International Journal of Remote Sensing
Vol. 29, No. 6, 20 March 2008, 1693–1706

Taylor & Francis
Taylor & Francis Group

Spectral prediction of *Phytophthora infestans* infection on tomatoes using artificial neural network (ANN)

X. WANG†§, M. ZHANG*‡, J. ZHU† and S. GENG§

†Dept. of Agronomy, Zhejiang University, Hangzhou, China 310029

‡Dept. of Land, Air and Water Resources, University of California Davis, CA 95616, USA

§Dept. of Plant Sciences, University of California Davis, CA 95616, USA

(*Received 26 June 2006; in final form 10 February 2007*)

Late blight (LB) is one of the most aggressive tomato diseases in California. Accurately detecting the disease will increase the efficiency of properly controlling the disease infestations to ensure the crop production. In this study, we developed a method to spectrally predict late blight infections on tomatoes based on artificial neural network (ANN). The ANN was designed as a back-propagation (BP) neural network that used gradient-descent learning algorithm. Through comparing different network structures, we selected a 3-25-9-1 network structure. Two experimental samples, from field experiments and remotely sensed image data sets, were used to train the ANN to predict healthy and diseased tomato canopies with various infection stages for any given spectral wavelength (μm) intervals. Results of discrete data indicated different levels of disease infestations. The correlation coefficients of prediction values and observed data were 0.99 and 0.82 for field data and remote sensing image data, respectively. In addition, we predicted the field data based on the remote sensing image data and predicted the remote sensing image data with field data using the same network structure, and the results showed that the coefficient of determination was 0.62 and 0.66, respectively. Our study suggested an ANN with back-propagation training could be used in spectral prediction in the study.

1. Introduction

Late blight (LB) is one of the economically most important tomato diseases that is caused by *Phytophthora infestans*. Since 1990, late blight has spread widely to cause crop economic losses across the United States and Canada (Fry and Goodwin 1997a, Goodwin *et al.* 1998). Late blight appears on potato or tomato leaves as pale green, water-soaked spots, often beginning at leaf tips or edges. Late blight can also develop on green tomato fruit and often appears on the sides or upper fruit surfaces. Moreover, if environmental conditions were favourable for disease spread in the production areas, it would potentially result in tremendous loss for production (Fry and Goodwin 1997b). Therefore, it is of great economic interest to control the spread of the late blight disease. Remote sensing offers a powerful tool for monitoring crop growth and health conditions (Abou-Ismail *et al.* 2004, Qin and Zhang 2005). For this study, the goal is to detect crop health conditions in the field

*Corresponding author. Email: mhzhang@ucdavis.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160701281007

by classifying the differentially diseased canopies as well as healthy patches using the artificial neural network (ANN).

Neural networks are very useful tools in pattern recognition because they allow classification without the need for explicit recognition rules (Bishop 1995). An ANN is a system that consists of a large number of simple processing units, called neurons, as in the nervous system. A neuron generally has a high-dimensional input vector and a simple output signal. The function to be performed on the input vector is hence defined by the non-linear function and the weight vector of the neuron. The strength of an ANN is that it trains itself and operates by a pattern of recognition of the data and arrives at a conclusion in an unbiased manner. ANN is traditionally used in the control researches, and in recent years, it has been successfully applied across a large range of domains such as image recognition, medicine, molecular biology and, more recently, ecological and environmental sciences (Iglesias *et al.* 2004, Tutu *et al.* 2005).

A number of researchers have attempted to develop classification and prediction techniques for spectral estimations (Mathers *et al.* 2002 Ding *et al.* 2005, Le *et al.* 2005). Zhang *et al.* (2002, 2004) also reported that the separation of the spectra of healthy tomato plants from infected ones with infection stage 3 or above is reliable due to the consistency of the results from Principle Components Analysis (PCA) and cluster analysis as well as image analysis. Those traditional methods, such as PCA, Discriminate Function Analysis (DFA) and Hierarchical Cluster Analysis (HCA), using spectral data alone do not generally produce high classification accuracy (Tso and Mather 2001). Moreover, these statistical analyses are based on the following assumptions: normal distribution and absence of collinearity among input variables (Johnston 1980). Hence, the objective of this study is to explore the capability of an ANN to build a spectral prediction model for discriminating *Phytophthora infestans* infection on tomatoes.

2. Materials and methods

2.1 *Data collection*

Spectral data were collected from tomato fields in the Salinas Valley, California. The tomato field used for the collections was approximate 40 acres in size and the spectra were collected at an early to middle fruit stage. A detailed description and experimental data can be found in Zhang *et al.* (2004).

We measured the spectral reflectance of canopies at 1 m above the canopy using spectrometer GER-2600. The wavelength of the measurement was in the spectral region range of 400–2500 nm. To take the impact of soil background into account, we also collected the soil spectra in the field. We rated late blight infection severity into five stages (table 1).

Table 1. Late blight infection stages and symptoms.

Infections	Infection Stages	Symptoms
Healthy	No infections	No infection symptom was observed on the leaves
LB1	Infection stage one	One lesion on one or two leaves
LB2	Infection stage two	One lesion on more than two leaves
LB3	Infection stage three	Two lesions can be seen on one or more leaves
LB4	Infection stage four	Two lesions on over half the canopy leaves

Table 2. The ANN influential parameters used during training.

Layer of hidden neurons*	Activation function	Learning rate	Momentum constant	Convergence criterion	Epochs
1	Tan-sigmoid	0.05	0.9	1e-10	2000
2	Tan-sigmoid, Linear	0.05	0.9	1e-10	2000

Note: *represents the layer number of hidden neurons, for example, 1 layer denotes one hidden layer (i.e. 4, 8, 12 neurons) and 2 is two hidden layers (i.e. 25-12, 25-9, 12-8, 12-4 neurons).

Advanced Visible Infrared Imaging Spectrometer (AVIRIS) hyperspectral images were acquired at low flight altitude for the fields containing late blight infections in Salinas Valley, California. The spectra had 224 bands with a band span of ~10 nm at 400–2500 nm. However due to inadequate quality of some bands in the original image, we conducted a preliminary image process to eliminate those low quality bands which resulted in a hyperspectral image with 180 bands for the analysis. The image had a spatial resolution of 4 m. The ATmospheric REMoval program (ATREM) developed by University of Colorado (Center for the Study of Earth From Space 1999) was used to calibrate the image before analyses. The spectra were extracted from the pixels with known disease infection stages and these AVIRIS spectra were then used for developing the prediction models using ANN.

2.2 *The Artificial Neural Networks (ANNs)*

In the early 1940s, McCulloch and Pitts (1943) explored the competitive abilities of networks made up of theoretical mathematical models when applied to the operation of simple artificial neurons. The structure of an ANN defines the overall architecture of the network, including one input layer, one output layer, and usually one or two hidden layers (figure 1). Each neuron receives a weighted sum from each neuron in the preceding layer and provides an input to each neuron of the next layer. Thus,

$$net = \sum_{i=1}^{n} W_i X_i \tag{1}$$

where *net* is the summation of the input signal, W_i denotes an element of the weight vector **w**, and X_i is an element of the input vector **x**. For a given network and input vector, the output vector is totally determined by the weights. The process of finding

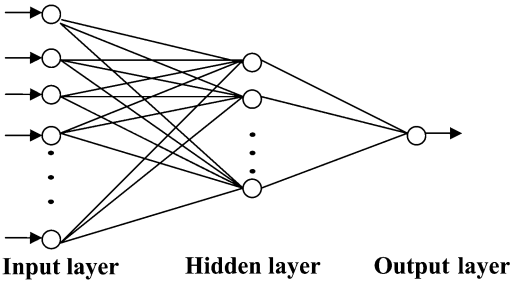

Figure 1. Typical ANN structure.

optimal weights is called "training". The training algorithm used in this study was back-propagation. In this process, the input units and their desired output value are set for the network. The activations of the units are then calculated, feeding forward layer-by-layer from the inputs to the output. A logistic threshold function was used,

$$O = \frac{1}{1+e^{-net}} \tag{2}$$

where O is the output of the network. Once the network output value has been produced, it is compared with the target output specified in the training data set. Following this comparison, a backwards adjustment of the weights is performed and the training is stopped when the minimum error for test data found.

$$MSE = \frac{1}{n}\sum_{1}^{n}(P-O)^2 \tag{3}$$

where MSE is the mean square error considering prediction (P) and observed values (O) for n testing data vectors. The correlation coefficient (r) or determination coefficient (R^2) was used to evaluate the prediction. When $r=1$, there is a perfect positive linear correlation between P and O. When $r=-1$, there is a perfectly linear negative correlation between P and O. When $r=0$, there is no correlation between P and O. Intermediate values present partial correlation.

In our study, the input vector consists of three values (wavelength, soil value and spectral value). These variables were all normalized. The output values were encoded as 0–1 [1 for healthy tomatoes, 0.75 for tomatoes with infection stage 1 (LB1), 0.5 for tomatoes with infection stage 2 (LB2), 0.25 for tomatoes with infection stage 3 (LB3) and 0 for tomatoes with infection stage 4 (LB4), respectively] for different levels, respectively. The field data and remote sensing image data were divided into training and testing data sets. The ANN parameters were: learning rate 0.05, the momentum constant 0.9 and the values of MSE were 10^{-10}. The only difference is the activation function, Tan-sigmoid for one hidden layer and Tan-Sigmoid and Linear for two hidden layers, respectively. The success of training was determined with the average sum square value between desired output vector and the predicted value.

3. Results

3.1 *Waveband selection*

The spectra used in this study cover the full range of wavelengths, 400–2500 nm. The Visible/Near Infrared (VNIR) portion of the spectrum (400–1050 nm) has a spectral resolution of approximately 3 nm at 700 nm wavelength. The short-wave infrared (SWIR) region is measured by two detectors: SWIR1 (900–185 nm) and SWIR2 (1700–2500 nm). The higher reflectance is normally seen in the NIR range than in visible range for the vegetation. In order to find the best discriminating wavebands between diseased and healthy spectra, we select the representative spectral reflectance, whose wavelength is from 750 to 1350 nm for field collected data and from 700 to 1105 nm for image extracted data in this study (figure 2(*a*), 2(*b*)). The reflectance of the vegetation canopies in these regions shows an obvious difference between different infection stages. Compared with field spectra, the differences in remote sensing spectra are smaller in the full range of wavelengths. The spectra differences of healthy plants and plants with LB1 are higher for the field collected

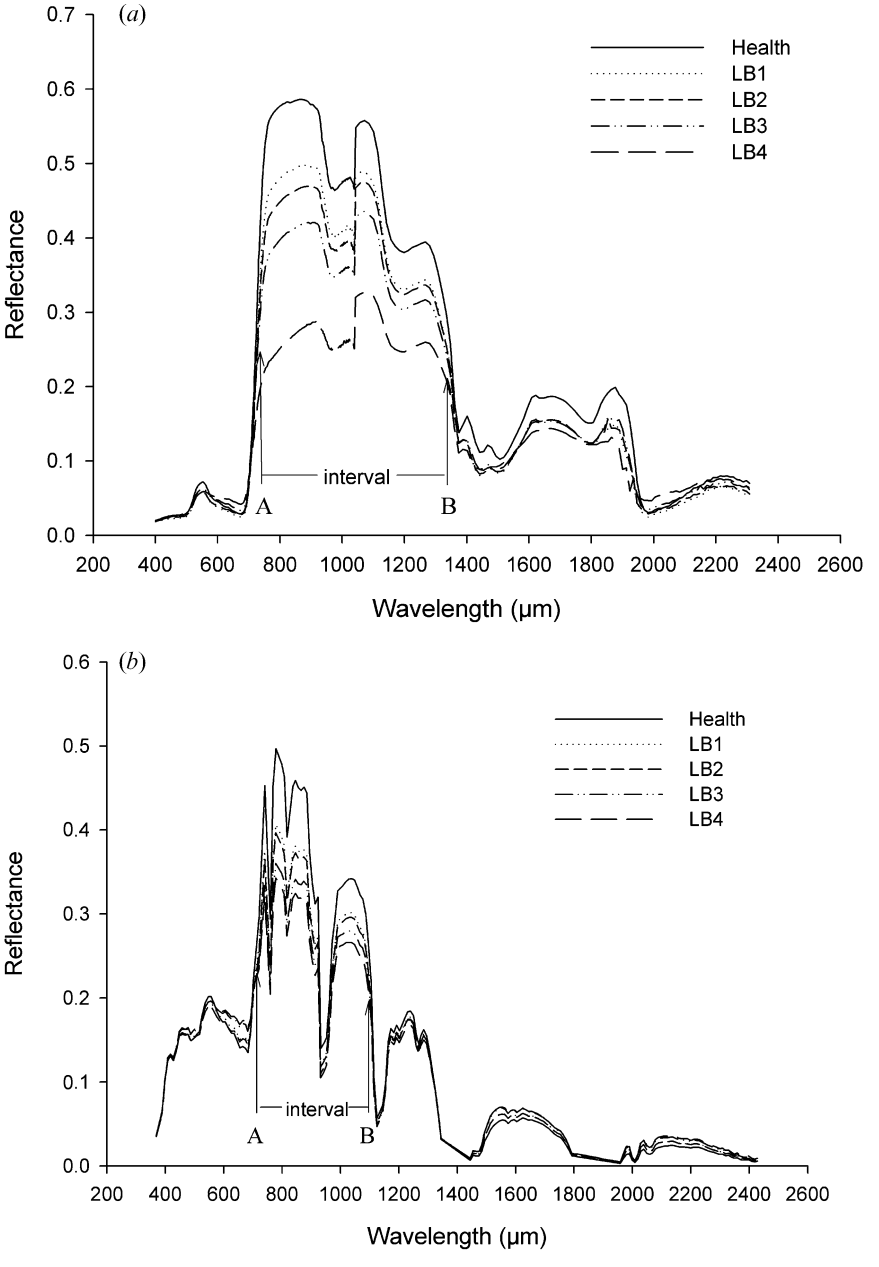

Figure 2. Mean spectra of the healthy and infected tomato plants at infection stages 1, 2, 3 and 4. The range of two arrows denotes sampling interval. (*a*) Represent the field data and (*b*) denote the remote sensing data.

data than for the image extracted data. Hence, the ultimate selection of the spectra data is then used to build for ANN analysis.

3.2 *Network selection*

A neural network that uses gradient-descent error learning is designed and used in our prediction. The neural network has one input and one output. In the training of

a BP neural network, three inputs (wavelength, soil value and spectral value) and one output vector sets are generated from the experiments, half of the total number of field data points (537/1074) and of image data points (192/392) are used as the training set. Due to the characteristic of sigmoid activation function, training data are standardized and thus the training data are scaled to a range of 0.1 ± 0.9 (rather than 0.0 ± 1.0). The number of neurons in the hidden layer is generally selected from the different levels, such as one hidden layer (i.e. 4, 8, 12 layers) and two hidden layers (i.e. 25-12, 25-9, 12-8, 12-4), and 25-9 neurons in the hidden layer were found to be successful in the training process for both field data and image data. The training process has been completed in approximately 2000 iterations. When the training is completed, a neural network is designed using the obtained weights. The logarization of MSE for the different networks are presented in figure 3(*a*) and figure 3(*b*) for the two layers network, the log MSE is much bigger than for the three layers. In the three layers network, a 3-25-9-1 network is chosen and it can be successfully modelled for the current data.

3.3 *ANN prediction*

We conducted the prediction of field data sets and image data sets using the 3-25-9-1 ANN. Training data setting is based on the 50%–50% sampling technique. Initially, one-half of the cases are randomly selected and used with the BP network. Subsequently, the BP networks are trained for 2000 iterations. The root mean squared training error achieved by the networks is approximately 10^{-10}. Determining the training ending point for the BP network is a tricky task. After training, the correlation coefficient between training and predictive values reaches 0.98 for field data and 0.99 for image data, respectively. Because of large sample size (98 training data points) of the image data, the correlation coefficient for image data was higher.

As shown in figure 4, when the output values calculated by ANN are compared with the desired values, the correlation coefficient reached 0.99 for field data and 0.82 for image data. The correlation of $r=0.99$ indicates a reasonable agreement between these values. However, this model is weaker ($r=0.82$) in predicting the disease infection in image data. This is because, in the image extracted data, there were two predicted values (one for healthy plants and the other for the plants with LB4 infections) deviated from the observed values. When we remove the two values, the correlation coefficients increased to 0.93.

The sensitivity, specificity, and area under the ROC curve across different cutoff scores indicate that a cutoff score of 4 results in the least error of classification (figure 5). The ROC curves rise rapidly and then reach a plateau for the two samples. The analysis of field data area under ROC curve (0.994) did not differ from that of image data (0.992). This indicates a high sensitivity and a low false-positive rate and, therefore, an adequate degree of accuracy.

3.4 *Using image extracted spectra data to predict the field disease conditions*

Due to fewer surveyed points in image data, we divided the field data into discrete intervals and averaged the points within each one, and two data sets therefore could be matched, as illustrated in figure 6. As mentioned before, the data set is randomly divided in two subsets with 196 spectral data points. Initially, the first data set is used to train the BP network and determine the *MSE* value. It is found that 500 iterations are sufficient for the task. A linear regression analysis comparing the

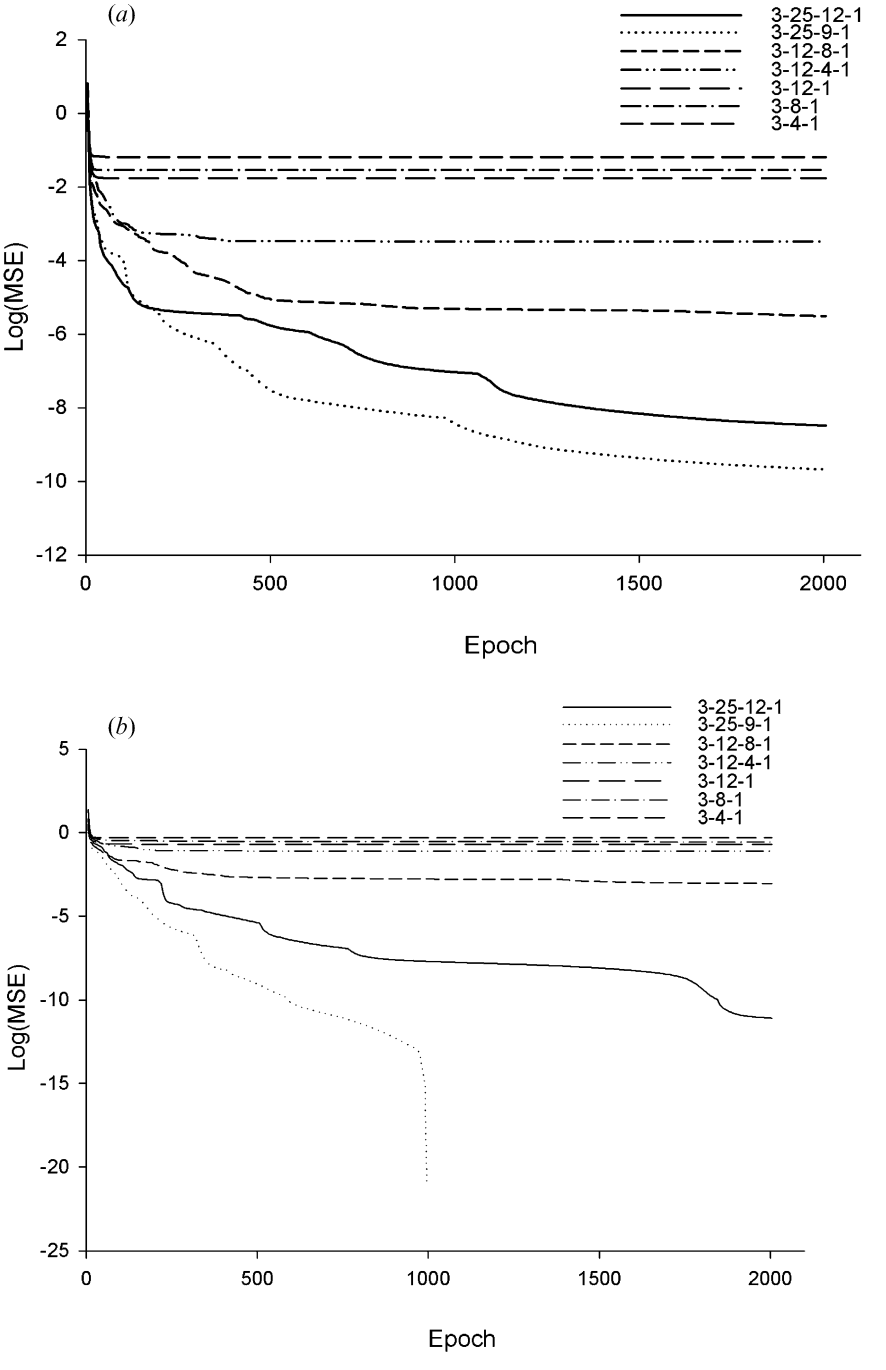

Figure 3. Legalisation of training MSE curve of backpropagation neural network for field data and remote sensing data.

outputs to observed data produces a coefficient of correlation of 0.78, indicating that approximately 61% (determination coefficient, $R^2=0.61$) of the variability of field data is explained by remote sensing image data alone (figure 7), and that only approximately 39% of the variability is caused by other factors. However, the

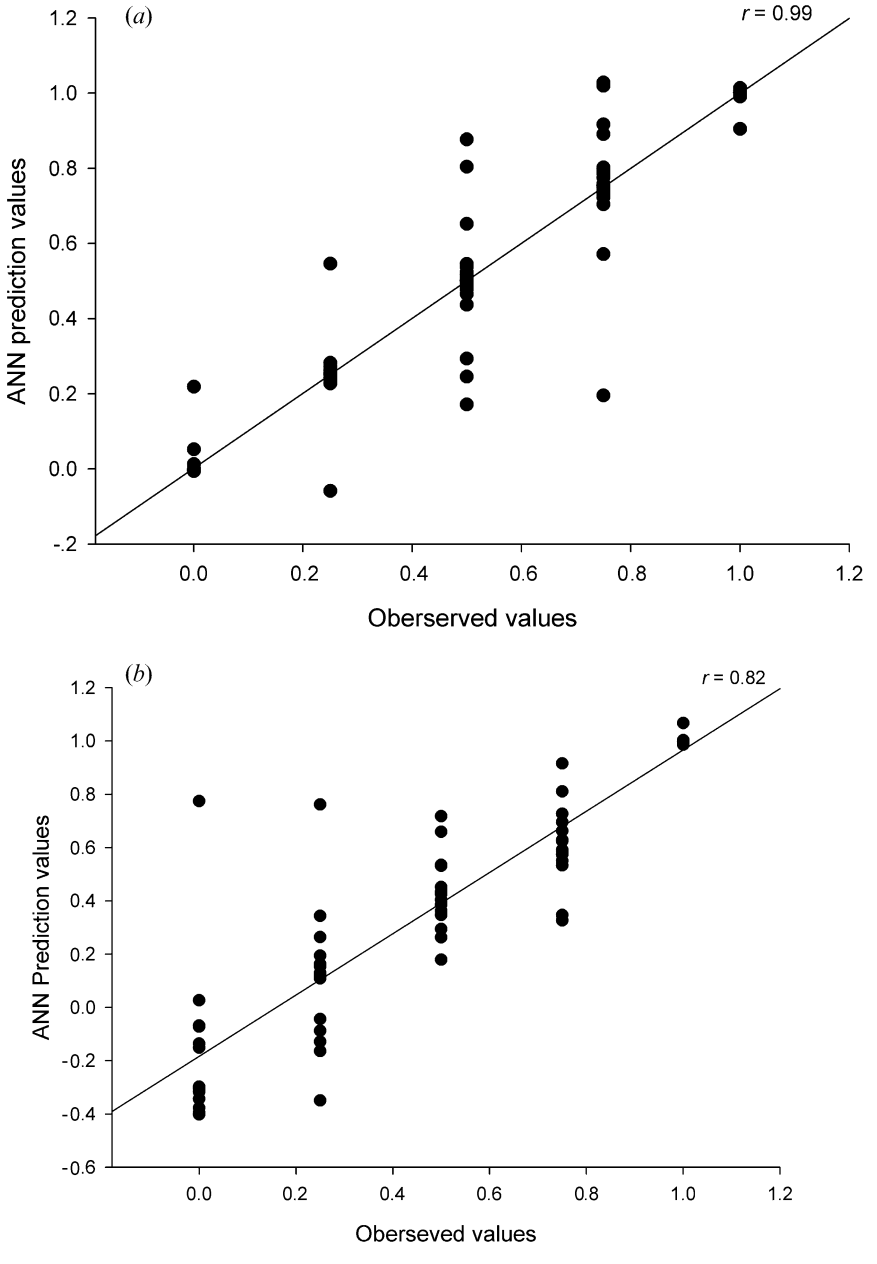

Figure 4. Oberved versus predicted values of artificial neural network for field data and remote sensing data.

application of remote sensing is usually used for image classification based on ground truth information. This backward approach is not often used in applications.

3.5 *Using field data to predict disease conditions in the image extracted data*

When we used the same network structure, and used field spectra data to predict the field disease conditions from the image extracted spectra data, the results were

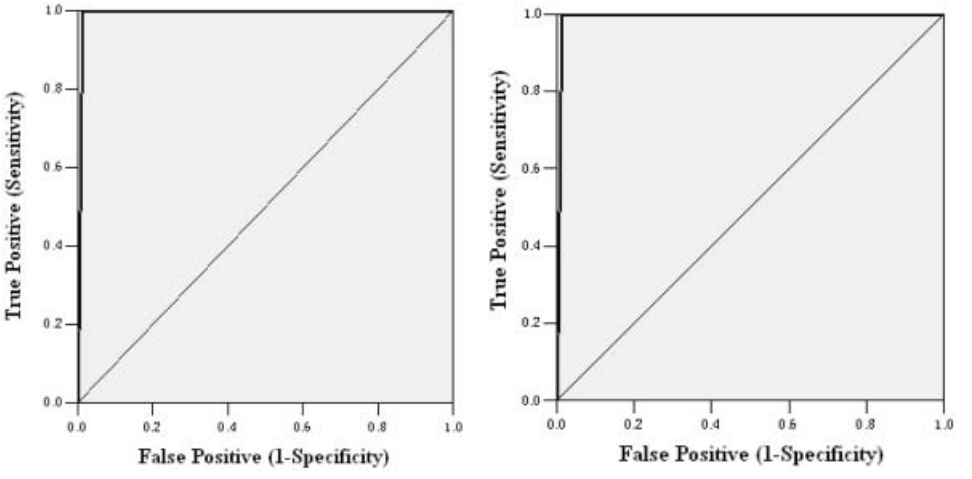

Figure 5. ROC curves for ANN prediction of field and image data. The diagonal line indicates an index that performs no better than chance.

relatively successful. The results reveal a higher correlation coefficient ($r=0.81$) than prediction of field data using image data (figure 8). In other words, the accuracy of prediction for the disease infection increased when we predicted the infection of field disease conditions through image extracted data using the field collected data as the training set. Figure 8 also indicates that it is possible to use remote sensing to monitor large field disease conditions and the health of the crop growth.

4. Discussion

In our study, ANNs have been shown to be able to model and predict infection levels of late blight on tomatoes based on the input spectral and soil data. This approach provides quite highly accurate classification of infection for both field collected spectra and image extracted spectra. Regarding the use of ANN method in spectral analysis, several authors have reported its superiority in predicting infection stage and over other statistic methodologies (Uno *et al.* 2005). From table 3, we can see that the prediction for field spectra data is much better than that for the image extracted spectra data. In this study, the correct classification rate of ANN prediction for stage 1 are 100% for both field data and image data. Therefore, it can be said that the ANN method can accurately classify the healthy and diseased tomato plants. Since stage 2 contains 3.7% and 4% of spectra samples with the healthy tomato plants, 0.9% and 28% with the infected plants of stage 3, 0.9% and 8% stage 4 for field data and image data, respectively. The stage 3 includes 2.8% and 20% of spectra samples with the infected plants of stage 4. For the image data, it is difficult to distinguish the spectra with infection stage 4 and stage 5. The stage 5 contains only 1 stage 4 for both field data and image data, respectively.

It is also important for many applications to utilize the appropriate classification methods. To some extent, the hierarchical cluster is limited to classifications, which shows the drawbacks of unsupervised classification. For comparison purposes, table 4 summarizes the results of various cluster classification methods. The Healthy, LB1, LB2, LB3, LB4 are clustered separately under complete linkage, unweighted average distance (UPGMA), weighted average distance (WPGMA),

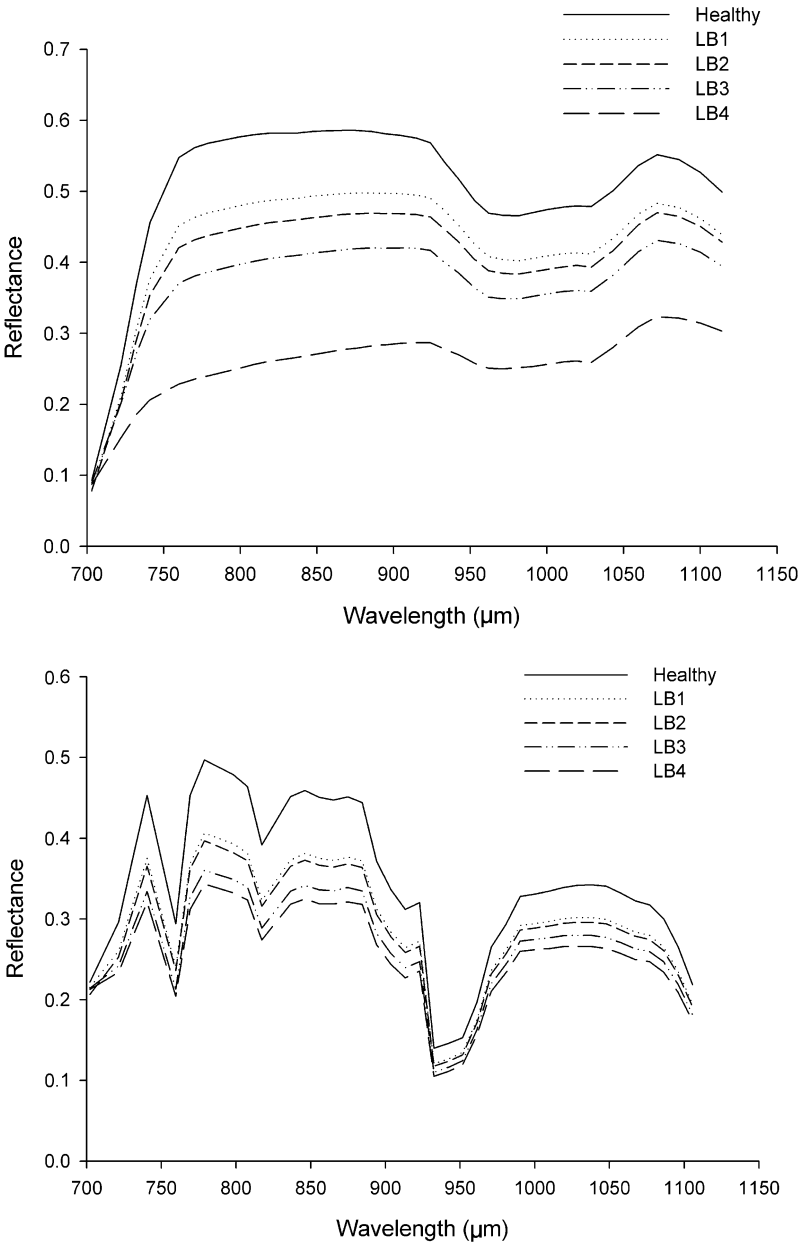

Figure 6. Sampling spectra for field data and remote sensing data.

unweighted centre of mass distance (UPGMC), and weighted centre of mass distance (WPGMC) methods. It is clear that the cluster effects are much smaller than the ANN method which was described in our study.

Although spectral reflectance depended on amount of green vegetation, variation in leaves and/or canopies alone did not explain all the variability in spectral values. There are several possible reasons for the inconsistencies between the field spectra and the image spectra. First of all, soil reflectance affects spectral measurements. Second, light responses vary with different biological features, such as green leaf,

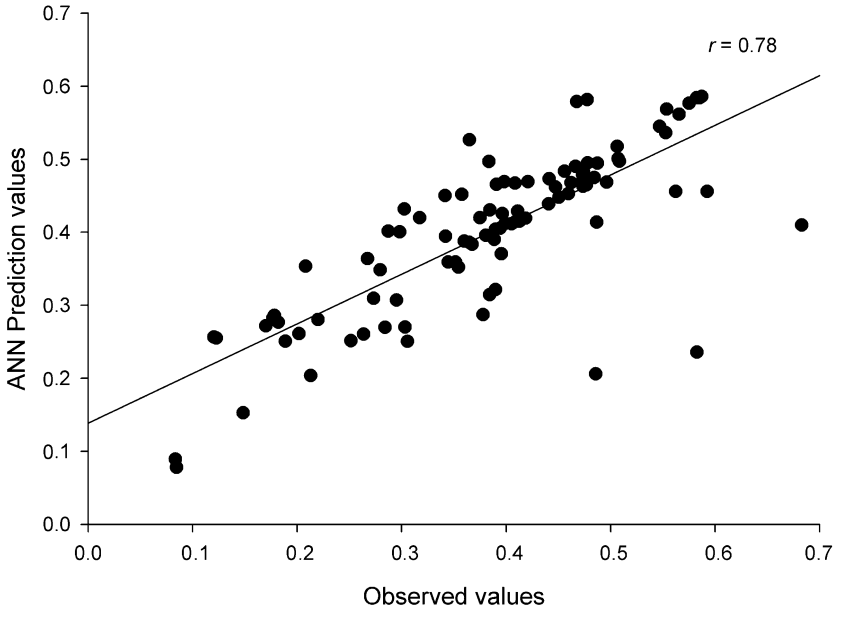

Figure 7. Scatter plot of remote sensing data, field data and least-squares regression line (r=0.78).

leaf litter, and bark optical properties. Finally, spectral collection for field and image data was performed at different times.

Generally speaking, all modelling methods have embedded inherent strengths and weaknesses in some aspects. ANNs act as a "black box" approach to the description

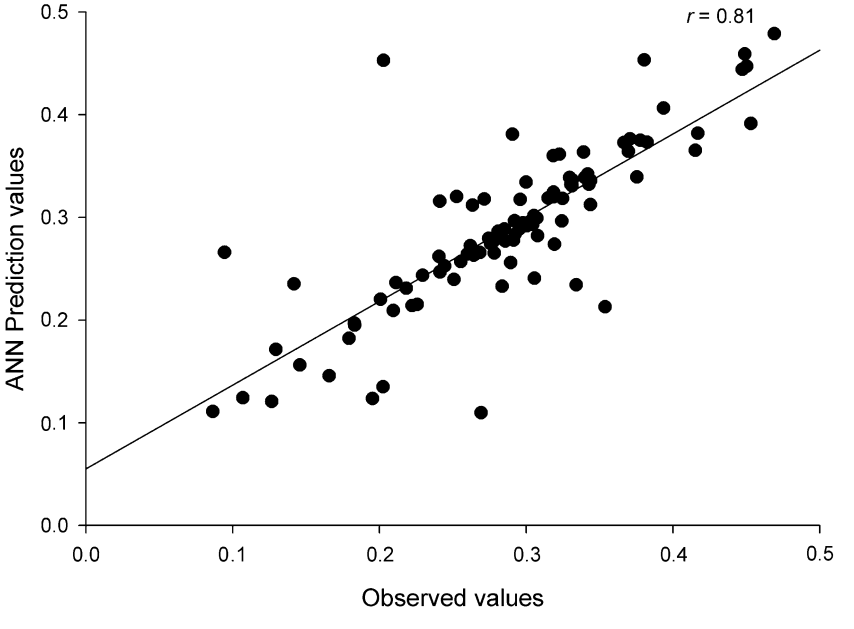

Figure 8. Scatter plot of field data, remote sensing data and least-squares regression line (r=0.81).

Table 3. Number and percentage of uncorrect classification of spectral samples.

Level	Field data (107)	Image data (25)
1		
2	LB1:4 (3.7%)	LB1:1 (4%)
	LB3:1 (0.9%)	LB3:7 (28%)
	LB4:1 (0.9%)	LB4:2 (8%)
3	LB1:1 (0.9%)	LB2:2 (8%)
	LB2:2 (1.9%)	LB4:5 (20%)
	LB4:3 (2.8%)	
4	LB3:1 (0.9%)	LB5:8 (32%)
	LB5:1 (0.9%)	LB3:1 (4%)
5	LB4:1 (0.9%)	LB3:1 (4%)

of the relationship between the input variables and the output layer. Even though ANN predicts the biological features, it will not tell us about the functional form between the input and output variables. The weight matrices of the network do not have any direct meaning. Namely, the performance of ANN depends on the quality and the number of training samples, rather than its own good generalization capability. Considering the limited number of image extracted data sets used to train the network, additional data points would undoubtedly improve the performance of the model for the remote sensing prediction. In our study, we only obtain 195 sample data sets in the range of 700–1130 nm which is only 1/10 of field data sets. Accounting for the difference between field measurement and remote sensing sampling techniques, we should concentrate more on how to improve the accuracy remote sensing data, such as adding sampling data points. In our study, although the ANN model showed a better predictability, we could not obtain an accuracy greater than 90%. This may be influenced by training data, and further studies should be conducted to improve these parameters.

The analysed results of the remote sensing image data and field spectra data provide useful information for agricultural field management (Bryant and Moran

Table 4. Comparison between performances of the five hierarchical cluster methods.

Cluster / Method	Cluster 1 (215)	Cluster 2 (215)	Cluster 3 (215)	Cluster 4 (215)	Cluster 5 (215)
complete	LB1: 134	LB2: 139	LB2: 126	LB3: 79	LB3: 7
	LB2: 69	LB3: 11	LB3: 11	Lb4: 136	Lb5: 208
	Lb4: 12	Lb4: 65	Lb4: 78		
UPGMA	LB1: 134	LB2: 139	LB2: 126	LB3: 79	LB3: 7
	LB2: 69	LB3: 11	LB3: 11	Lb4: 136	Lb5: 208
	Lb4: 12	Lb4: 65	Lb4: 78		
WPGMA	LB1: 139	LB2: 202	LB2: 178	LB2: 96	LB3: 7
	LB2: 64	LB3: 11	LB3: 11	LB3: 12	Lb5: 208
	Lb4: 12	Lb4: 2	Lb4: 26	Lb4: 107	
UPGMC	LB1: 134	LB2: 127	LB2: 87	LB3: 81	LB3: 7
	LB2: 69	LB3: 11	LB3: 11	Lb4: 134	Lb5: 208
	Lb4: 12	Lb4: 77	Lb4: 117		
WPGMC	LB1: 142	LB1: 71	LB2:126	LB2:12	Lb4: 7
	Lb2: 61	LB2: 71	LB3: 11	LB3: 3	Lb5: 208
	Lb3: 12	LB3: 62	Lb4:78	Lb4: 200	
		Lb4: 11			

1999 Zhang *et al*. 2002, Karpouzli and Malthus 2003). This study is aimed at testing the feasibility of looking for spectral differences between healthy and *Phytophthora infestans* infected tomatoes with ANN combining remote sensing image data with field spectra data. The fungus *Phytophthora infestans* caused the late blight that spreads quickly in tomato fields. Once the tomatoes are infected, the symptoms of late blight first appear on leaves, which will gradually change colour from green to yellow, then the fungus will infect the stems and fruits. The spectral reflectance of green vegetation in the red band (0.6–0.7 μm) is most sensitive to leaf chlorophyll and pigment contents while the near infrared (NIR) band (0.7–0.9 μm) is most sensitive to biomass (Kurschner et al. 1984, Blakeman 1990). The sampling strategy is affected by the leaf canopy of healthy and diseased plants (Barker and Pinard 2001). The reflection of incident energy depends on both the properties of tomatoes leaves and the structure of its cover. If the cover is less than 100%, then the reflectance properties of the underlying soil will also be captured in the remotely sensed data. Therefore, future work is necessary to validate these findings for canopy reflectance with greater variability before this method will be ready for adoption by the farmers. Understanding the impact of disease on levels of pigments will also assist in future studies. In this study, our results showed that the ANN method is successful for discriminating between healthy and unhealthy tomatoes with high accuracy. Farmers will be able to use the predicted disease result from hyperspectral remote sensing images to prevent the spread of disease before significant economic loss occurs. Our study provides a means for precise disease management for tomatoes.

References

ABOU-ISMAIL, O., HUANG, J.F. and WANG, R.C., 2004, Rice yield estimation by integrating remote sensing with rice growth simulation model. *Pedosphere*, **14**, pp. 519–526.

BARKER, M.G. and PINARD, M.A., 2001, Forest canopy research: sampling problems and some solutions. *Plant Ecology*, **153**, pp. 23–38.

BISHOP, C., 1995, *Neural Networks for Pattern Recognition* (New York: Oxford University Press).

BLAKEMAN, R.H., 1990, The identification of crop disease and stress by aerial photography. In Application of Remote Sensing in Agriculture, M.D. Steven and J.A. Clark (Eds), pp. 229–254 (London: Butterworths).

BRYANT, R.B. and MORAN, M.S., 1999, Determining crop water stress from crop temperature variability. *Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition/21st Canadian symposium on Remote Sensing, Ontario, Canada, 21–24 June 1999*, pp. 19–26 (Ann Arbor, MI: ERIM International Inc.).

Center for the Study of Earth From Space (CSES) 1999, ATmosphere REMoval Program (ATREM) User's Guide Version 3.1. Center for the Study of Earth From Space (CSES), Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder.

DING, C., HE, X.F. and SIMON, H.D., 2005, Nonnegative Lagrangian relaxation of K-means and spectral clustering. *Lecture Notes in Artifical Intelligence*, **3720**, pp. 530–538.

FRY, W.E. and GOODWIN, S.B., 1997a, Reemergence of potato and tomato late blight in the United States. *Plant Disease*, **81**, pp. 1359–1357.

FRY, W.E. and GOODWIN, S.B., 1997b, Resurgence of the Irish potato famine fungus. *Bioscience*, **47**, pp. 363–370.

GOODWIN, S.B., SMART, C.D., SANDROCK, R.W., DEAHL, K.L., PUNJA, Z.K. and FRY, W.E., 1998, Genetic change within populations of Phytophthora infestans in the United States and Canada during 1994 to 1996: role of migration and recombination. *Phytopathology*, **88**, pp. 939–949.

IGLESIAS, A., VARELA, B.A., COTOS, J.M., TABOADA, J.A. and DAFONTE, C., 2004, A comparison between functional networks and artificial neural networks for the prediction of fishing catches. *Neural Computing and Applications*, **13**, pp. 24–31.

JOHNSTON, R.J., 1980, *Multivariate Statistical Analysis in Geography: A Primer on the General Linear Model* (New York: Longman Scientific and Technical).

KARPOUZLI, E. and MALTHUS, T., 2003, The empirical line method for the atmospheric correction of IKONOS imagery. *International Journal of Remote Sensing*, **24**, pp. 1143–1150.

KURSCHNER, E., WALTER, H. and KOCH, W., 1984, Measurements of spectral reflectance of leaves as a method for assessing the infestation with powdery mildew. *Journal of Plant Disease Protection*, **91**, pp. 71–80.

LE, M.L., VALENSI, P., CHARNIOT, J.C., HANTZ, E. and ALBERTINI, J.P., 2005, Serum H1-nuclear magnetic spectroscopy followed by principal component analysis and hierarchical cluster analysis to demonstrate effects of statins on hyperlipidemic patients. *NMR in Biomedicine*, **18**, pp. 421–429.

MATHERS, N.J., XU, Z.H., BERNERS-PRICE, S.J., SENAKE PERERA, M.C. and SAFFIGNA, P.G., 2002, Hydrofluoric acid pre-treatment for improving C-13 CPMAS NMR spectral quality of forest soils in south-east Queensland, Australia. *Australian Journal of Soil Research*, **40**, pp. 655–674.

MCCULLOCH, W.S. and PITTS, W., 1943, A logical calculus of the ideas immanent in nervous activity. *Bulletin Mathematics Biophysics*, **5**, pp. 115–133.

QIN, Z. and ZHANG, M., 2005, Detection of rice sheath blight for in-season disease management using multispectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, **7**, pp. 115–128.

TUTU, H., CUKROWSKA, E.M., DOHNAL, V. and HAVEL, J., 2005, Application of artificial neural networks for classification of uranium distribution in the Central Rand goldfield, South Africa. *Environmental Modeling & Assessment*, **10**, pp. 143–152.

TSO, B. and MATHER, P.M., 2001, *Classification methods for remotely sensed data* (London: Taylor and Francis).

UNO, Y., PRASHER, S.O., LACROIX, R., GOEL, P.K., KARIMI, Y., VIAU, A. and PATEL, R.M., 2005, Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Computers and Electronics in Agriculture*, **47**, pp. 149–161.

ZHANG, M., LIU, X. and O'NEILL, M., 2002, Spectral Discrimination of Phytophthora infestans infection on tomatoes based on principal component and cluster analyses. *International Journal of Remote Sensing*, **23**, pp. 1095–1107.

ZHANG, M. and QIN, Z., 2004, Spectral analysis of tomato late blight infections for remote sensing of tomato disease stress in California. *IEEE 2004 International Geosciences and Remote Sensing Symposium, VI: 4091–4094*, Anchorage, Alaska, USA.