Journal of Biological Control, 25 (2): 134–142, 2011

Research Article

Neural-network classifier for the prediction of occurrence of *Helicoverpa armigera* (Hübner) and its natural enemies

M. PRATHEEPA[1*#], K. MEENA[2], K. R. SUBRAMANIAM[1], R. VENUGOPALAN[3] and H. BHEEMANNA[4]

[1]*Dept. of Computer Science, Shrimathi Indira Gandhi College, Tiruchirappalli 620 002, Tamil Nadu, India.*
[2]*Bharathidasan University, Tiruchirappalli 620 024, Tamil Nadu, India.*
[4]*Indian Institute of Horticultural Research, Hessaraghatta Lake Post, Bangalore 560 089, Karnataka, India.*
[5]*Department of Entomology, University of Agricultural Science, Raichur 584 102, Karnataka, India.*
[*]*Corresponding author E-mail:* mpratheepa@rediffmail.com

ABSTRACT: The cotton bollworm, *Helicoverpa armigera* (Hübner) is an important pest in India damaging cotton crop and resulting in economic loss. Accurate and timely prediction of the pest, considering biotic and abiotic factors is essential to reduce the crop loss. In this paper, we present a neural-network classifier for predicting the pest incidence on cotton by considering the season, crop phenology, biotic factors (spiders and *Chrysoperla zastrowi sillemi*) and abiotic factors such as maximum temperature, minimum temperature, rainfall and relative humidity. Single layer perceptron neural-network with back-propagation algorithm was utilized for the design of the presented intelligent system. Decision tree is presented from the proposed trained neural-network. The results showed that the supervised neural network system could classify or predict the pest incidence as either 'high' or 'low' based upon economic threshold level with high degree of accuracy. Extracting rules from the decision tree helps the user to understand the role of biotic and abiotic factors on *H. armigera* incidence.

KEY WORDS: Back-propagation algorithm, biotic and abiotic factors, *Helicoverpa armigera,* knowledge extraction, neural-network classifier; pest prediction

(Article chronicle: Received: 11.03.2011; Sent for revision: 08.04.2011; Accepted: 02.05.2011)

INTRODUCTION

The infestation of bollworm (*Helicoverpa armigera* (Hübner)) in cotton crop is one of the most important constraints to crop production globally (Zalucki *et al.*, 1986). Frequent outbreaks of *H. armigera* are common in India leading to various social and economical problems. A suitable pest management systen with timely forewarning could help the farmers to take up control measures against this pest. The currently available prediction models developed mainly based on simple regression equations do not predict the pest incidence accurately. There is a good scope to use the emerging information and communication technology (ICT) to accurately predict the pest build-up. There is a need for linking the pest incidence with weather parameters, crop phenology and relative abundance of natural enemies (Trivedi *et al.*, 2005). Therefore, an attempt was made for developing a model for predicting the pest incidence in cotton crop by using data mining technique – artificial neural-networks (Han and Kamber, 2001) considering factors like season, crop stage, natural enemies, *viz.* spiders (*Pardosa* sp., *Tetragnatha* sp., *Lycosa* sp.) and *Chrysoperla zastrowi sillemi* and abiotic factors like maximum temperature, minimum temperature, rainfall and relative humidity. lSingle layer perceptron neural-network with back-propagation algorithm was used for the classification of the pest incidence as 'HIGH' or 'LOW' based upon economic threshold level (ETL) for finding the occurrence of *H. armigera* in cotton crop.

[#]This is part of the Ph. D. work of the first author submitted to Shrimathi Indira Gandhi College, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

MATERIALS AND METHODS

Database

The data sets were obtained from Regional Agricultural Research Station, Raichur, Karnataka, India from the unsprayed experimental plots under All India Coordinated Cotton Improvement Project (AICCIP) on NCS-145 (non Bt-Cotton). The sample size was 25 plants/500 sq. m^2 area. Weekly observations on mean number of *H. armigera* larvae present per five plants were recorded for the period 2005 to 2009. Natural enemies – spiders (NE1) and *Chrysoperla zastrowi sillemi* (NE2) per plant were recorded during this period. Weather parameters like maximum temperature (MaxT), minimum temperature (MinT), relative humidity (RH) and rainfall (RF) were taken based on weekly mean values. Pest incidence in relation to previous week's abiotic and biotic factors was considered for analysis. Observations at different stages of the cotton crop like 1-5 weeks age of the crop, square initiation, flowering and boll formation, boll maturity and boll bursting were taken for analysis. The sample data is given in Table 1 along with attributes/features and Class/Target variable.

Table 1. Sample records / tuples from the database

Tuples/ Sample	Variables / Features / Attributes								Class/ Target
Number	Crop stage	Season	NE1	NE2	MaxT	MinT	RF	RH	PI
1	1	Monsoon	0.00	0.00	31.81	23.24	5.8	72	Low
2	1	Monsoon	0.00	0.00	31.38	22.58	62.4	74	Low
3	1	Monsoon	0	0	30.67	22.67	47.6	82	Low
4	1	Monsoon	0	0	31.5	22.27	6	73	Low
5	2	Monsoon	0.00	0.00	32.45	23.08	59.2	74	Low
6	2	Monsoon	0.00	0.00	31.15	22.9	101	79	Low
7	2	Monsoon	0.22	0.00	31.34	22.25	0	72	High
8	2	Monsoon	0.26	0.00	29.9	22.04	50.2	76	High
9	3	Postmonsoon	0.32	0.2	31.3	22.5	18.2	71	High
10	3	Postmonsoon	0.36	0.26	33.1	22.72	8.6	68	High
11	3	Postmonsoon	0.39	0.2	32.2	21.24	46.8	71	High
12	3	Postmonsoon	0.41	0.26	30.21	22.58	81.4	77	High
13	3	Postmonsoon	0.45	0.22	31.11	21.45	1.4	68	High
14	3	Postmonsoon	0.46	0.28	28.3	17.8	21.4	76	High
15	4	Postmonsoon	0.47	0.31	30.11	16.25	0	59	High
16	4	Postmonsoon	0.52	0.36	30.22	13.41	0	57	High
17	4	Postmonsoon	0.58	0.39	30.48	16.28	0	66	High
18	4	Postmonsoon	0.59	0.49	29.85	16.51	0	60	High
19	4	Postmonsoon	0.62	0.58	31.04	17.08	0	55	High
20	5	Postmonsoon	0.71	0.61	30.81	14.42	0	55	High
21	5	Postmonsoon	0.75	0.97	31.28	15.57	0	62	High
22	5	Winter	0.82	0.22	29.85	14.37	0	59	High
23	5	Winter	0.92	0.11	29.27	14.9	0	61	Low
24	5	Winter	0.11	0.09	31.45	17.54	0	63	Low
25	5	Winter	0.1	0	33.77	15.87	0	54	Low

Assigning class label / pattern

The pest incidence was classified into two classes – 1 larva/10 plants considered as HIGH class and < 1 larva/ 10 plants considered as LOW class based on ETL (Dhaliwal and Arora, 1996). The class labels were assigned into the database for the training set of data.

Operation of an ANN

An artificial neural-network (ANN), usually called "neural-network" (NN), is a mathematical model based on biological neural networks of what goes in our brain and it can be applied for predictive modeling and classification. The most popular ANNs are back-propagation (BP) networks (Rumelhart, *et al.*, 1986) and are used in many fields.

The trained neural-network operates in a feed forward manner. However, the weight adjustments enforced by the learning rules propagate exactly backward from the output layer through the so-called "hidden layers" toward the input layer.

The threshold is one of the key components of the perceptron. It determines based on the inputs whether the perceptron fires or not. Basically, the perceptron takes all the weighted input values and adds them together. When the computed sum value is above or equal to threshold value, then the perceptron fires. Otherwise, the perceptron does not. So, it fires whenever the following equation is true (where 'W' represents the weight, and there are '*n*' inputs). The firing rule for back propagation network with perceptron concept is defined with the function (1).

$$Y = f(net) = \begin{cases} 1 & W_1X_1 + W_2X_2 + \dots + W_nX_n > \theta \\ 0 & W_1X_1 + W_2X_2 + \dots + W_nX_n < \theta \end{cases} \quad (1)$$

The activation function is f(net), θ is the threshold value and Y is the output value. The functioning of the artificial neural network is given in Fig. 1.

Network architecture

A back propagation network typically comprises three types of neuron layers - an input layer, one or more hidden layers and an output layer, each including one or more neurons. The proposed network has an input layer (on the left) with 8 input neurons, one hidden layer (in the middle) and an output layer (on the right) with 2 output neurons.

Input layer

A vector of predictor variables ($X_1 \dots X_n$) was presented to the input layer. The input layer (or processing before the input layer) standardized these values so that the range of each variable is –1 to 1 or 0 to 1. The input layer distributes the values to each of the neurons in the hidden layer.

Hidden layer

Arrived at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (W_{ij}), and the resulting weighted values are added together producing a combined value *net*. The weighted sum (*net*) is fed into a transfer function, which is *sigmoid* ($Y = f(net) = (1+e^{-net})^{-1}$) outputs a value Y. The outputs from the hidden layer are distributed to the output layer.

Output layer

The 'Y' values received from the hidden layer after activating the sigmoid activation function are the final outputs of the network.

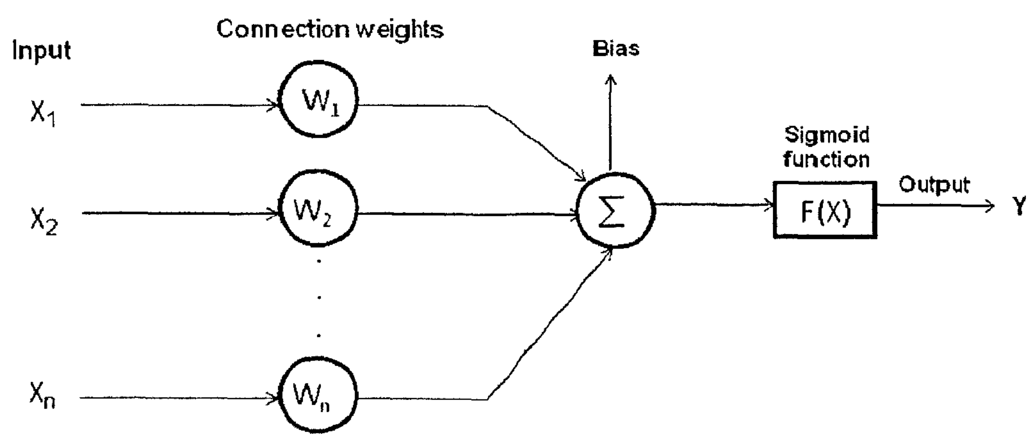

Fig 1. Operation of an artificial neural-network

Input neurons

There were eight input neurons like crop stage, season, NE1, NE2, MaxT, MinT, RF and RH, which were assigned as X_1, X_2,....,X_8. So, X_i denotes the input vector, and W_{ij} denotes the weight vector.

Data preprocessing

The normalization technique min-max method used with the function (2), so all the attribute values scaled between 0 and 1 (Obach, *et al.*, 2001). Min(x) and Max(x) denotes minimum and maximum values in the input/ attribute array X_i and 'x' denotes the input/attribute value. Min-max normalization has been chosen since it preserves the relationships among the original data values.

$$\mathrm{MinMax}(x) = \frac{x - \mathrm{Min}(x)}{\mathrm{Max}(x) - \mathrm{Min}(x)} \tag{2}$$

Weights and learning parameters

In order to optimize the performance of back-propagation network, it is essential to note that the performance is a function of several internal parameters including the transfer function, error function, learning rate and momentum term. In the proposed model weights had been initialized to small, random values between ±0.2 in the first iteration. The learning rate parameter, η = 0.25 had been assigned as constant to ensure that the network would settle/converge to a solution. Tolerance value had been set assigned as 0.0001.

Training phase and Testing phase

The leave-one-out method (Efron, 1983) was chosen for partitioning database into training set and testing set. The database for the period from monsoon, 2005 to Winter, 2008 (74 observations – 3 years data) was used as a training set. The database for the period from monsoon, 2008 to winter, 2009 (27 observations) was used as testing set to validate the model.

Hidden neurons

Starting with single hidden neuron for each class category, training of the network started. Hidden neurons are then added one at a time in an attempt to improve model performance. The total number of hidden neurons used in this network was six (0.75 * 8) and the number of hidden layer was one. The total number of hidden neurons was chosen as 3 for each class and it was confirmed with earlier report of Lenard *et al.* (1995) where 0.75* N, where N represents the number of input nodes. The proposed neural network architecture with 8 input neurons x 6 hidden neurons x 2 category/class output is shown in Fig. 2.

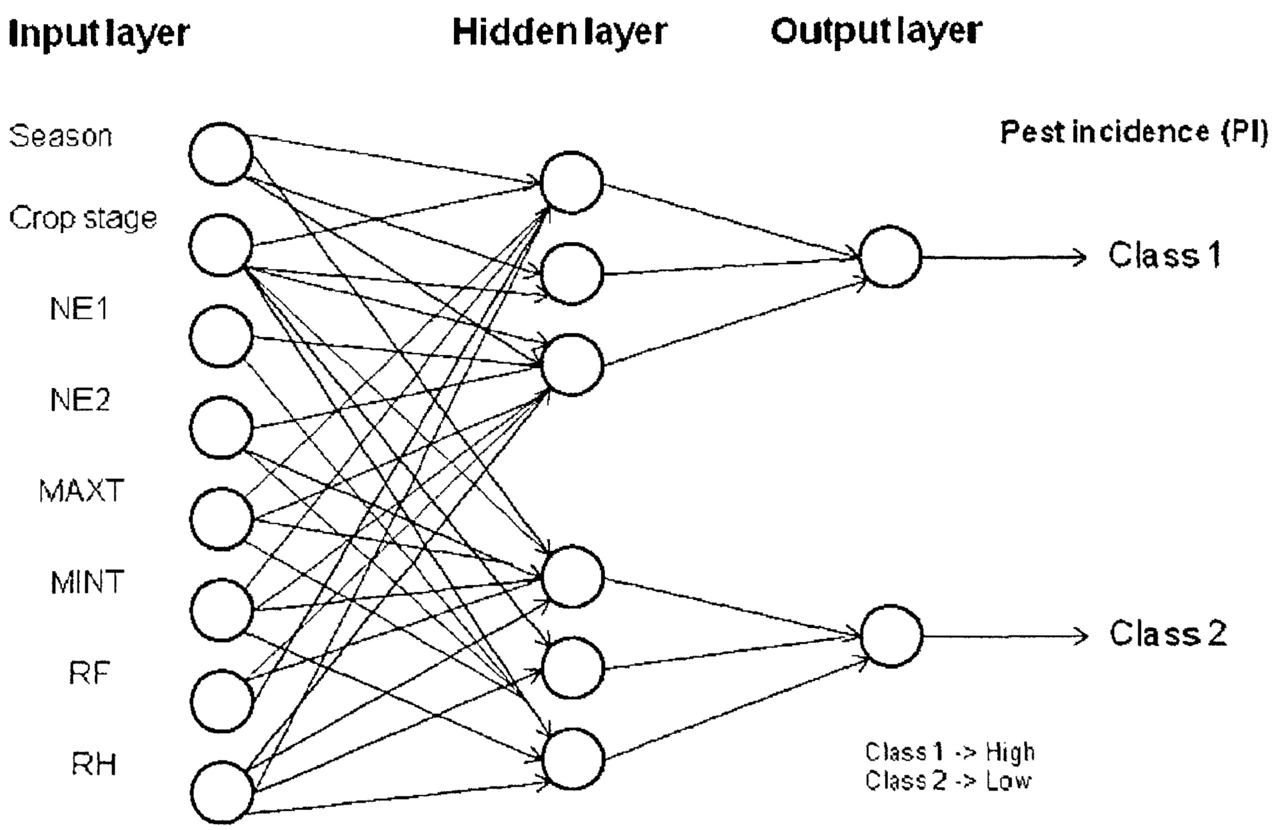

Fig. 2. BP network architecture with perceptron for prediction of *H. armigera* on cotton

Model performance

Model performance was determined using the correlation coefficient (r) between observed and estimated values of the output variable. In the training and testing procedure, as recommended by Lek *et al.* (1996), the MSE between observed and predicted values has been used to determine the optimal training zone.

The statistical model 'logistic regression' is used to validate our BPN model. Since the process of classification is dichotomous (two classes – High or Low), logistic regression analysis has been used to validate our model for both training set and testing set by using Statistical Package for Social Sciences (SPSS) Ver. 18.0.

Decision Tree

According to Andrews *et al.* (1995), the use of ANN – Multi-layer Perceptron's (MLP's) greatest weakness is their lack of transparency. Unlike decision trees, which show their seasoning explicitly, MLPs hide their knowledge in the complex interrelationships of their weights. This means that although MLPs often provide excellent models for prediction, they provide no insight into the relationships between input values and output values that the model may have found (Andrews, *et al.*, 1995). This is because there is no explanation of the mechanism inside the models. The study showed that the decision trees generated from the trained network had higher accuracy than decision trees created directly from the data (Njubai *et al.*, 2009). The optimum set of proposed single layer backpropagation network has been used to derive decision tree as a rule extraction method. The tree was constructed in the form of binary tree. The tree had decision node as condition and output/result of that condition is derived as yes or no options. The 'yes' option always grows as left child and 'no' option always grows as right child in the tree. The node ended when the condition is not able to proceed further in the 'no' option as right child. The end of the leaf node denoted the class label of pest incidence. The classes always defined in the left child of the tree, i.e., 'yes' option of the decision node. Root had been fixed as PI and the categorical variable crop stage had been taken as the first attribute/variable starting from the level-1. Maximum-minimum values used to find out the range for continuous data attributes. Mode value had been chosen for categorical attributes. The decision tree diagrams of *H. armigera* incidence at 'High' or 'Low' are given in Fig. 4 and Fig. 5, respectively.

RESULTS AND DISCUSSION

The line diagram by plotting the computed actual values (Y) and the desired values (Fig 3) revealed that the proposed neural network model predicted the pest incidence more accurately and the average of mean square (Table 4) reveals that the pest incidence was HIGH when the neural network output ('Y') was > 0 and the pest incidence was LOW when the neural network output ('Y') was d" 0.

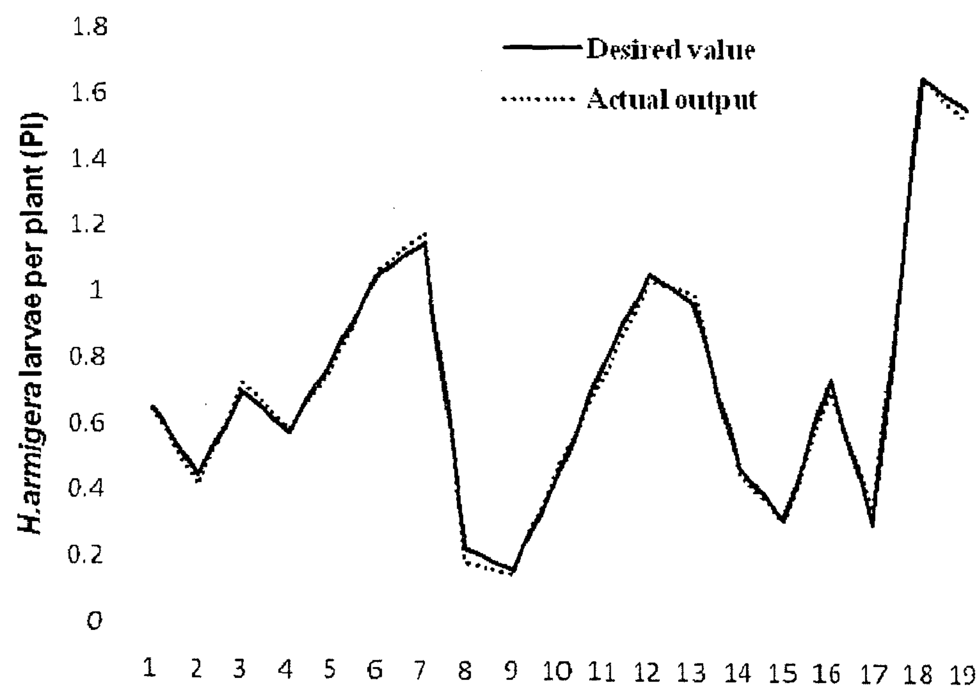

Fig. 3. Prediction performance of trained network by using back propagation algorithm

Table 4. Set of observed and computed values of pest incidence (PI) with MSE

Pest Incidence	Observed value	Computed 'Y' value	Mean square error (MSE)
High	0.7	0.73	0.000
High	0.8	0.77	0.000
High	0.65	0.64	0.000
High	1.06	1.07	0.000
Low	0.084	0.09	0.0003
Low	0.04	−0.037	0.003
Low	0	0.0295	0.0004

It has been observed that 'Y' value for class1 category was > 0 and 'Y' value for class 2 category was d" 0. Here, the trained neural network learned the concepts of classifying the pest incidence as either HIGH or LOW. If the actual output, 'Y' value derived as > 0, then there was a chance of high level of pest incidence (above ETL). If the actual output, 'Y' value derived as < 0, then there was no chance or less chance or low level occurrence of pest incidence (below ETL). The number of training records was 74 and the goodness of fit is 47.908 and R^2 value is 0.65. The no. of testing records was 27 and the goodness of fit is 7.517 and R^2 value is 0.887. The correctly classified instances (CCI) for training set is 79 percent and for testing set is 92 percent.

Njubai *et al.* (2009) developed mating decision support system by using neural network model in Kenyan Holstein-Friesian dairy cattle. The results of decision tree (Fig. 4) revealed that the pest incidence was high when the crop was in square initiation stage, flowering and boll formation stage and boll bursting stage during post monsoon season. The pest incidence was low during monsoon when the crop is 1–5 weeks old and in square initiation stage. Rules can be extracted based on decision tree and results are confirmed with earlier reports of Leonardo and Miriam (2002) and Prasad *et al.* (2008a). According to Leonardo and Miriam (2002) *H. armigera* larvae were confined to succulent plant parts like growing tips, small squares, big squares and bolls. Prasad *et al.* (2008b) observed that the peak activity of *H. armigera* adults was from September to November in Andhra Pradesh. Similar work has been carried out for predicting the occurrence of *H. armigera* and its natural enemies on cotton by using decision tree analysis with Shannon information theory (Pratheepa *et al.*, 2011).

When the maximum temperature ranged from 29.7°C to 33°C, the pest incidence would be high. When the minimum temperature ranged from 14.37°C to 22°C, the pest incidence will be high. When relative humidity ranged from 54% to 81% the pest incidence was high. These findings are in corroboration with those of earlier reports by Dahiya (1997), Dubey *et al.* (2004) and Dhaliwal *et al.* (2004). The pest incidence was high when there was no rainfall as well as when rain fall was more, which was in concurrence with the reports of Bhatti *et al.* (2007).

The prediction model with back propagation artificial neural network method has been used for forecasting the paddy stem borer (*Scirpophaga incertulas*) and the proposed method predicted the pest incidence more accurately (Lin-nan *et al.*, 2009). Neural network method has been used to understand the population dynamics of *H. armigera* on chickpea (*Cicer arietinum* L.) by considering the weather factors such as minimum temperature, maximum temperature, relative humidity and rainfall and neural network method successfully predicted the pest incidence one week in advance (Gupta *et al.*, 2003). In our proposed prediction model, back propagation neural network with perceptron concept has been used for predicting the *H. armigera* incidence on cotton by considering the major factors like crop stage, season, natural enemies (spiders and *C. zastrowi sillemi*), weather factors like minimum temperature, maximum temperature, relative humidity and rainfall (Trivedi *et al.*, 2005).

When the output of the ANN model consists of the species abundance, richness, diversity, density or a derived index, commonly used performance measures are the correlation (r) or determination (R^2) coefficient and the mean square error (MSE). The correlation coefficient (r) between observed and predicted values is 0.997 and it is significant ($P<0.01$). The no. of training records were 74 and the goodness of fit is 47.908 and R^2 value is 0.650. The no. of testing records was 27 and the goodness of fit is 7.517 and R^2 value is 0.887. The correctly classified instances for testing set were 92.59%. The average mean square error value was 0.000378. Hence, it has been found that the single layer perceptron neural network model with back propagation algorithm could be used for finding the occurrence of the *H. armigera* incidence as either HIGH or LOW.

The proposed single layer perceptron neural network with 8 x 6 x 2 architecture and back propagation algorithm can be the best method for finding the occurrence of the pest incidence. The trained neural network learned the

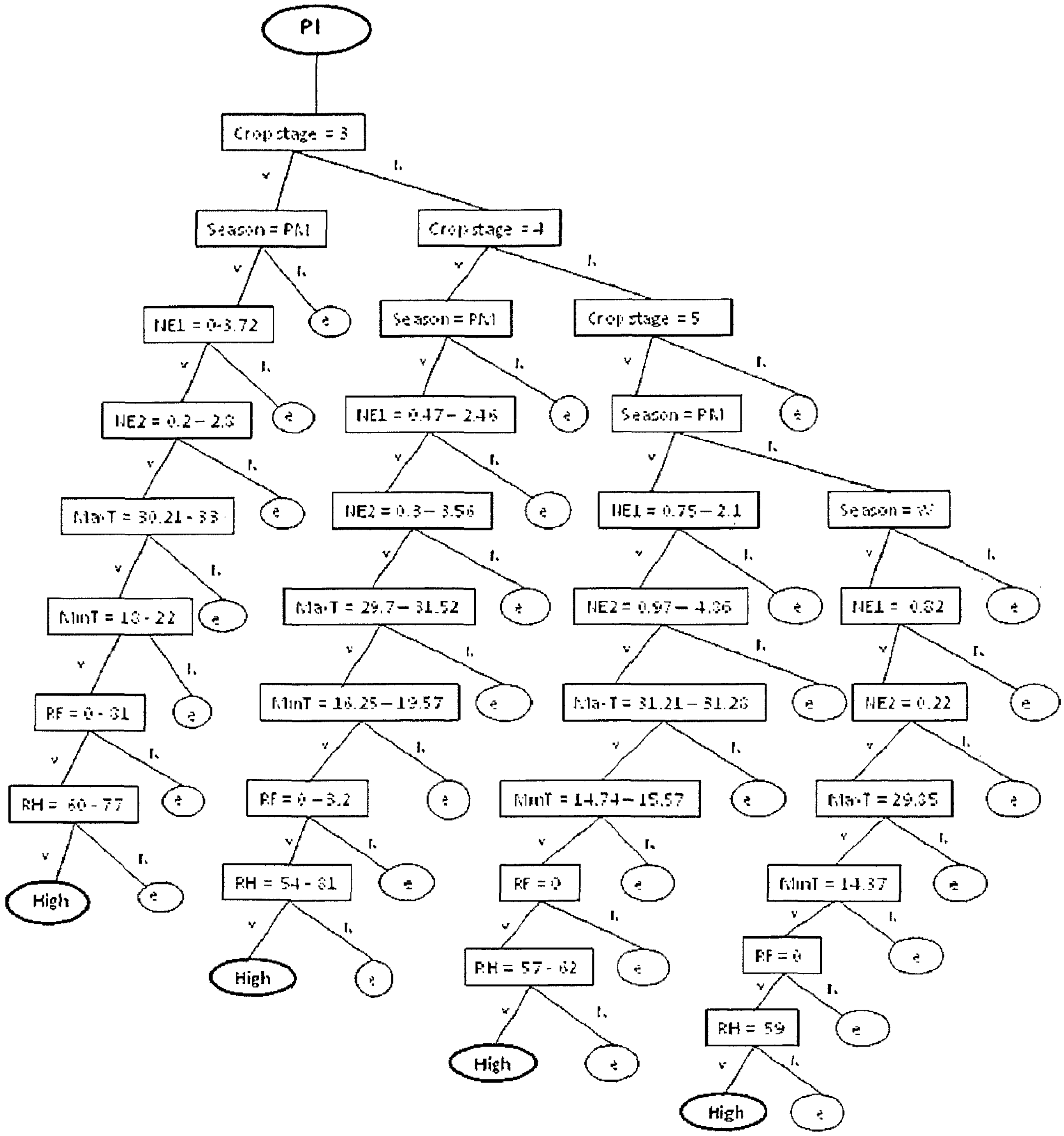

Fig. 4. Decision tree when *H. armigera* incidence was 'HIGH'

concepts of classifying the pest incidence as either HIGH or LOW. If the actual output, 'Y' value is derived as > 0, then there was a chance of high level occurrence of pest incidence (above ETL). If the actual output, 'Y' value is derived as < 0, then there was no occurrence or less occurrence or low level occurrence of pest incidence (below ETL). Therefore, it could be concluded that the developed prediction model computed on ANN considering the major variables like crop stage, season, natural enemies, (spiders and *C. zastrowi sillemi*), weather factors like minimum temperature, maximum temperature, relative humidity and rainfall, which affect the *H. armigera* incidence in cotton crop could be used accurately for predicting *H. armigera* incidence in cotton crop and thus it would be helpful to take up the control measures in advance to reduce the crop loss. The decision tree

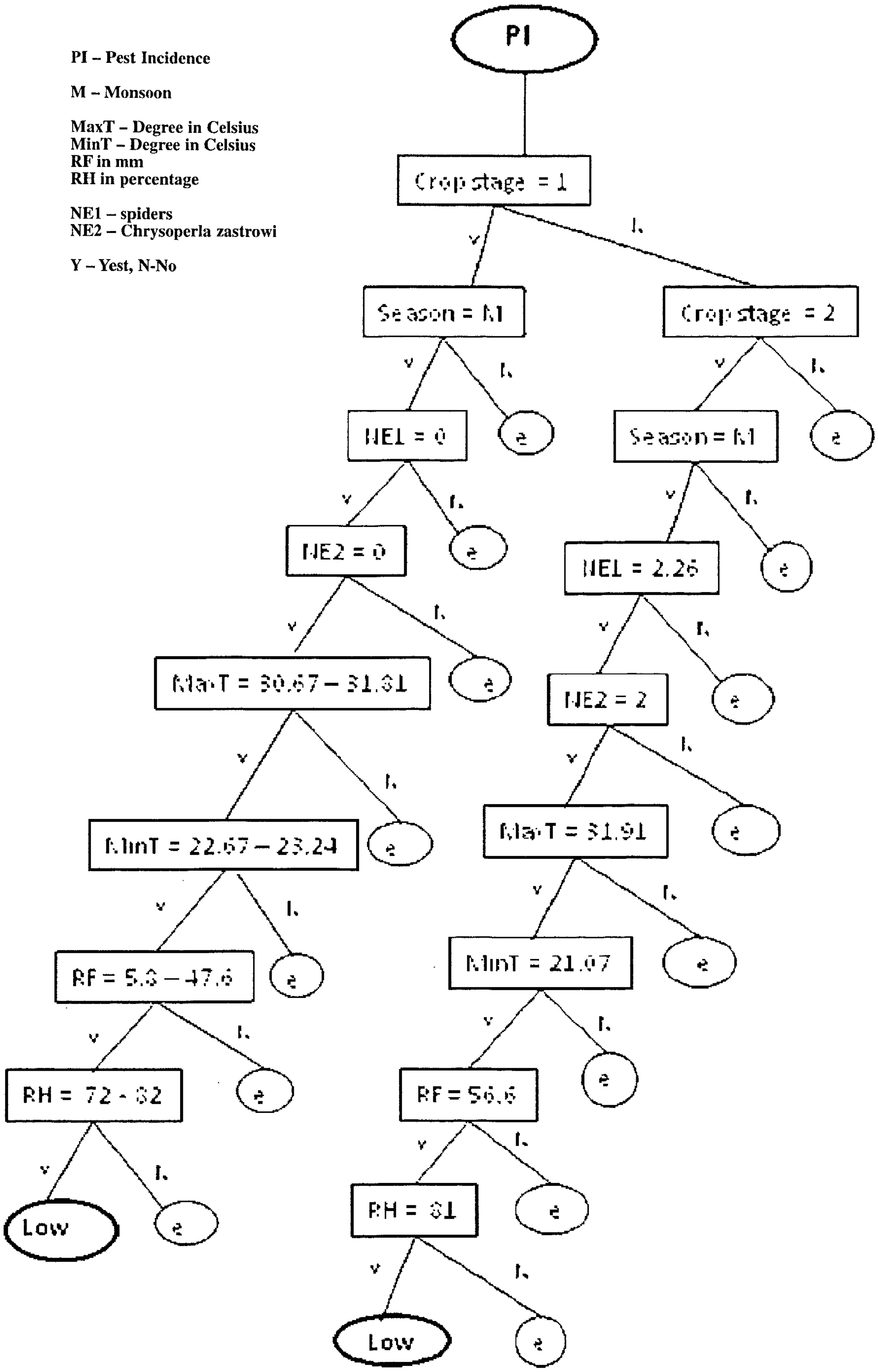

Fig. 5. Decision tree when *H. armigera* incidence was 'LOW'

derived from the optimal set of ANN helps to understand the role of biotic and abiotic factors in the occurrence of *H. armigera* on cotton crop.

ACKNOWLEDGEMENT

The authors thank an anonymous referee for valuable comments that helped to improve the paper.

REFERENCES

Andrews, R., Diederich, J. and Tickle, A. B. 1995. Survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge-Based Systems*, **8**: 373–389.

Bhatti, J. A., Khan, M. A., Murtaza, M. A., Majeed, M. Z. and Jamil, F. F. 2007. Response of American bollworm (*Helicoverpa armigera* Hüb.) to weather factors in cotton under unsprayed conditions. *Journal of Agricultural Research*, **45**: 209–214.

Dahiya, K. K., Chauhan, R., Ombir and Malik, V. S. 1997. Effect of abiotic factors on the incidence of pod borer, *Helicoverpa armigera* (Hübner) on early maturing varieties of red gram. *Crop Research Hisar*, **14**: 151–154.

Dhaliwal, G. S. and Arora R. (Eds.). 1996. Integrated pest management: Achievements and Challenges, pp. 308-355. In: Principles of Insect Pest Management, NATIC, India.

Dhaliwal, L. K., Kooner, B. S., Singh, J. and Sohi, A. S. 2004. Incidence of *Helicoverpa armigera* (Hübner) in relation to meteorological parameters under Punjab conditions. *Journal of Agrometeorology*, **6** (Special Issue): 115–119.

Dubey, A., Kanaujia, K. R. and Kanaujia, S. 2004. Efficiency of different pheromonal blends to monitor *Helicoverpa armigera* (Hübner) moth catches. *Indian Journal of Plant Protection*, **32**:147–148.

Efron, B. 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of American Statistical Association*, **78**: 316–330.

Gupta, R., Narayana, B. V. L., Krishna Reddy, P., Ranga Rao, G. V., Gowda, C. L. L., Reddy, Y. V. R. and Rama Murthy, G. 2003. Understanding *Helicoverpa armigera* pest population dynamics related to chickpea crop using neural-networks, pp. 723–726. In: Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03).

Han J. and Kamber, M. 2001. Classification and Prediction, pp. 285–375. In: Gray, J. (Ed.). *Data Mining Concepts and Techniques*, Morgan Kaufmann.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga J. and Aulagnier, S. 1996. Application of neural networks to modeling non linear relationships in ecology. *Ecological Modelling*, **90**: 39–52.

Lenard, M. J., Alam, P. and Madey, G. R. 1995. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences*, **26**: 209–227.

Leonardo, T. and Miriam E. Pascua. 2002. The distribution and movement of cotton bollworm, *Helicoverpa armigera* Hubner (Lepidoptera:Noctuidae) larvae on cotton. *Philippine Journal of Science*, **131**: 91–98.

Lin-nan, Y., Lin P., Li-min Z., Li-lian Zhang, Shi-sheng Y. 2009. A prediction model for population occurrence of paddy stem borer (*Scirpophaga incertulas*), based on back propagation artificial neural network and principal components analysis. *Computers and Electronics in Agriculture*, **68**: 200–206.

Njubai, D. M., Wakhungu, J. and Badamana, M. S. 2009. Mating decision support system using computer neural network model in Kenyan Holstein-Friesian dairy cattle. *Livestock Research for Rural Development. Volume 21, Article # 45.* Retrieved May 31, 2010, from http://www.lrrd.org/lrrd21/4/njub21045.htm.

Obach, M., Wagner, R., Werner, H., Schmidt, H. H., 2001. Modelling population dynamics of aquatic insects with artificial neural networks. *Ecological Modeling*, **146**: 207–217.

Prasad,T.V., Nandagopal,V. and Gedia, M. V. 2008a. Effect of abiotic factors on the population dynamics of *Aphis craccivora* Koch in groundnut in Saurashtra Region of Gujarat. *Indian Journal of Entomology*, **70**: 309–313.

Prasad, N. V. V. S. D, Mahalakshmi, M. S. and Rao, N. H. P. 2008b. Monitoring of cotton bollworms through pheromone traps and impact of abiotic factors on trap catch. *Journal of Entomological Research*, **32**: 187–192.

Pratheepa, M., Meena, K., Subramaniam, K. R., Venugopalan, R. and Bheemanna, H. 2011. A decision tree analysis for predicting the occurrence of the pest, *Helicoverpa armigera* and its natural enemies on cotton based on economic threshold level. *Current Science*, **100**: 238–246.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. 1986. Learning representations by back-propagation errors. *Nature*, **323**: 533–536.

Trivedi, T. P., Yadav, C. P., Vishwadhar, Srivastava, C. P., Dhandapani, A., Das, D. K., Singh, J. 2005. Monitoring and forecasting of *Heliothis/Helicoverpa* population, pp. 119–140. In: Sharma, H. C. (Ed.). *Heliothis / Helicoverpa* Management – Emerging Trends and Strategies for Future Research. Oxford & IBH, New Delhi, India.

Zalucki, M. P., Daglish, G., Firempong, S. and Twine, P. 1986. The biology and ecology of *Heliothis armigera* (Hübner) and *H. punctigera* Wallengren (Lepidoptera: Noctuidae) in Australia: What do we know? *Australian Journal of Zoology*, **34**: 779–814.