ANFIS BASED CROP YIELD PREDICTION MODEL

Menaka .K
M.E. (Computer Science and Engineering)
KPR Institute of Engineering and Technology
Arasur, Coimbatore

N.Yuvaraj M.E, MISTE,(Ph.D),
Assistant professor,
Department of Computer Science and Engineering
KPR Institute of Engineering and Technology
Arasur.

Abstract— The Prediction of crop yield is more important for increasing the yield of the crop. Many advanced technologies are incorporated in the agricultural processes which enhances the crop yield production efficiency. The process of predicting the crop yield can be done by taking agriculture data which helps to analyze and make important decisions during cultivation. Accurate prediction results will help the agriculturalists to face and solve various problems related to farming. The models of PSO and ANFIS are used for predicting the yield of crops by considering all the essential input parameters which plays important role in the growth and crop yield. In this paper the implementation of crop yield prediction model-ANFIS is discussed with various essential input attributes. Those attributes are pre-processed using feature selection algorithm (PSO). The performance of the proposed model can be estimated by comparing the results with the prediction models like SVM and ANN.

Keywords—*ANFIS; PSO; Prediction; Crop yield*

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 is based on various kinds of data collected and extracted by using data mining techniques different sources which are useful for growth of the crop .It is a 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. They can contract their crop prior to harvest, if they have an idea of the amount of yield they can expect which gives 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 do better planning the logistics of their business as the benefit from yield predictions.

2) CROP YIELD PREDICTION

2.1) Crop yield

Crop yield is a branch of agriculture that deals with growing crops for use as food and fiber. Crop yield includes rice, wheat, maize, grains, cotton, tobacco, fruits, vegetables, nuts and plants. Different crops grow best in different areas of the country. Crop producers usually work from sunrise to sunset during planting and harvesting seasons. They sell the crops they've produced, plan crops for the next season and repair machinery.

2.2) Need for Prediction in Agriculture

The demand for world agriculture output will grow exponentially over coming decades due to world population growth and expanding world economies. At the same time, the agriculture sector will be impacted by changes in many factors that will challenge the productivity of the world's agriculture resources. To meet this expanding world demand, agriculture must become more adept at anticipating changes and variations and finding ways of adapting to these changes. Below is a discussion of the needs of farmers and agribusinesses for quality information relating to predictions.

The prediction information has the potential to reduce the impact of adverse events as shown in Fig.1.This will occur because the advance notice will allow decision makers the opportunity to implement plans to minimize the impact of adverse events and find opportunities within favourable events.

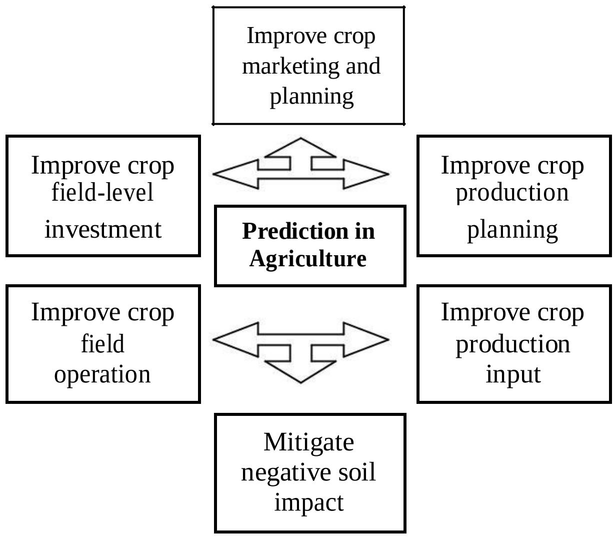


Figure.1 Crop yield prediction needs

2.3) Attribute Selection

The dependable attributes can be difficult to find. Various methods of predicting and modeling of the crop yields are used from the past with varying success. The planning made by farmers in advance gets affected due to various factors. 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. 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. Hence all the essential parameters should be considered during the crop yield prediction, which helps in increasing the accuracy of the prediction of crop and crop yield.

2.4) 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.

3) INPUT ATTRIBUTES

Input Attributes

The input attributes are the parameters that are important for the growth of the crop. These inputs can be based on climatic condition, topographical factors, soil properties, biotic factors etc. Each of them plays an important role in the process of determining the crops to be cultivated and the amount of the crop yield.

3.1) Soil Attributes

Soil attributes are the inputs that are to be considered from the soil. For example Ph of soil, nutrients of soil, soil type, organic matter etc.

3.1.1) Ph of Soil

Technically, pH is defined as the negative (-) log or base 10 value of the concentration of hydrogen ions (H+). Neutral pH value will be seen in pure water, which has the value 10 to the minus 7 concentration of H+ ions (10-7 [H+]). This concentration of the pH is expressed as 7.

The Ph value above the neutral Ph means the H+ ion concentration is lower than the value 7 and the tested component is said to be alkaline and there are more hydroxyl (OH-) ions present than H+ ions. The Ph value below the neutral pH means the H+ ion concentration is greater than that of more hydroxyl (OH-) ions and the solution is acidic. Soils Below a pH of 5 are considered acidic, and below a pH of 4 are very acidic. Conversely, soils with a pH of 7.5 and above are considered alkaline and above a pH of 8 the soil is said to be very alkaline Table 1 explains the pH values of the soil.

Table.1pH of soil

Soil Acidity	Nitrogen	Phosphorus	Potassium
Extremely Acid -4.5 pH	30%	23%	33%
Very Strong Acid- 5.0 pH	53%	34%	52%
Strongly Acid — 5.5 pH	77%	48%	77%
Medium Acid — 6.0 pH	89%	52%	100%
Neutral — 7.0 pH	100%	100%	100%

3.1.2) Rainfall

The amount of rainfall in that area is the most important determinant of the type of crop raised. Wet crops are cultivated in wet zone and dry crops in the dry zone. High temperature and heavy rainfall are required for the cultivation of crops like rice, jute, sugarcane, etc. Therefore these crops are mostly cultivated in summer. Moderate temperature and rainfall are required for the cultivation of crops like wheat, barley etc. Therefore these are cultivated in winter. Table.2 explains the cardinal rainfall.

Table.2 Cardinal Rainfall

Crop's Seed	RAINFALL (OC)		
	Minimum	Optimum	Maximum
Rice	175	225	300
Maize	60	80	110
Wheat	110	125	150

3.1.3) Temperature

Temperature greatly influences germination of seeds. The inhibitions phase of germination is primarily physical and shows low sensitivity to temperature as shown in Table.3. The subsequent phases of germination are temperature dependent due to biochemical processes involved. Cardinal temperature of different crop's seed germination

Table.3 Cardinal Temperature

Crop's Seed	Temperature (OC)		
	Minimum	Optimum	Maximum
Rice	11	32	41
Maize	9	33	42
Wheat	4	25	32

3.1.4) Nutrients

A bisection is made between those nutrients required in greater quantities are said to be macronutrients, and those elements required in smaller quantities are said to micronutrients. This partition of nutrients does not mean that one nutrient is more important than the other nutrient, it is just that they are required in different amounts and concentrations. This is shown in the Table 4

Macronutrients: N, K, Ca, Mg, P, and S, and
Micronutrients: Cl, Fe, B, Mn, Zn, Cu, Mo, and Ni

Table.4 General requirements of plants

Nutrient	Low	Sufficient	High
Nitrogen (N) %	1.25	1.75 - 3.0	3.0 - 4.0
Phosphorous (P) %	0.15	0.26 - 0.5	0.5 - 0.8
Potassium (K) %	1.0	1.5 - 3.0	3.0 - 5.0
Sulphur (S) %	0.1	0.15 - 0.40	0.40 - 0.8
Calcium (Ca) %	0.10	0.2 - 1.0	1.0 - 1.5
Magnesium (Mg) %	0.1	0.15 - 0.50	0.5 - 1.0
Zinc (Zn) ppm	10	15 – 70	70 - 150
Copper (Cu) ppm	2.3	3.7 – 25	25 - 50
Iron (Fe) ppm	15	20 – 250	250 - 500
Manganese (Mn)	10	15 – 100	100 - 250
Boron (B)	3	5 – 25	25 - 75
Molybdenum (Mo)	0.01	.03 – 5	5 - 10

3.2) Major Crops

Some of the major crops that are cultivated in wide range depend on the soil properties, rainfall, nutrients, temperature and some other factors. In India, major crops are rice, wheat, maize, jute, sugarcane etc. Those crops have their own external factors which are needed for their high yield.

3.2.1) Rice

Rice is predominantly a Kharif or crop. The one third of total cultivated area of India is rice. For more than half of the Indian population it provides food. With low labour costs and high rainfall countries and regions, rice cultivation is well-suited, as it is labour-intensive to cultivate and requires ample water. However, with the use of water-controlling terrace systems, rice can be grown practically anywhere, even on a steep hill or mountain area. Centuries of trade and exportation are made in Asia and certain parts of Africa as its parent species are their native.

3.2.2) Wheat

In India, Wheat is the second most important crop after Rice. It's a Rabi Crop. In north and north western India, it is the staple food. It's a winter crop p and needs low temperature. The temperature between 10-15°C is the ideal temperature for wheat cultivation and at the time of harvesting the temperature should be 21-26°C. In rainfall less than 100 cm and more than 75 cm wheat thrive well. Well drained fertile loamy soil and clayey soil are the most suitable soil for cultivation of wheat. Plain areas are also most suitable. The wheat crop cultivation may need less labour and is highly mechanization oriented.

3.2.3) Maize

Maize production depends on the correct application of production inputs that will sustain the environment as well as agricultural production. These inputs are adapted cultivars, soil tillage, inter alia, fertilization, disease control, plant population, insect and weed, harvesting, marketing and financial resources

4) REVIEW OF CROP YIELD PREDICTION MODELS

4.1) Attribute Selection

The dependable attributes can be difficult to find. With varying levels of success, different methods of predicting and modeling of the crop yields have been used from the past. 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 [17] 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.

4.2) 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 [2]. 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.

4.3) 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 [13].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 considered and selected depend on the heuristics. It also decreases the performance by degrading the efficiency of the prediction process.

4.4) Statistical Modeling

The model [14] 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 first six harvesting years FFB yield of the young mature oil palm. As shown in Table.5 quick approach for planters modeling with limited meteorological data is assessed when lacking weather measuring instruments.

Table.5 Model with significant predictor variables developed for monthly yield forecast (Y) using a stepwise regression approach

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

4.5) ANFIS- Intelligent System

The model presented adaptive neuro-fuzzy inference system [15] 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 as discussed in Table.6. The ANFIS system is the combination of both fuzzy system and ANN.

Table 6 Contribution of input variables to the output for wheat production

Sensitivity	**Yield**
N	0.09
P2O5	0.10
K2O	0.03
FYM	0.00
Labour	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.

4.6) 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) [16] 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:

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

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.

4.7) 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) [19]. The ANFIS system gets different inputs like radiation, temperature, yield, vapour 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 which is 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 this Table.7.

Table 7 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

4.8) Adaptive Neuro Fuzzy Inference System (ANFIS).

The model introduces Adaptive Nero Fuzzy Inference System for Jatropha seed yield prediction [24]. 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 which is explained in Fig.2

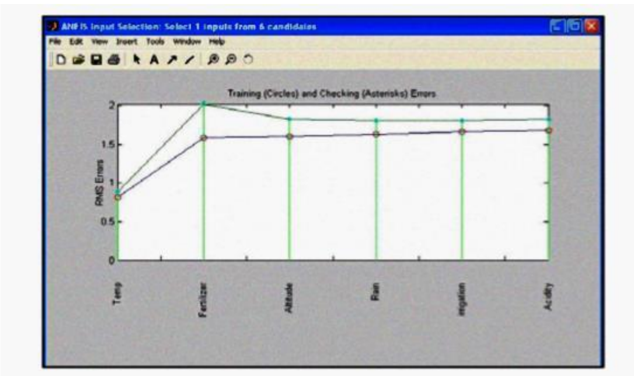


Figure.2 Two attribute selection from inputs

5) PROPOSED METHODOLOGY

The proposed model is used for the crop and crop yield prediction with high accuracy and precision. The models like PSO and ANFIS are used for the highly efficient classification. As mentioned in Fig.3, PSO is used for the feature selection process.

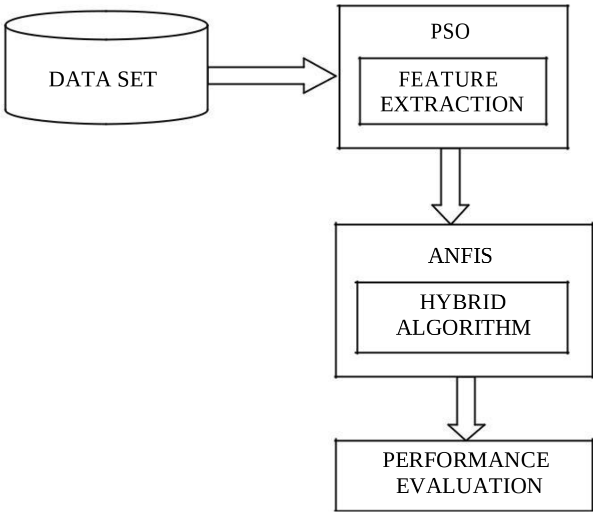


Figure.3 Architecture of Proposed Model

This model involves reduction of unwanted input parameters by considering only the parameters which are essential for the crop growth and which plays vital role in improving the crop yield. The reduced amounts of parameters are given as the inputs to the ANFIS model. Anfis model classifies the data and it will conclude by the crop which is suitable for the cultivation, which is based on the value of the inputs and the crop's threshold values **.**

5.1) System Implementation

The models which is proposed. Includes three modules. They are as follows:

List of Modules

1. Feature Extraction
2. ANFIS Model
3. Performance Evaluation

5.1.1) Feature Extraction

It describes the input parameters considered for the crop yield prediction. Input parameters which are biomass, esw, radiation, amount of rainfall during sowing, pH value of soil, soil temperature, Organic Carbon, percentage of Phosphorous, Potassium, Iron, Zinc, Manganese, Copper and Boron along with soluble salts, amount of organic matter and carbon-nitrogen ratio, numbers of microorganisms and soil fauna are considered for crop yield prediction.

The feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations

When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then it can be transformed into a reduced set of features (also named a feature vector . Feature selection is the process of determining a subset of the initial features. The relevant information that are expected to contain in the input data are selected, so that instead of the complete initial data desired task can be performed by using the reduced representation.

5.1.1.1) Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is an algorithm modeled on swarm intelligence, in a search space, or model it finds a solution to an optimization problem and predict social behavior in the presence of objectives. The PSO is a population-based stochastic computer algorithm, modeled on swarm intelligence. Swarm intelligence is based on social-psychological principles and it provides insights into social behavior, as well as it contributes to many engineering applications.

. A problem is given, and a specific method to evaluate a proposed solution to that is by finding and forming a fitness function. A communication structure or social network is also defined, assigning neighbours for each individual particle to interact with each other. Then a population of individual particles can be defined by making the random guesses at the problem solutions which can be initialized. These solutions to the individual are candidate solutions. They are also known as the particles, hence the name particle swarm is obtained. For the improvement of these candidate solutions, iterative process is to be done. The particle’s fitness of the candidate solutions is iteratively evaluated and the location where they had their best success is remembered. The best solution of the individual particle is called the particle best or the local best. Each particle makes this information available to their neighbours.

Each particle can also able to see where their neighbours have had success. Movements through the search space are guided by these successes, with the population usually converging, by the end of a trial, on a problem solution better than that of non-swarm approach using the same methods. A candidate solution to the optimization problem IS represented by each particle. The position of a particle is influenced by the best position visited by the particle itself i.e. the particle’s own experience and the position of the best particle in its neighbourhood i.e. the experience of their neighbouring particles. When the neighbourhood of a particle is the entire swarm, the global best particle is considered as the best position in the neighbourhood and the resulting algorithm is referred to as the gbest PSO. , the algorithm is generally referred to as the lbest PSO. When smaller neighbourhoods are used, the performance of each particle is measured using a fitness function that varies depending on the optimization problem.

Each Particle in the swarm is represented by the two of their characteristics namely :

1. The current position of that particular particle
2. The current velocity of that particular particle

The particle swarm optimization which is one of the latest evolutionary optimization techniques that makes searches among the population of particles participating.

Each particle is considered individually in evolutionary algorithm. Each particle has two vectors namely position and velocity vectors, updating both the position vector and velocity vector by moving through the problem space.

$$V_i^{k+1}=wV_i^k+c_1rand_1()x(pbest_i-s_i^k)+c_2rand_2()x(gbest-s_i^k) \quad (1)$$

$$S_i^{k+1}=S_i^k+V_i^{k+1} \quad (2)$$

Where, V_i^k is the velocity of I at iteration k, S_i^k is the current position of I at iteration k. c_1 and c_2 are positive constants and $rand_1$ and $rand_2$ are uniformly distributed random number in [0,1]. The velocity vector is range of [-Vmax,Vmax]. In velocity updating equations (1) and (2) creates the new velocity are,

1. Inertia term, forces the particle to move in the same direction as before by adjusting the old velocity.
2. The particle to go back to the previous best position as it is forced by Cognitive term (Personal best).
3. The particle moves to the best previous position of its neighbours as it is forced by the Social Learning term.

5.1.1.2) PSO Algorithm

1. Initialize location and velocity of each of the particle
2. Repeat the same
3. For each of the particle
4. Evaluate objective function for each of the particle
5. For each particle
6. Update the best solution
7. Update the best global solution
8. For each of the particle
9. Update their velocity
10. Compute the new locations of the articles
11. Until finished()

Fig.4 shows the step by step working of the particle Swarm Optimization algorithm. The final output of the process will be the optimized one that helps in Feature selection process. The number of essential parameters will be filtered or reduced by using this method.

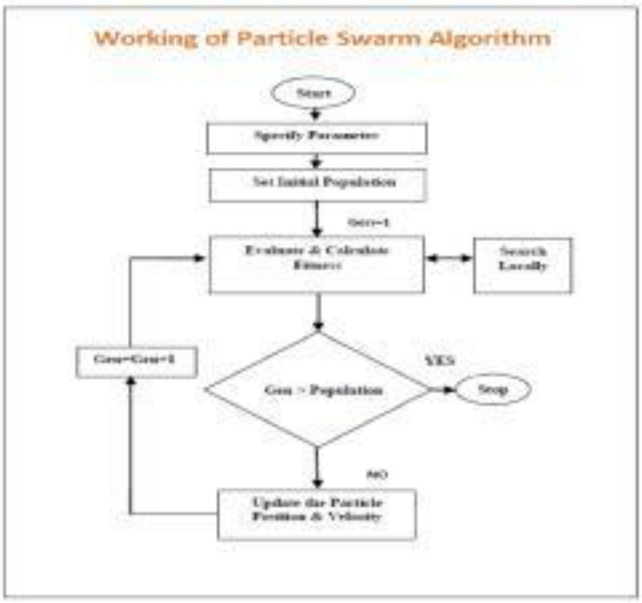


Figure.4 Working of PSO

5.1.2) 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 .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 considered and selected depend on the heuristics. It also decreases the performance by degrading the efficiency of the prediction process.

5.1.2.1) ANFIS structure

For simplicity, it is assumed that the fuzzy inference system under consideration has two inputs and one output. The rule base contains the fuzzy if-then rules of Takagi and Sugeno's type as follows: If x is A and y is B then z is f(x,y) where A and B are the fuzzy sets in the antecedents and z = f(x, y) is a crisp function in the consequent. Usually f(x, y) is a polynomial for the input variables x and y. But it can also be any other function that can approximately describe the output of the system within the fuzzy region as specified by the antecedent.

A zero order Sugeno fuzzy model is formed which may be considered to be a special case of Mamdani fis where each rule consequent is specified by a fuzzy singleton when f(x, y) is a constant. If f(x, y) is taken to be a first order polynomial a first order Sugeno fuzzy model is formed. For a first order two rule Sugeno fuzzy inference system, the two rules may be stated as:

Rule 1: If x is A1 and y is B1 then f1 = p1x + q1y + r1 Rule
2: If x is A2 and y is B2 then f2 = p2x + q2y + r2

Here type-3 fuzzy inference system proposed by Takagi and Sugeno is used. A linear combination of the input variables added by a constant term is the output of each rule in this inference system. The final output is the weighted average of each rule's output. The corresponding equivalent ANFIS structure is shown in Fig.5.

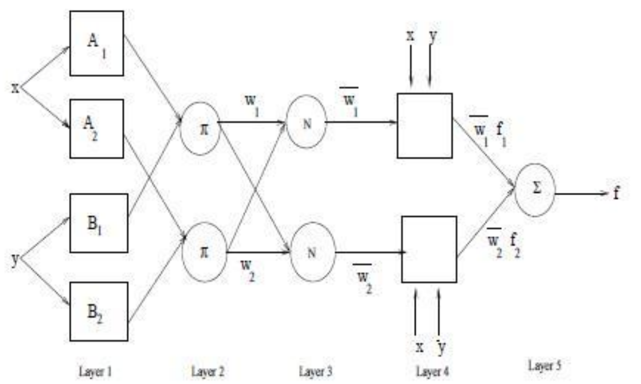


Figure.5 ANFIS Structure

Layer 1

$O_{l,i}$ is the output of the i^{th} node of the layer l. Every node i in this layer is said to be an adaptive node with a node function for each of its node.

$$O_{1,i} = \mu A_i(x) \quad for\ i = 1,2$$
$$O_{1,i} = \mu B_{i-2}(x) \quad for\ i = 3,4$$

x (or y) is the input node i and A_i (or B_{i-2}) is a linguistic label associated with this node. Therefore $O_{1,i}$ is the membership grade of a fuzzy set (A_1, A_2, B_1, B_2). Typical membership function:

$$\mu_A(x) = \frac{1}{1 + \left|\frac{x - c_i}{a_i}\right| 2b_i}$$

Where a_i, b_i, c_i is the parameter set. Parameters are referred to as premise parameters.

Layer 2

Every node in this layer is a fixed node labelled.

The output of this layer is the product of all the input signals supplied from the previous layer.

$$O_{2,i} = w_i = \mu A_i(x).\mu B_i(y), \quad for\ i = 1,2$$

The fire strength of the rule is used to represent each node. AND operator performing any other T-norm operator that can be used.

Layer 3

Every node in this layer is a fixed node labelled Norm. The i^{th} node calculates the ratio of the i^{th} rule's firing strength to the sum of all rule's firing strengths.

$$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2}, i = 1,2$$

Outputs are called normalized firing strengths.

Layer 4

Each and every individual node i in this layer 4 is an adaptive node with a respective node function:

$$O_{i,4} = \overline{w_i} f_i = \overline{w_i}(p_i + q_i y + r_i)$$

Where w is the normalized firing strength from the previous layer $\{p_i, q_i, r_i\}$ is the parameter set of this adaptive node. These parameters are referred to as consequent parameters.

Layer 5

This layer has a single node, which has the fixed labelled sum of overall output as the summation of all the individual inputs form the previous layer.

$$overall\ output = O_{5,i} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

Hybrid Learning Algorithm

The ANFIS can be trained by a hybrid learning algorithm. The least-squares method is implemented to identify the consequent parameters on the layer 4 in this forward pass the algorithm. In the backward pass the errors are propagated backward and the premise parameters are updated by gradient descent.

ANFIS structure (for input 1 to 4) is shown in Fig.6 ANFIS1 is having 4 inputs and only one output.

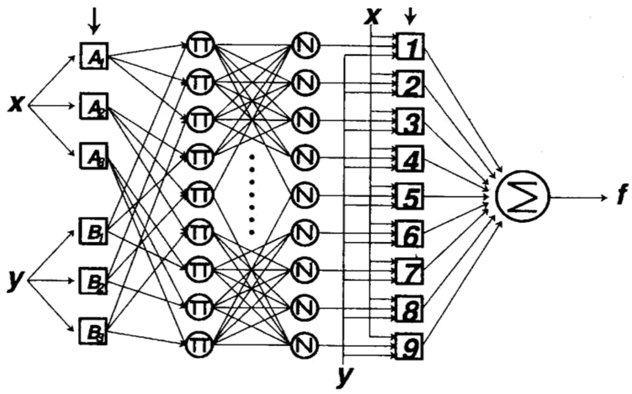


Figure.6 ANFIS structure (for input 1 to 4)

6) RESULT AND DISCUSSION

The performance are calculated by the comparison parameter such that precision, recall, F-measure, Accuracy.

6.1) Precision

Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage. Precision is the agreement among several determinations of the same quantity. Precision is defined as the measurement of fraction from true value and its scatter.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive + False\ Positive\ (FP)}$$

Fig.7 shows that the comparison of SVM, ANN and ANFIS Classification algorithms in terms of precision values . The result shows that the ANFIS approach provides higher precision than other algorithms.

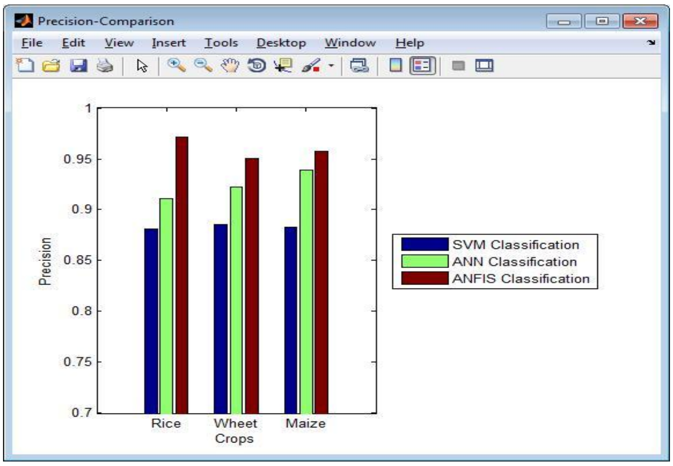


Figure.7 Comparison of Precision

6.2) Recall

Recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage. Recall in this context is also referred to as the true positive rate or sensitivity. Recall is defined as the statistical measure such that the fraction of relevant material is returned by the search.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative\ (FN)}$$

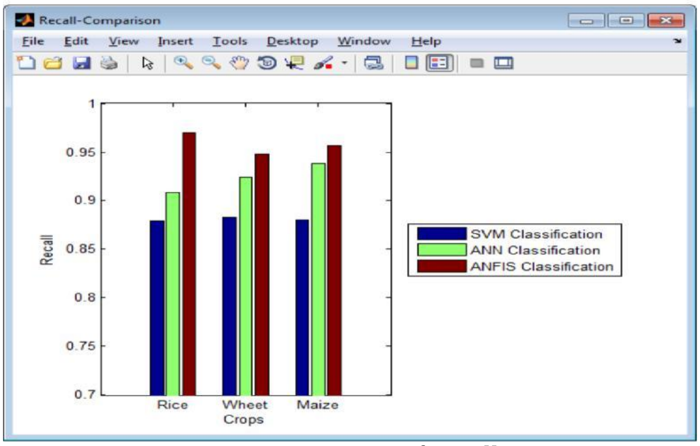


Figure.8 Comparison of Recall

Fig .8 shows that the comparisons of SVM, ANN and ANFIS Classification algorithms in terms of recall values. The result shows that the ANFIS approach provides higher recall value than other algorithms

6.3)F-Measure

F-measure is a measure of a test's accuracy. It considers both the precision *p* and the recall *r* of the test to compute the score: *p* is the number of correct positive results divided by the number of all positive results, and *r* is the number of correct positive results divided by the number of positive results that should have been returned-measure is calculated from the precision and recall value. It is calculated as:

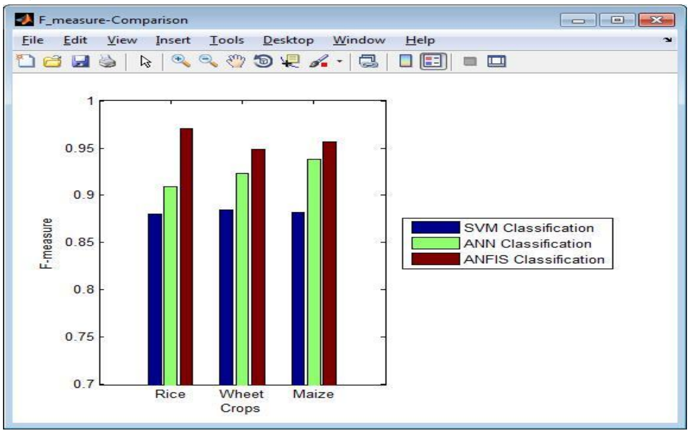


Figure.9 Comparison of F-Measure

Fig.9 shows that the comparison of SVM, ANN and ANFIS Classification algorithms in terms of F-Measure values.

$$f - measure = 2 \times \left(\frac{precision \times recall}{precision + recall}\right)$$

The result shows that the ANFIS approach provides higher F-Measure value than other algorithm

6.4) Accuracy

The accuracy is the agreement between an experimental value, or the average of several determinations of the value, with an accepted or theoretical (―true‖) value for a quantity. Accuracy is usually expressed as a percent difference .It is the proportion of true results (both true positives and true negatives) among the total number of cases examined. Accuracy can be calculated from formula given as follows:

$$Accuracy = \frac{\text{TP} + \text{True Negative (TN)}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Fig.10 shows that the comparison of SVM, ANN and ANFIS Classification algorithms in terms of accuracy values. The result shows that the ANFIS approach provides higher accuracy than other algorithms.

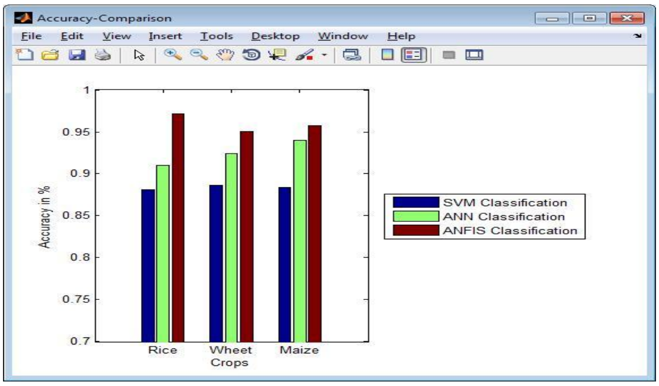
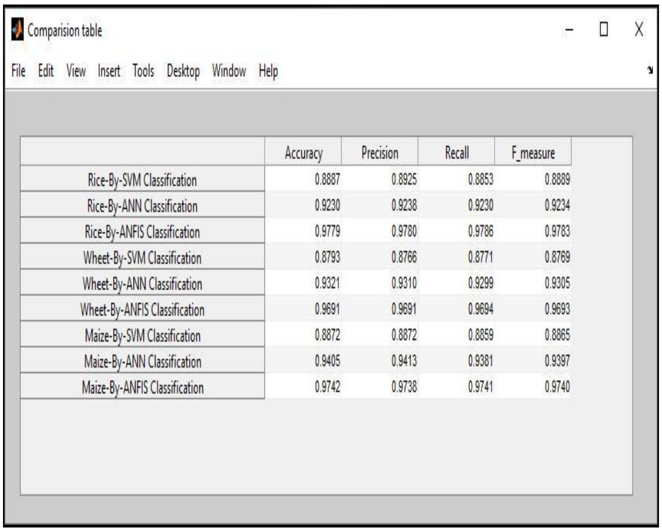


Figure.10 Comparison of Accuracy

6.5) Performance Evaluation

The performance evaluation of the prediction models are compared with the other similar model to calculate and find out the difference.

Fig.11 shows the performance evaluation of the prediction models- ANFIS, SVM and ANN. The performance evaluation parameters like accuracy, precision, recall and F-measure are calculated. From the result, the prediction performance of the ANFIS model is higher.


Comparision table
File Edit View Insert Tools Desktop Window Help

	Accuracy	Precision	Recall	F_measure
Rice-By-SVM Classification	0.8887	0.8925	0.8853	0.8889
Rice-By-ANN Classification	0.9230	0.9238	0.9230	0.9234
Rice-By-ANFIS Classification	0.9779	0.9780	0.9786	0.9783
Wheet-By-SVM Classification	0.8793	0.8766	0.8771	0.8769
Wheet-By-ANN Classification	0.9321	0.9310	0.9299	0.9305
Wheet-By-ANFIS Classification	0.9691	0.9691	0.9694	0.9693
Maize-By-SVM Classification	0.8872	0.8872	0.8859	0.8865
Maize-By-ANN Classification	0.9405	0.9413	0.9381	0.9397
Maize-By-ANFIS Classification	0.9742	0.9738	0.9741	0.9740

Figure.11 Comparison with similar models

7) CONCLUSION

In our proposed system, classification performance is improved by considering various attributes using ANFIS method. Those attributes are pre-processed by using PSO for the feature selection process by which essential parameters to be considered for the crop yield can be selected. ANFIS combines Neural Network with fuzzy classification algorithm in which fuzzy supports large data sets and also enhances the classification result. It can handle 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.

Prediction results of the model are compared with some prediction algorithms such as SVM and ANN based on the Precision, Recall, F-measure and accuracy values. From the comparison result, we conclude that, by adding and processing all the essential attributes, the classification accuracy will increase many times than the normal prediction models. Hence, ANFIS can be used to predict the yield of the crops more accurately than any other model.

REFERENCES

1. *Agus Buono and Mushthofa. "An Implementation of Fuzzy Inference System for Onset Prediction Based on Southern Oscillation Index for Increasing the Resilience of Rice Production Against Climate Variability", in Proc. IEEEICACSIS, Dec. 2012, pp. 281-286.*
2. Al-Jamimi, H. A. (2016, August). Toward comprehensible software defect prediction models using fuzzy logic. In *Software Engineering and Service Science (ICSESS), 2016 7th IEEE International Conference on* (pp. 127-130). IEEE.
3. Ananthara, M. G., Arunkumar, T., & Hemavathy, R. (2013, February). CRY—an improved crop yield prediction model using bee hive clustering approach

for agricultural data sets. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on* (pp. 473-478). IEEE.

4. Cao, W., Tang, L., Zhu, Y., Pan, J., Li, W., & Chen, B. (2006, November). Development of growth model-based decision support system for crop management. In *Plant Growth Modeling and Applications, 2006. PMA'06. Second International Symposium on* (pp. 181-184). IEEE.
5. Chaudhari, O. K., Khot, P. G., Deshmukh, K. C., & Bawne, N. G. (2012). Anfis based model in decision making to optimize the profit in farm cultivation. International Journal of Engineering Science and Technology, 1(4), 442-448.
6. Cheng-lin, L., & Jian-jun, W. (2008, July). Crop drought monitoring using MODIS NDDI over mid-territory of China. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International* (Vol. 3, pp. III-883). IEEE.
7. Emamgholizadeh, S., Parsaeian, M., & Baradaran, M. (2015). Seed yield prediction of sesame using artificial neural network. European Journal of Agronomy, 68, 89-96.
8. Fan, W., Chong, C., Xiaoling, G., Hua, Y., & Juyun, W. (2015, December). Prediction of Crop Yield Using Big Data. In *2015 8th International Symposium on Computational Intelligence and Design (ISCID)* (Vol. 1, pp. 255-260). IEEE.
9. Hira, S., & Deshpande, P. S. (2015). Data Analysis using Multidimensional Modeling, Statistical Analysis and Data Mining on Agriculture Parameters. *Procedia Computer Science*, *54*, 431-439
10. Jawad, F., Choudhury, T. U. R., Sazed, S. A., Yasmin, S., Rishva, K. I., Tamanna, F., & Rahman, R. M. (2016, June). Analysis of Optimum Crop Cultivation using Fuzzy System. In *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on* (pp. 1-6). IEEE.
11. Jayashree, L. S., Rajathi, N., & Thirumal, A. (2016, November). Precision agriculture: On the accuracy of multilevel and clustered ANFIS models for sugarcane yield categorization. In *Region 10 Conference (TENCON), 2016 IEEE* (pp. 1983-1987). IEEE.
12. Jin, S., Ge, J., & Peng, J. (2016, August). Terrorism risk assessment using hierarchical bidirectional fuzzy rule interpolation. In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2016 IEEE 15th International Conference on* (pp. 403-410). IEEE.
13. Kadir, M. K. A., Ayob, M. Z., & Miniappan, N. (2014, August). Wheat yield prediction: Artificial neural network based approach. In *Engineering Technology and Technopreneuship (ICE2T), 2014 4th International Conference on* (pp. 161-165). IEEE.
14. Keong, Y. K., & Keng, W. M. (2012). Statistical modeling of weather-based yield forecasting for young mature oil palm. APCBEE Procedia, 4, 58-65.
15. Khoshnevisan, B., Rafiee, S., Omid, M., & Mousazadeh, H. (2014). Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. Information processing in agriculture, 1(1), 14-22.
16. Muazu, A., Yahya, A., Ishak, W. I. W., & Khairunniza-Bejo, S. (2014). Yield prediction modeling using data envelopment analysis methodology for direct seeding, wetland paddy cultivation. Agriculture and Agricultural Science Procedia, 2, 181-190.
17. Murmu, S., & Biswas, S. (2015). Application of fuzzy logic and neural network in crop classification: a review. *Aquatic Procedia*, *4*, 1203-1210.
18. Paul, M., Vishwakarma, S. K., & Verma, A. (2015, December). Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach. In *Computational Intelligence and Communication Networks (CICN), 2015 International Conference on* (pp. 766-771). IEEE.
19. Qaddoum, K., Hines, E., & Illiescu, D. (2011). Adaptive neuro-fuzzy modeling for crop yield prediction. parameters, 16(17), 18.
20. Romero, J., & Smith, K. (2016, April). Crop profit optimization for farmers. In *Systems and Information Engineering Design Symposium (SIEDS), 2016 IEEE* (pp. 289-291). IEEE.
21. Shastry, A., Sanjay, H. A., & Hegde, M. (2015, June). A parameter based ANFIS model for crop yield prediction. In Advance Computing Conference (IACC), 2015 IEEE International (pp. 253-257). IEEE.
22. Soria-Ruiz, J., Fernandez-Ordonez, Y., McNairm, H., & Bugden-Storie, J. (2007, July). Corn monitoring and crop yield using optical and RADARSAT-2 images. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International* (pp. 3655-3658). IEEE.
23. Soto, J., Melin, P., & Castillo, O. (2013, August). Optimization of interval type-2 and type-1 fuzzy integrators in ensembles of ANFIS models with Genetic Algorithms. In Nature and Biologically Inspired Computing (NaBIC), 2013 World Congress on (pp. 41-46). IEEE.
24. Srinivasan, S. P., & Malliga, P. (2010, October). A new approach of adaptive Neuro Fuzzy Inference System (ANFIS) modeling for yield prediction in the supply chain of Jatropha. In Industrial Engineering and Engineering Management (IE&EM), 2010 IEEE 17Th International Conference on (pp. 1249-1253). IEEE.