

Full Length Research Paper

Time series prediction of apple scab using meteorological measurements

Bayram Cetişli^{1*} and Eşref Büyükçingir²

¹Computer Engineering Department, Süleyman Demirel University, Isparta, 32260, Turkey.

²ARTEM Electric and Automation, Yeni Sanayi Sitesi, 36. Blok, No: 10, Isparta, 32200, Turkey.

Accepted 23 August, 2013

A new prediction model for the early warning of apple scab is proposed in this study. The method is based on artificial intelligence and time series prediction. The infection period of apple scab was evaluated as the time series prediction model instead of summation of wetness duration. Also, the relations of different measurements with apple scab infection time were analyzed. The important hours of duration were determined with the feature selection methods, such as Pearson's correlation coefficients (PCC), Fisher's linear discriminant analysis (FLDA) and an adaptive neuro-fuzzy classifier with linguistic hedges (ANFC_LH). The experimental dataset with selected features was classified by ANFC_LH, and predicted by an adaptive neural network (ANN) model. The proposed ANN model successfully predicts the apple scab infection time with 2 to 5% error rates compared to the traditional weather station predictions. The results show that the last 24-hour period is important to determine the apple scab infection at any time.

Key words: Apple scab (*Venturia inaequalis*), early warning, time series prediction, feature selection, artificial intelligence.

INTRODUCTION

The plant protection activities against apple pest are very important for growers. Meteorological conditions directly affect the crop quality of productions. For that reason, the meteorological conditions are monitored to protect the orchards against pests and diseases. The aims of the early warning and disease modeling systems are to decrease the use of pesticides, to eliminate crop losses caused by diseases and insects, and to protect the natural environment by analyzing the meteorological conditions. This study focuses on predicting the apple scab infection times.

The meteorological conditions with time factor affect the growth cycle of apple scab. This phenomenon was first

detected by Keitt and Jones (1926) who denoted the relation between temperature and wetting duration for apple scab infection. After that, Mills (1994) prepared a table about the meteorological effects on apple scab infections. This table shows the summation of wetness duration with different temperatures. Then, this table was revised by MacHardy and Gadoury (1989). According to researches, the temperature, leaf wetness and humidity affect the appearance time of the infection and intensity of the apple scab. Different apple scab models based on meteorological measurements have been developed, evaluated and successfully used for disease control in apple orchards (Holb, 2003; Jones and Sundin, 2006).

Corresponding author. E-mail: bayramcetisli@sdu.edu.tr; Tel: (+90) 246 2111563; Fax: (+90) 246 2370859.

Abbreviations: PCC, Pearson's correlation criterion; FLDA, Fisher's linear discriminant analysis; LWS, Luftt weather station; ANFC_LH, adaptive neuro-fuzzy classifier with linguistic hedge; ASD, apple scab degree; MLPN, multi-layer perceptron network; RMSE, root mean square error.

The meteorological parameters can be monitored by traditional weather stations. By this way, the appearance time of diseases can be easily determined. Some computer-based early warning systems, such as WELTE, METOS, HP-100, ADEM and RIMpro were designed to determine the infection periods (Hindorf et al., 2000; Holb, 2003; Berrie and Xu, 2003). These systems have been used in the orchards, successfully.

In order to determine the infection periods in nature, some mathematical prediction models were proposed, besides the summation of wetness duration. Some of them are polynomial models (Mills, 1944; Gadoury and MacHardy, 1989). Some statistical methods, such as Bayes theory, were approved for disease prediction (Yuen and Hughes, 2002; Turechek and Wilcox, 2005; De Wolf and Isard, 2007). Rossi et al. (2007) estimated the risk of apple scab infection by a dynamic simulation model. They suggested an algorithm for different situations.

The infection times of diseases can also be considered as the time series problem as in weather or stock exchange forecasting. The time series analysis is a well known method, and it is generally used to forecast the next value of events or identify the nature of the phenomenon represented by a sequence of observations using a series of past measurements. Some of these studies are based on statistical methods, such as auto-regressive, auto-regressive moving average, auto-regressive integrated moving average (De Gooijer and Hyndman, 2006).

The last approaches to estimate the infection times are based on artificial intelligence, such as neural networks, fuzzy systems; support vector machines (Rojas et al., 2008). However, these methods are rarely used in agricultural problems. One of them is about an early warning system for a Ross River virus in Australia (Hu et al., 2006). Kolhe et al. (2011) suggested diagnosis of diseases of oilseed-crops based on fuzzy approach. Bregaglio et al. (2011) modeled the leaf wetness of plant disease using artificial intelligence methods. Although most of the apple scab disease models have been used in the summation of wetness duration with different temperatures, the time series analysis of apple scab with respect to meteorological variables is firstly proposed in this study.

In this study, the meteorological data, obtained by Lufft HP-100 model weather station, (Lufft, 2010) was used. The best measurement variables that are correlated with apple scab infection were determined by Pearson's correlation criterion (PCC). The most important hours of selected measurements were also determined by the feature selection algorithms such as the adaptive neuro-fuzzy classifier with linguistic hedge (ANFC_LH) and Fisher's linear discriminant analysis (FLDA) to determine the infection period. The apple scab was predicted with selected sequences upon adaptive neural networks. Although the mathematical model of Lufft is not known, the proposed early warning model easily follows the apple scab degree of Lufft. Also, the apple scab problem was eva-

evaluated as classification problem to give simple early warning, in this study.

MATERIALS AND METHODS

The Lufft model gives apple scab infection degrees between 0 and 100% according to meteorological measurements. The previously collected data, monitored from the Lufft weather station (LWS) in 2004, was used to predict the infection times of apple scab. The weather station was set up at Afşar Village of Gelendost in Isparta. In this study, it is assumed that the weather station always gives right apple scab warning because the facility to do experimental studies in orchards or in any laboratory is not possible.

The LWS gives five meteorological measurements, such as relative humidity, leaf wetness, temperature, day light and rainfall. The LWS is sending the meteorological measurements to the Center of Plant Protection Department in Agricultural Provincial Directorate of Isparta for each 12 min. The dataset was analyzed with the measurements of previous 12 and 24 h time intervals from the observed apple scab times. Consequently, the original data was arranged as: Instead of all 12-min measurements, the average values of measurements in one-hour time interval was approved; only possible infection seasons were evaluated instead of all year measurements. Therefore, the evaluated dataset consisted of measurements from 1 February 2004 to 31 October 2004; times that are before and after the occurrence of apple scab infection were also considered to correctly determine the infection timing of apple scab.

As a result, the large dataset was shortened for better prediction and classification. There are 5415 measurement points with five measurement features. Also, another important parameter is the time that was added to dataset for the time series prediction. The proposed method consists of three stages: Marking the correlations between the apple scab and meteorological measurements; detecting the appropriate times for infections; prediction and classification. These stages are depicted in Figure 1.

Feature selection

The two-stage feature selection procedure was applied to the original dataset. In the first stage of the proposed model, the correlations among the determined apple scab values and measurement features were evaluated with the PCC. As a result, the important measurement features were determined. Then, the feature selection methods, such as FLDA and ANFC_LH determined the important time (hour) of selected measurements in the second stage of the model.

Pearson's correlation criterion (PCC)

The PCC is very simple and effective method to determine the relevant features without training data (Theodoridis and Kotroumbas, 2008). PCC is generally utilized for ranking of features according to their relation capability with the output.

Adaptive neuro-fuzzy classifier with linguistic hedges (ANFC_LH)

In this paper, ANFC_LH was approved for feature selection and classification (Cetişli, 2010a). The ANFC_LH method includes the linguistic hedge concept and the adaptive neuro-fuzzy classifier. The adapted linguistic hedge values of fuzzy classes denote the importance degree of features (Cetişli, 2010b).

Let A be a continuous linguistic term for any input variable. Then A^p is interpreted as a modified version of the original linguistic

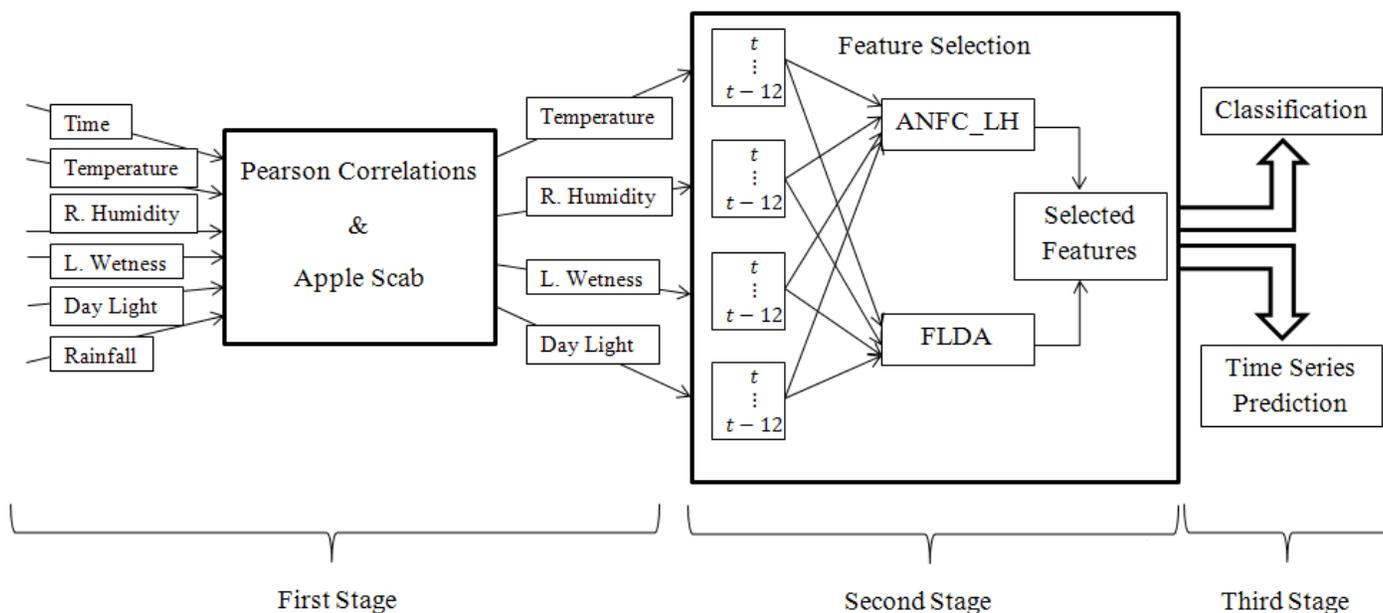


Figure 1. The stages of apple scab early warning model.

term, where p denotes the linguistic hedge value of the linguistic term A . It has a power effect to membership function. When the LH values are forced to binary values, such as $p = 0$ or $p = 1$, the modified A term be either itself or X space. The obtained linguistic hedge properties in fuzzy classification rules can be used for feature selection (Cetişli, 2010b).

Fisher's linear discriminant analysis (FLDA)

The other feature selection algorithm is the Fisher's linear discriminant analysis (FLDA). In this algorithm, the distances between classes and within classes for each feature are determined. The ratio of these distances is evaluated as feature selection criterion (Theodoridis and Kotroumbas, 2008).

Classification

In the third stage of proposed model, the prepared time series dataset can also be evaluated as any classification problem to select the important features easily and to give a simple warning. For this aim, the apple scab degree (ASD) ranges were divided into four classes: Non-scab, low, moderate and high scab as in the Mills' table intuitively. These classes are created as

$$class = \begin{cases} non, & \text{if } ASD = 0\% \\ low, & \text{if } 0\% < ASD \leq 33.3\% \\ moderate, & \text{if } 33.3\% < ASD \leq 66.6\% \\ high, & \text{if } 66.6\% < ASD \end{cases} \quad 1$$

The selected features and their selected hours can be used for classification and prediction of apple scab. The ANFC_LH was approved for the scab classification.

Prediction

The ANNs is based on human neural system and is very suitable to predict the ASD of Luftt model.

Adaptive neural networks

ANNs are computational models that try to simulate the structure and functional aspects of biological neural networks (Jang et al., 1997). ANNs consist of an interconnected group of artificial neurons and process information using a connectionist approach for computation (Theodoridis and Kotroumbas, 2008). Modern neural networks are evaluated as the non-linear statistical data modeling tools. They are generally used to determine complex relationships between inputs and outputs (Jang et al., 1997). ANN can be used for classification and prediction.

The multi-layer perceptron network (MLPN), which is special and popular model of ANNs, is created by using the simple neurons as shown in Figure 2. In the MLPN, various numbers of neurons can be assigned for different layers. One can define the adequate number of neurons after a few trials. In each layer, the neurons show the similar properties and the inputs of hidden layers are the outputs of previous layer. Also, the neuron connections that are called as weights can be adapted by different training methods, such as gradient based algorithms, heuristic or evolutionary methods (Jang et al., 1997; Theodoridis and Kotroumbas, 2008).

RESULTS AND DISCUSSION

The important meteorological measurements of the used data should be selected for robust prediction. Table 1 shows the PCCs of the features with respect to the determined ASDs. It can be seen in Table 1 that the best relation is between the leaf wetness and the ASDs. The others are ranked as relative humidity, day light, temperature,

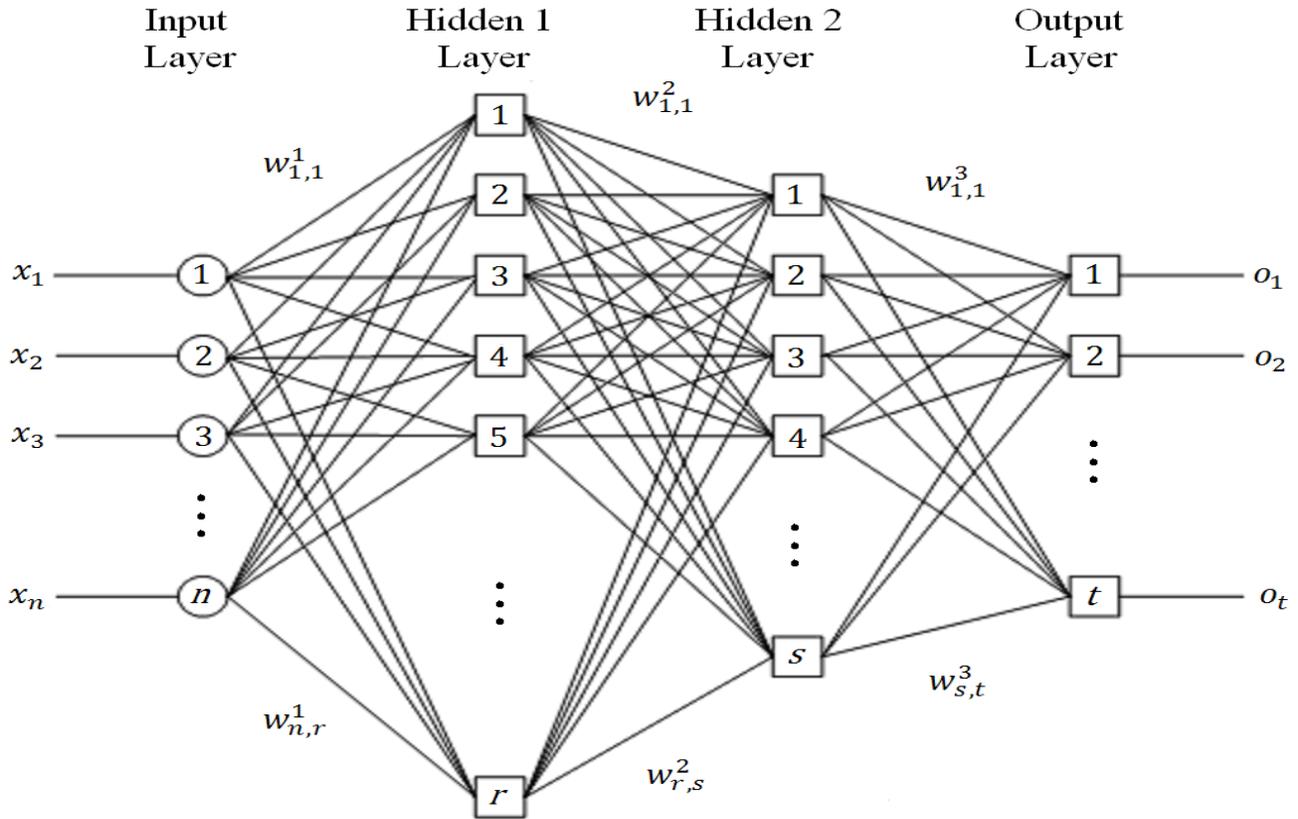


Figure 2. A multilayer perceptron network model.

Table 1. The Pearson's correlation coefficients of measurements respect to apple scab degrees.

Parameter	Time (h)	Temperature	Relative humidity	Leaf wetness	Day light	Rainfall	Apple scab
Time (h)	1	0.24	-0.33	0.18	0.01	-0.01	-0.01
Temperature	0.24	1	-0.71	0.27	0.44	-0.02	-0.04
Relative humidity	-0.33	-0.71	1	-0.56	-0.47	0.03	0.18
Leaf wetness	0.18	0.27	-0.56	1	0.09	-0.05	-0.42
Day light	0.01	0.44	-0.47	0.09	1	-0.01	0.06
Rainfall	-0.01	-0.02	0.03	-0.05	-0.01	1	0.02
Apple scab	-0.01	-0.04	0.18	-0.42	0.06	0.02	1

rainfall, and time. Moreover, there are very important relations among the features. For instance, there are inverse correlations between the temperature and relative humidity; between the leaf wetness and relative humidity. Despite the widespread knowledge about the relation between the leaf wetness and rainfall, the leaf wetness is not related with the rainfall. These results can also be seen in Figure 3. In this case, the leaf wetness, relative humidity, daylight and air temperature measurements can be selected to prepare the time series dataset.

The important times in hours of selected measurements could be determined by ANFC_LH and FLDA methods. In both previous 12 and 24 h measurement time intervals, the most important times were selected. Each obser-

vation of prepared dataset is composed of 72 or 144 features which were obtained by six different measurements at the last 12 and 24 hours, respectively. The dataset was equally divided into training and testing sets. Figures 4 and 5 show the mean values of classes with respect to 12 and 24 h measurements. When Figure 3 is analyzed, the duration between $t-1$ and $t-6$ in temperature is distinctive for the high and moderate scab classes. While the non-scab and high scab classes are easily separated, the low and moderate scab classes are generally mixed.

In Figure 4, it can be seen that the scab classes denote periodicity with respect to time duration except for the high level scabs. The time values of high class keep the

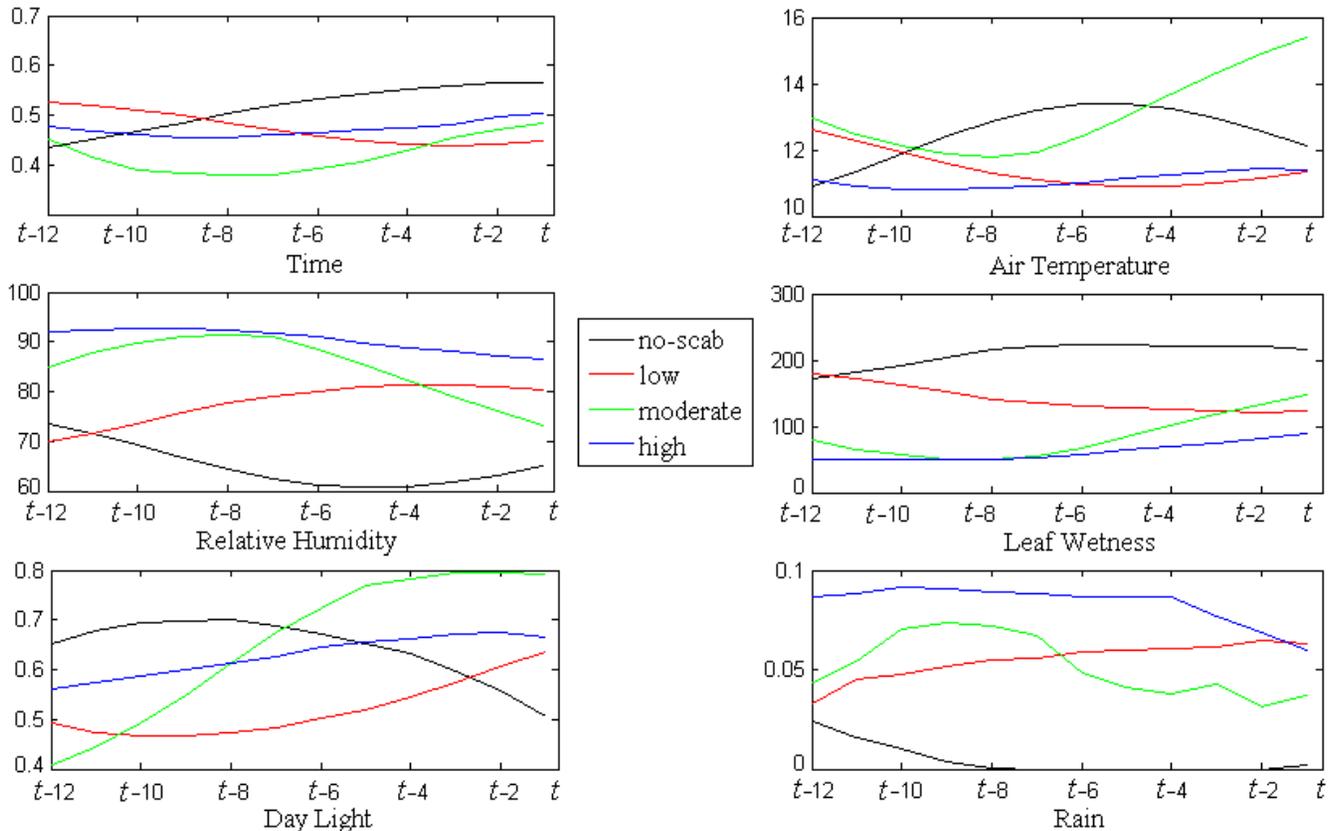


Figure 3. The mean values of 12-hour measurements for each class.

Table 2. The important times of measurements for 12-hour measurements.

Feature	Selected times (h)	Number of features
Temperature	$t-9, t-8, t-7, t-1$	4
Relative humidity	$t-11, t-8, t-6$	3
Leaf wetness	All	12
Day light	$t-1$	1

values through the duration. The leaf wetness, temperature, and relative humidity maintain their discriminative properties. In the feature selection, the important times of selected measurements were sorted by linguistic hedge values and FLDA scores. The important times of selected measurements are given in Tables 2 and 3. There are 20 and 23 distinctive hours of measurements for previous 12 and 24 h time intervals, respectively.

It is observed in Tables 2 and 3 that the leaf wetness values in all hours are important for classification. It means that the Luft model uses the summation of wetness duration. When the other measurements are analyzed, there are no remarkable results with regard to important hours. The selected features were also classified by ANFC_LH, MLPN, and Bayes for comparison. The Bayes classifier is a simple and popular probabilistic method

based on Bayes' theorem (Theodoridis and Kotroumbas, 2008). The obtained classification results are given in Table 4. ANFC_LH is sufficient to give right warning alarm. If Table 4 is analyzed, there is no significant difference between the 12- and 24- time durations according to classification results. As a result, the 12 h time duration is sufficient to discriminate the apple scab classes.

If the apple scab degree as percentage is very important instead of scab classes, the prediction methods are necessary. The obtained prediction results by using MLPN, ANFIS and auto-regressive moving average (ARMA) methods are given in Table 5 for 12- and 24-hour measurements. The ANFIS is a popular approximation method based on fuzzy rules (Jang et al., 1997). The ARMA method is an iterative statistics based method (Rojas et al., 2008). In this study, the ARMA(q,t), where q is the

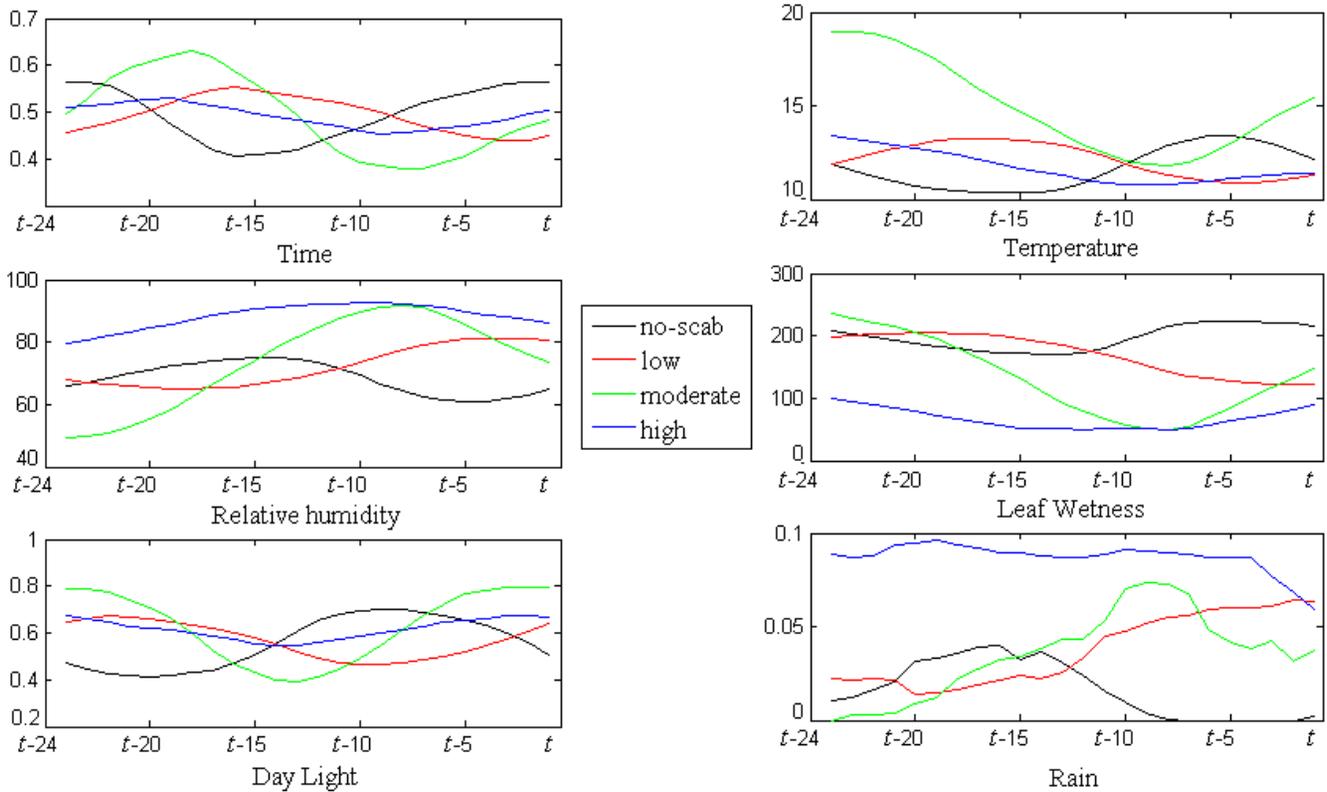


Figure 4. The mean values of 24-hour measurements for each class.

Table 3. The important times of measurements for 24-hour measurements.

Feature	Selected times (h)	Number of feature
Temperature	$t-23, t-12$	2
Relative humidity	$t-1$	1
Leaf wetness	$t-1, t-3, t-15, t-21, t-24$	15
Day light	$t-1, t-9, t-16, t-18, t-24$	5

Table 4. The classification results of early warning dataset.

Method	12 h duration classification rate (%)	24 h duration classification rate (%)	Explanations
ANFC_LH	95.86	95.97	20 fuzzy rules, 400 epochs
MLPN	89.54	92.42	20-40-20-4 network, 400 epochs
Bayes	80.31	81.57	-

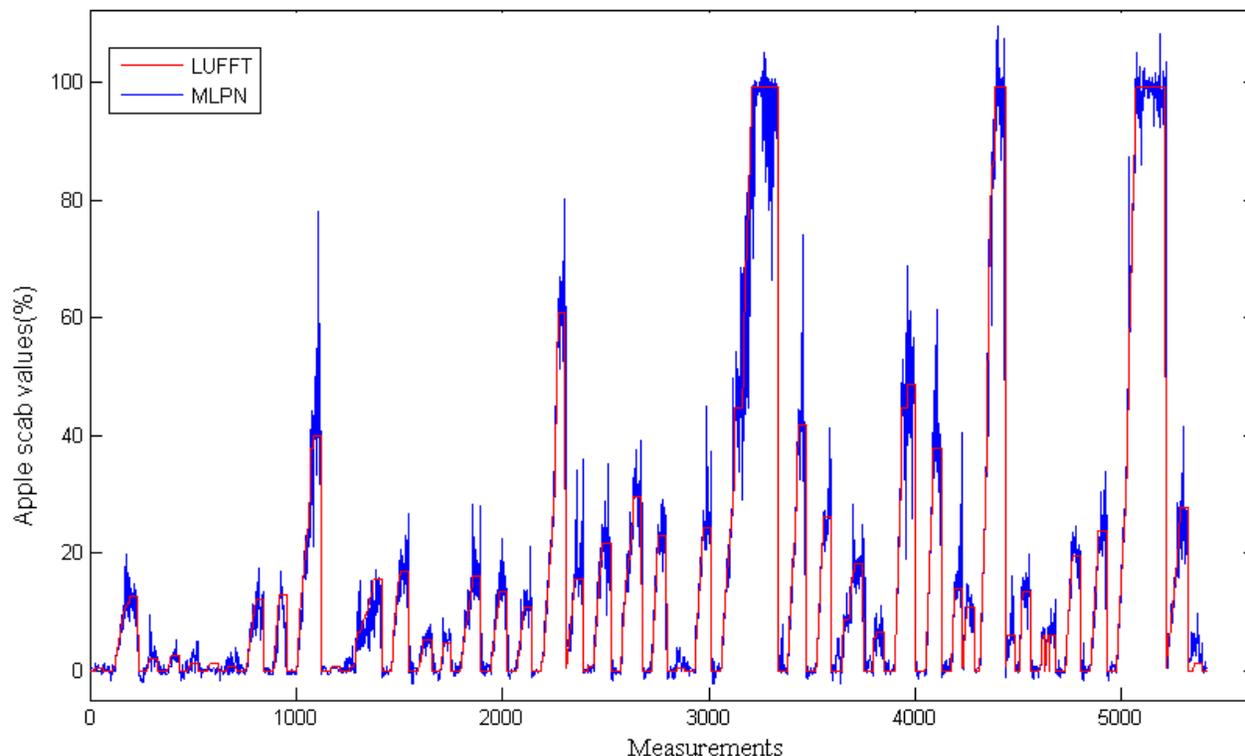
number of autoregressive parameters, t is the number of moving average parameters, was adapted using the Levenberg-Marquardt method.

In the table, root mean square error (RMSE) criterion was evaluated for comparison (Theodoridis and Kotroumbas, 2008). Among the used methods, the ARMA produced the worst results. The ANFIS estimation results could not be lower than 10% (RMSE) rates (Büyükçingir, 2009).

For this reason, the MLPN was preferred as the main predictor in this study. The MLPN has four layers: input, two hidden and output layers. There is a difference between the time durations according to RMSE values; 24-hour measurements give better results than 12-hour measurements with same features. The apple scab predictions of 12 and 24 h measurements by using MLPNs are depicted in Figures 5 and 6, respectively.

Table 5. Prediction results of the apple scab for 12- and 24-hour measurements.

Parameter	12 h times RMSE	24 h times RMSE	Explanation
MLPN	4.94	2.64	20-200-100-1 network, 1000 epochs
ANFIS	10.95	7.36	20 fuzzy rules, 1000 epochs
ARMA	36.51	27.48	ARMA (0, 12) and ARMA (0, 24)

**Figure 5.** Apple scab predictions with 12-hour time durations.

In most cases for 12 h durations, the MLPN predictions can approximate to the Lufft prediction. However, in some cases, there are big differences between the predictions and real values, which may cause false warning. The false warning rate can be decreased by 24 h time duration. In this time interval, the prediction errors are lower than 5%.

Conclusions

In this study, the predicting of apple scab infection time is performed by using meteorological measurements. Although most researchers have used the summation of wetness or other measurement duration as in the Mills' table to determine the apple scab, the meteorological measurements can also be evaluated as the time series prediction problems. For this purpose, a new early warning model of apple scab is designed by using artificial intelligence based methods, in this study. The model consists of three stages: Correlation, feature selection

and prediction-classification. In the first stage, the correlated measurements with the apple scab degrees are determined. Then the important hours of correlated measurements are selected. Lastly, the degree of apple scab is predicted, or is classified. While the 12 h time duration is sufficient to determine the apple scab classes, the 24 h time duration is necessary for the prediction of apple scab attack degree.

It is shown that the leaf wetness is the most important measurement for determination of apple scab degree. In addition, the relative humidity, daylight and temperature measurements were also determined as important features. As per literature, the time and rainfall measurements are important features to determine the scab; however, there is no important relation between the mentioned features and scab attack degrees. Also, the experimental studies show that the Lufft early warning system monitors the summation of leaf wetness durations.

In future, this method can be applied to design regional and different apple varieties of Isparta in Turkey. After

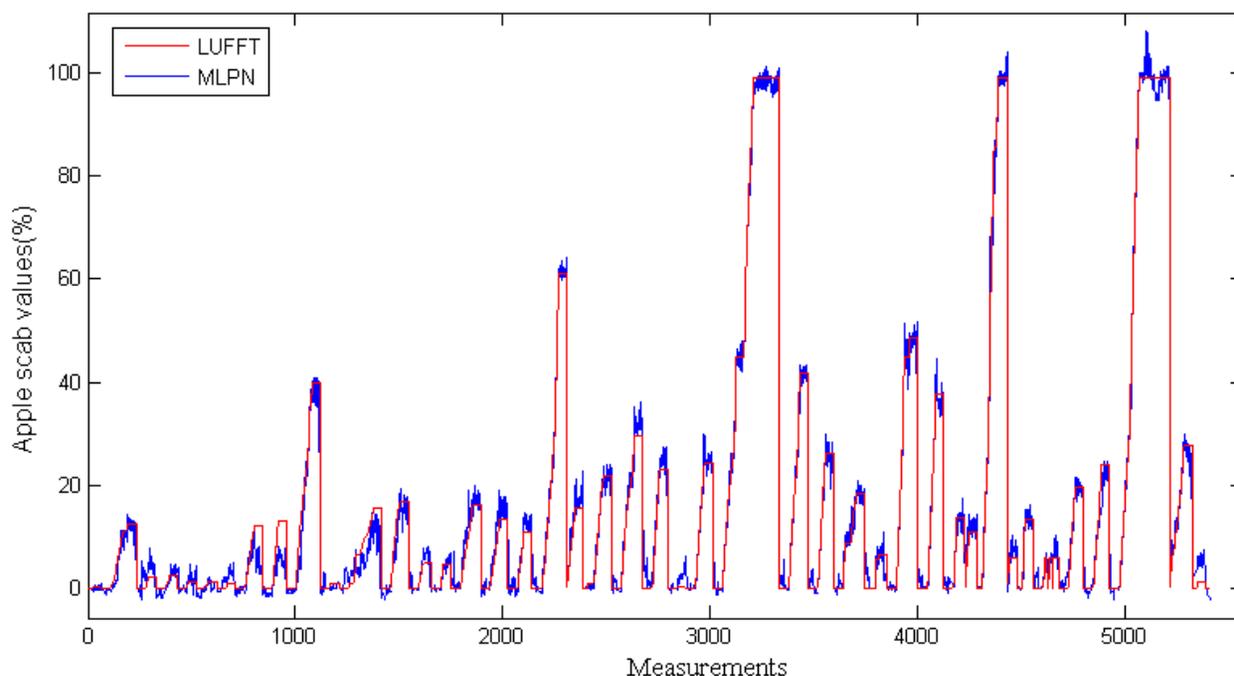


Figure 6. Apple scab predictions with 24-hour time durations.

designing the scab models, an early warning system will be designed using electronic embedded systems for apple orchards.

ACKNOWLEDGEMENT

We thank Isparta Provincial Directorate of Agriculture for experimental dataset.

REFERENCES

- Berrie AM, Xu XM (2003). Managing apple scab (*Venturia inaequalis*) and powdery mildew (*Podosphaera leucotricha*) using Adem™. *Int. J. Pest Managt.* 49(3):243-249.
- Bregaglio S, Donatelli M, Confalonieri R, Acutis M, Orlandini S (2011). Multi metric evaluation of leaf wetness models for large-area application of plant disease models. *Agr. Forest Meteorol.* 151(9):1163-1172
- Büyükçingir E (2009). Modelling of apple scab and apple codling moth using adaptive network-based fuzzy inference system (ANFIS) and designing with programmable logic controller. M.Sc. Thesis, Süleyman Demirel Un. Grad. Sch. Appl. Nat. Sc. Isparta, TR.
- Cetişli B (2010a). Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1. *Expert Sys. Appl.* 37(8):6093-6101.
- Cetişli B (2010b). The effect of linguistic hedges on feature selection: Part 2. *Expert Sys. with Appl.* 37(8): 6102-6108.
- De Gooijer JG, Hyndman RJ (2006). 25 years of time series forecasting. *Int. J. Forecasting* 22(3):443-473.
- De Wolf ED, Isard SA (2007). Disease cycle approach to plant disease prediction. *The Annual Review of Phytopathology* 45:203-220.
- Gadoury DM, MacHardy WE (1989). A revision of Mills's criteria for predicting apple scab infection periods. *Phytopathology* 79:304-310.
- Hindorf H, Rövekamp IF, Henseler K (2000). Decision aids for apple scab warning services (*Venturia inaequalis*) in Germany. *Bulletin OEPP/EPPO Bulletin* 30:59-64.
- Holb IJ (2003). Comparison of scab warning systems in integrated apple production. *J. Agri. Sc.* 11:53-57.
- Hu W, Tong S, Mengersen K, Oldenburg B (2006). Rainfall, mosquito density and the transmission of Ross River virus: A time-series forecasting model. *Ecol. Modelling* 196(3-4):505-514.
- Jang JSR, Sun CT, Mizutani E (1997). *Neuro-fuzzy and soft computing*. Prentice Hall, Upper Saddle River.
- Jones AL, Sundin GW (2006). Apple scab: role of environment in pathogen and epidemic development. Cooke BM, Jones DG Kaye B. (Eds.). *The Epidemiology of Plant Diseases*, 2nd Ed., ch. 18:473-489.
- Keitt GW, Jones LK (1926). Studies of the epidemiology and control of apple scab. *Wisconsin Agri. Exp. Station Res. Bulletin* 73:1-104.
- Kolhe S, Kamal R, Saini HS, Gupta GK (2011). A web-based intelligent disease-diagnosis system using a new fuzzy-logic based approach for drawing the inferences in crops. *Comp Elect. Agr.* 76(1):16-27.
- Lufft, HP-100 remote reading weather station and crop monitor, <http://www.tequipment.net/LufftHP100.html> (last accessed 10 November 2010).
- Mateo F, Sovilj D, Gadea R (2010). Approximate k-NN delta test minimization method using genetic algorithms: Application to time series. *Neurocomputing* 73(10-12):2017-2029.
- Mills WD (1944). Efficient use of sulphur dusts and sprays during rain to control apple scab. *Cornell University Extension Bulletin* 630:1-4.
- Rojas I, Valenzuela, O, Rojas F, Guillen A, Herrera LJ, Pomares H, Marquez L, Pasadas M (2008). Soft-computing techniques and ARMA model for time series prediction. *Neurocomputing* 71(4-6):519-537.
- Rossi V, Giosuè S, Bugiani R (2007). A-scab (Apple-scab), a simulation model for estimating risk of *Venturia inaequalis* primary infections. *Bulletin OEPP/EPPO Bulletin* 37:300-308.
- Theodoridis S, Kotroumbas K, (2008). *Pattern Recognition*. Academic Press, San Diego. 4th Ed.
- Turechek WW, Wilcox WF (2005). Evaluating predictors of apple scab with receiver operating characteristic curve analysis. *Phytopathology* 95:679-691.
- Yuen JE, Hughes G (2002). Bayesian analysis of plant disease prediction. *Plant Pathol.* 51:407-412.