ESA–MOST Dragon Cooperation

中国科技部–欧洲空间局“龙计划”合作

2018 DRAGON 4 MID-TERM RESULTS SYMPOSIUM

2018年“龙计划”四期中期成果国际学术研讨会

19–22 June 2018 | Xi’an, P.R. China

2018年6月19日–22日，中国 西安

ENHANCE WHEAT POWDERY MILDEW MONITORING WITH LIMITED SAMPLE DATA BASED ON OPTIMIZED TRADABOOST ALGORITHM

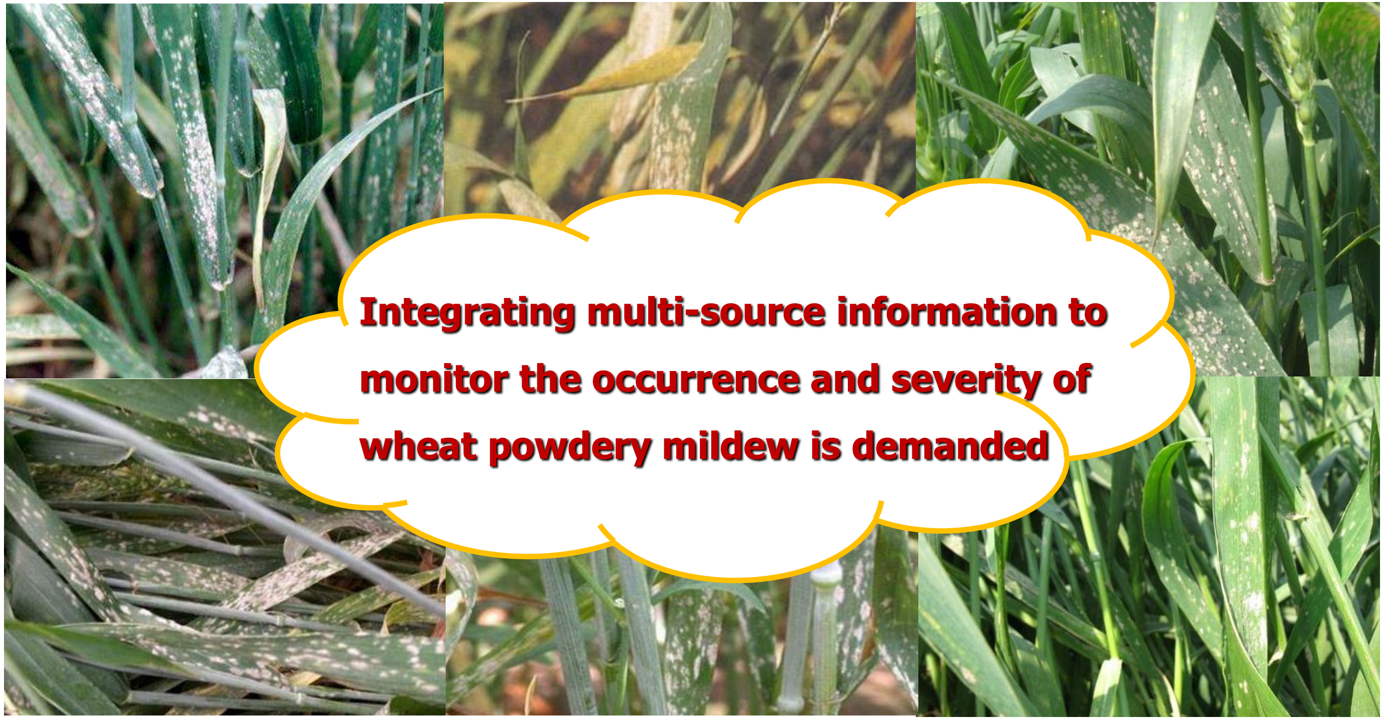
Linyi Liu, Wenjiang Huang, Giovanni Laneve, YueShi, Qiong Zheng, Huiqin Ma, Pablo Marzialetti

Outline

01 **Introduction**

02 **Materials and methods**

03 **Result**



Wheat powdery mildew is one of the most destructive diseases in China. It could lead to a significant yield loss and grain quality reduction

1.The available time window for monitoring the infection of powdery mildew is approximately only **one month**

start at the booting stage

end at the filling stage

2. The field inspection often requires large amounts of field work and much time for data postprocessing

The available ground samples are always not enough and this problem brings difficulties for high accuracy of disease monitoring

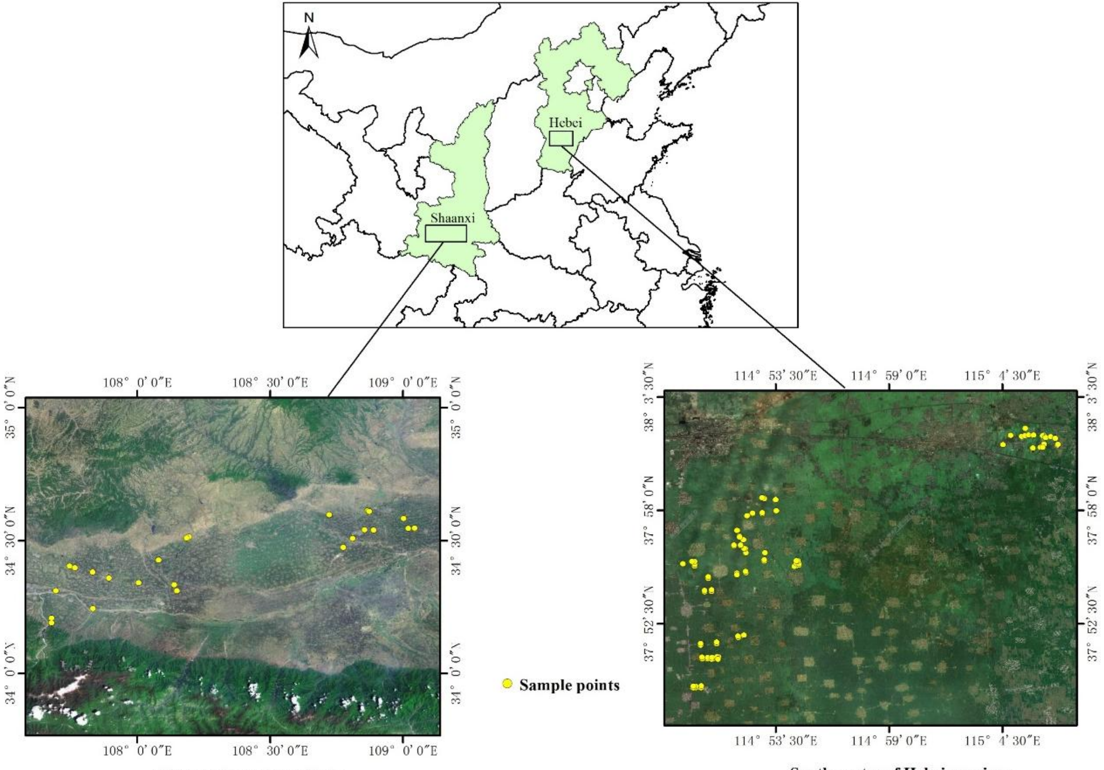
In order to enhance the accuracy of wheat powdery mildew under limited sampling, an optimized TrAdaBoost algorithm was constructed in this study to improve the representative and effective properties of the samples in our research area

Outline

01 Introduction

02 **Materials and methods**

03 Result



Western Guanzhong Plain

South-center of Hebei province

Source dataset: 39

western Guanzhong Plain, Shaanxi province

auxiliary dataset:106

south-center of Hebei province

Region	Type of data	Source of data	Acquired time	Spatial resolution	Time resolution
Western Guanzhong Plain	Remote sensing data	Landsat 8/OLI	2014.5.11	30m	16 days
	Meteorological data	Climate Hazards Group InfraRed Precipitation with Station data(CHIRPS)	2014.3.1—2014.5.11	0.05°	1 day
		The MODIS/Terra Land Surface Temperature and Emissivity (LST/E) product(MOD11A1)	2014.3.1—2014.5.11	1km	1 day
	Field survey data	Field work	2014.5.8—2014.5.10		
South-central of Hebei Province	Remote sensing data	Landsat 8/OLI	2014.5.22	30m	16 days
	Meteorological data	Climate Hazards Group InfraRed Precipitation with Station data(CHIRPS)	2014.3.1—2014.5.22	0.05°	1 day
		The MODIS/Terra Land Surface Temperature and Emissivity (LST/E) product(MOD11A1)	2014.3.1—2014.5.22	1km	1 day
	Field survey data	Field work	2014.5.23—2014.5.28		

Landsat 8 image is commonly used in crop disease monitoring. In this study, the preprocessing of Landsat 8 image included radiometric calibration and atmospheric correction

CHIRPS is a 30+ year **quasi-global rainfall dataset** and it incorporates 0.05° resolution satellite imagery with in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring

MOD11A1 provides per-pixel **temperature** and emissivity values, which are produced daily using the generalized split-window LST algorithm.

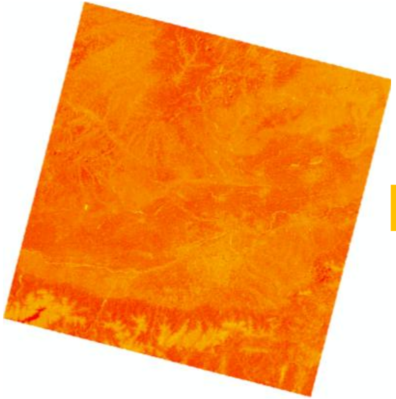
Variables indicating growth status and environmental conditions of wheat

NDVI

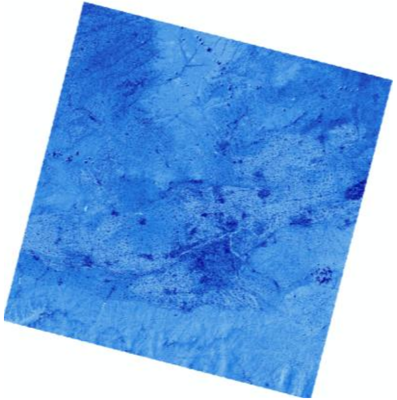
EVI

Wetness

Greenness

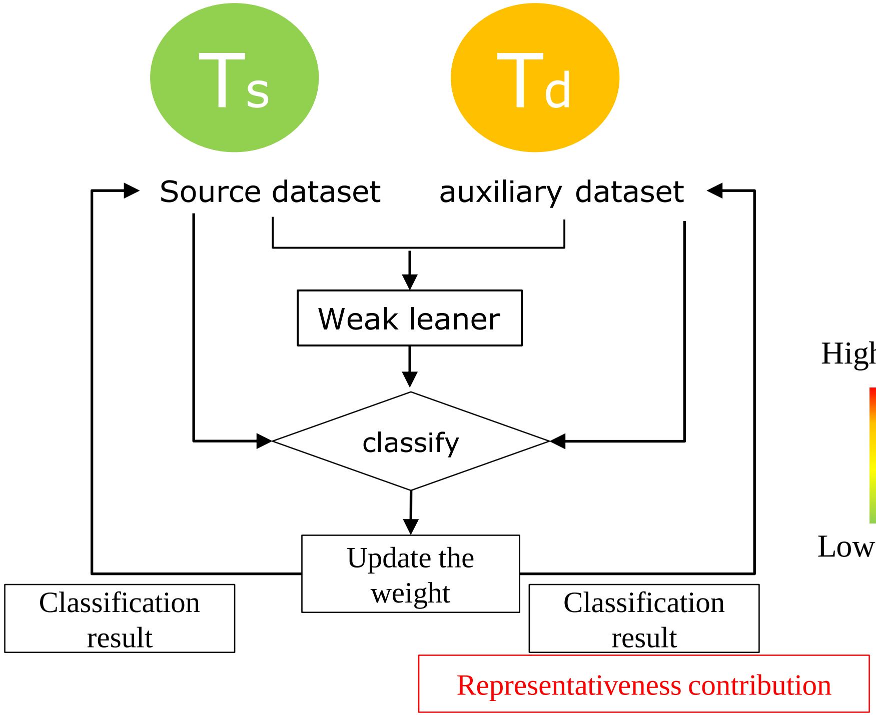


Average LST



Average precipitation

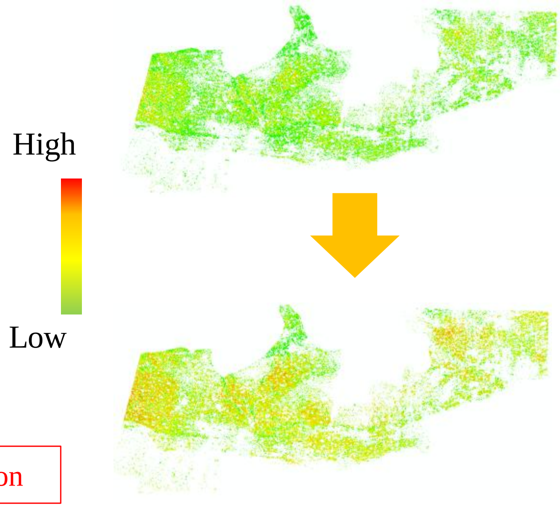
Optimized TrAdaBoost algorithm



Representativeness contribution

Analyze the representativeness of the source dataset added with each sample from auxiliary dataset

Zhu, A. X.; Liu, J.; Du, F.; Zhang, S. J.; Qin, C. Z.; Burt, J.; Behrens, T.; Scholten, T., Predictive soil mapping with limited sample data. *European Journal of Soil Science* **2015,** 66, (3), 535-547.



The number of pixels (Mi) at which the prediction uncertainty was reduced

Total decrement of prediction uncertainty (Vi)

Contribution= Mi* Vi

For t=1,2,…,N

Set $P^t = W^t/(\Sigma_{i=1}^{145} w_i^t)$

Call SVM, providing it the combined training set T with the distribution P^t over T, then get back a hypothesis h_t

Pick some samples(n) according to the P^t

Calculate the error of h_t on Ts:

$$e_t = \frac{w_i^t \cdot |h_t(x_i) - c(x_i)|}{\Sigma_1^{38} w_i^t}$$

(If e_t is less than 0.5, return to call SVM)

Set $\beta_t = e_t/(1-e_t)$ $\quad \beta = 1/(1+\sqrt{2In(n/N)})$

Update the new weight vector:

$$w_i^{t+1} = \begin{cases} w_i^t \beta^{|h_t(x_i)-c(x_i)|\times Contribution_i}, & 1 \le i \le n \\ w_i^t \beta_t^{-|h_t(x_i)-c(x_i)|}, & n+1 \le i \le n+m \end{cases}$$

Output the result:

$$h_f(x) = \begin{cases} 1, & \Pi_{t=\lceil N/2\rceil}^{N} \beta_t^{-h_t(x)} \ge \Pi_{t=\lceil N/2\rceil}^{N} \beta_t^{-1/2} \\ 0, & \text{otherwise} \end{cases}$$

Using one-versus-one method to get the final class

$$\text{RBFSVM} \begin{cases} \text{The Gaussian width: } \sigma \\ \text{The regularization parameter: C} \end{cases}$$

The number of iterations: N

The number of samples picked: n

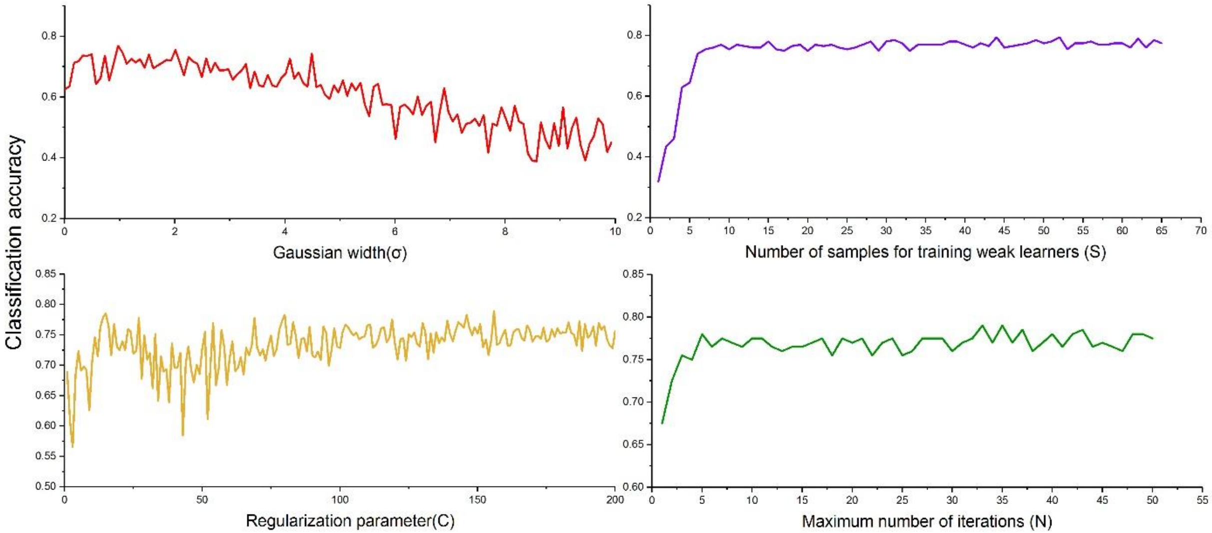
Outline

01 Introduction

02 Materials and methods

03 Result

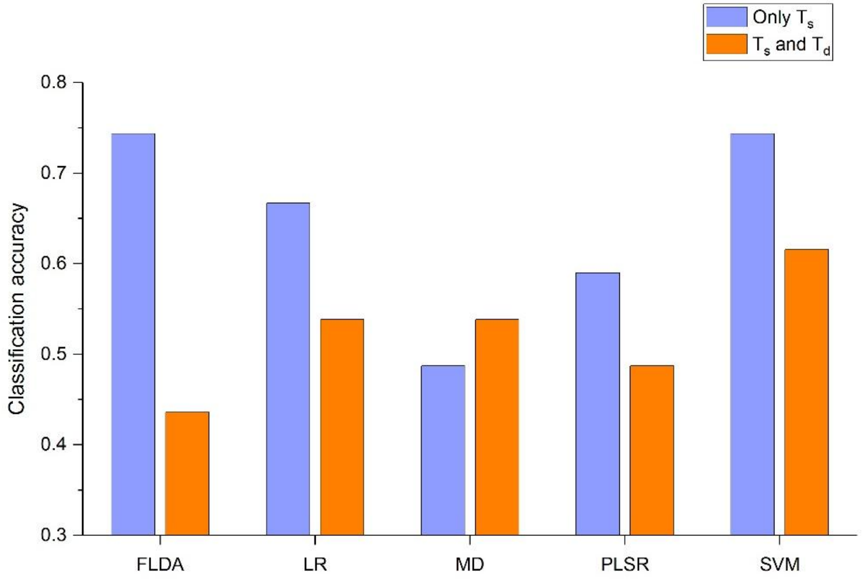
Parameter tuning of TrAdaBoost optimization algorithm



1. The variation of σ has more impact on the final performance of TrAdaBoost optimization algorithm than C

2. S and N have similar influence on TrAdaBoost optimization algorithm

Methods	Full name	Description	Literature
MD	mahalanobis distance	A direction-sensitive distance classifier that uses statistics for each class, which assumes all class covariances are equal.	Richards, 1999[49]
PLSR	partial least square regression	A statistical method that finds a linear regression model by projecting the predicted variables and the observable variables to a new space.	Herman, 1985[50]
FLDA	Fisher's linear discriminant analysis	A method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects.	Mclachlan, 2004[51]
LR	Logistic regression	A statistical method that is used to describe the relationship between a dependent variable and multiple independent variables. It has the advantage of being less affected by some non-normality of variables.	David, 2010[52]
SVM	Support Vector Machine	A supervised learning model that divide the examples of the separate categories by a clear gap that is as wide as possible	Hearst, 1998[53]



The performance of five commonly used algorithms with different training datasets

Overall verification results of five commonly used algorithms and TrAdaBoost optimization algorithm

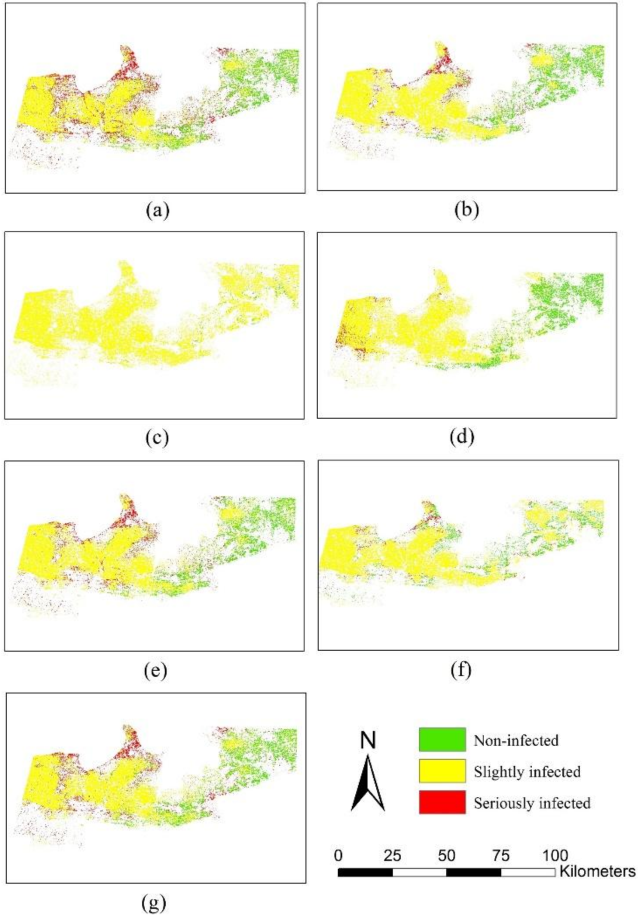
		Reference				User's accuracy(%)	Overall accuracy(%)	Kappa
		Normal	Slight	Serious	Sum			
FLDA	Normal	9	5	0	14	64.29	74.36	0.61
	Slight	2	11	0	13	84.62		
	Serious	0	3	9	12	75.00		
	Sum	11	19	9	39			
	Producer's accuracy (%)	81.82	57.89	100.00				
LR	Normal	8	3	0	11	72.73	66.67	0.48
	Slight	3	12	3	18	66.67		
	Serious	0	4	6	10	60.00		
	Sum	11	19	9	39			
	Producer's accuracy (%)	72.73	63.16	66.67				
MD	Normal	1	1	0	2	50.00	48.72	0.02
	Slight	10	18	9	37	48.65		
	Serious	0	0	0	0	0.00		
	Sum	11	19	9	39			
	Producer's accuracy (%)	9.09	94.74	0.00				
PLSR	Normal	7	3	0	10	70.00	58.97	0.31
	Slight	4	14	7	25	56.00		
	Serious	0	2	2	4	50.00		
	Sum	11	19	9	39			
	Producer's accuracy (%)	63.64	73.68	22.22				
SVM	Normal	9	2	0	11	81.82	74.36	0.59
	Slight	2	14	3	19	73.68		
	Serious	0	3	6	9	66.67		
	Sum	11	19	9	39			
	Producer's accuracy (%)	81.82	73.68	66.67				
TrAdaBoost optimization algorithm	Normal	10	3	1	14	71.43	82.05	0.72
	Slight	1	14	0	15	93.33		
	Serious	0	2	8	10	80.00		
	Sum	11	19	9	39			
	Producer's accuracy (%)	90.91	73.68	88.89				

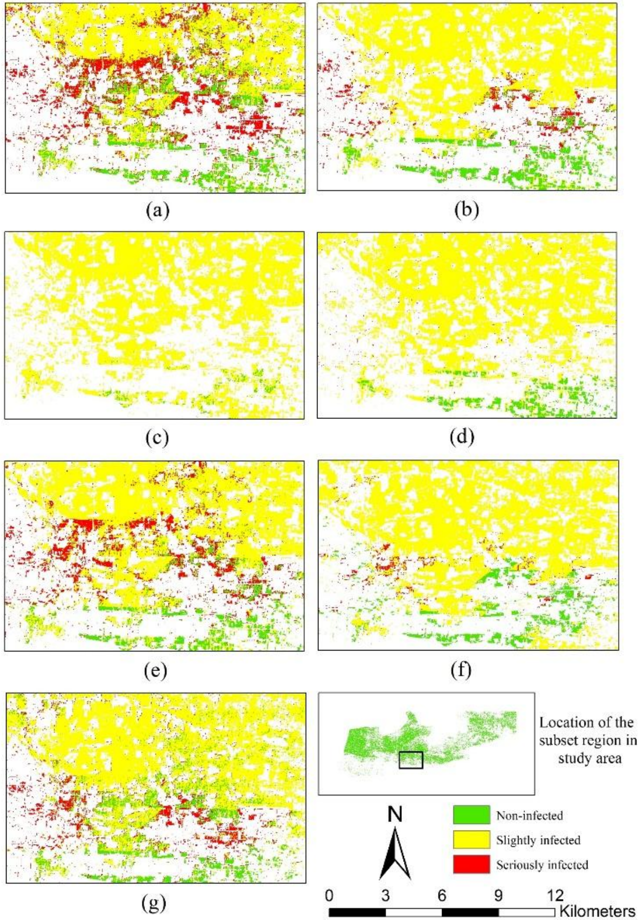
TrAdaBoost optimization algorithm performed the best among all algorithms with an overall accuracy of 82.05% and kappa coefficient of 0.72

Overall verification results of TrAdaBoost and TrAdaBoost optimization algorithm

		Reference				User's accuracy(%)	Overall accuracy(%)	Kappa
		Normal	Slight	Serious	Sum			
TrAdaBoost	Normal	10	5	2	17	58.82	74.36	0.61
	Slight	1	12	0	13	92.31		
	Serious	0	2	7	9	77.78		
	Sum	11	19	9	39			
	Producer's accuracy(%)	90.91	63.16	77.78				
TrAdaBoost optimization algorithm	Normal	10	3	1	14	71.43	82.05	0.72
	Slight	1	14	0	15	93.33		
	Serious	0	2	8	10	80.00		
	Sum	11	19	9	39			
	Producer's accuracy(%)	90.91	73.68	88.89				

The results indicates our new algorithm considering the representativeness and effectiveness of auxiliary samples could enhance the classification accuracy of the learner and provide high disease monitoring accuracy with limited sample data





Infection map of powdery mildew produced by LFDA (a), LR (b), MD (c), PLSR (d), SVM (e), TrAdaBoost (f) and TrAdaBoost optimization algorithm (g)

Thanks!