advanced Computer Science and Information Systems (ICACSIS).* 2012; 281-286.

[7] O.K. Chaudhari, P.G. Khot, K.C. Deshmukh, N.G. Bawne. Anfis based model in decision making to optimize the profit in farm cultivation. *International Journal of Engineering Science*. 2012; 4(2), 442-448.

[8] Samad Emamgholizadeh, M. Parsaeian, Mehdi Baradaran. Seed yield prediction of sesame using artificial neural network. *European Journal of Agronomy*. 2015;68, 89-96.

[9] Keong, Yong Kian, Wong Mum Keng. Statistical modeling of weather-based yield forecasting for young mature oil palm. *APCBEE Procedia.* 2012; 4, 58-65.

[10] Benyamin Khoshnevisan, Shahin Rafiee, Mahmoud Omid, Hossein Mousazadeh. Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. *Information Processing in Agriculture*. 2014; 1(1), 14-22.

[11] Kefaya Qaddoum, Evor Hines, Daciana Illiescu. Adaptive neuro-fuzzy modeling for crop yield prediction.*Recent Researches in Artificial Intelligence, Knowledge Engineering and D atabases.* 2011; 16(7), 1-6.

[12] Aditya Shastry, H.A. Sanjay, Madhura Hegde. A parameter based ANFIS model for crop yield prediction. *Advance Computing Conference (IACC).* 2015; 253-257.

[13] Jesus Soto, Patricia Melin, Oscar Castillo. Optimization of interval type-2 and type-1 fuzzy integrators in ensembles of ANFIS models with Genetic Algorithms. *World Congress onNature and Biologically Inspired Computing (NaBIC)*.2013; 41-46.

[14] S.P. Srinivasan, P. Malliga. A new approach of adaptive Neuro Fuzzy Inference System (ANFIS) modeling for yield prediction in the supply chain of Jatropha. *IEEE 17th International Conference onIndustrial Engineering and Engineering Management (IE&EM).*2010, 1249-1253.

[15] Swati Hira, P.S. Deshpande. Data analysis using multidimensional modeling, statistical analysis and data mining on agriculture parameters. *Procedia Computer Science*. 2015; 54, 431-439.

The Publication fee is defrayed by Indian Society for Education and Environment (iSee). www.iseeadyar.org

Citation:
K. Menaka, N. Yuvaraj. A survey on crop yield prediction models. *Indian Journal of Innovations and Developments.* 2016; 5(12), December.