Research Article

https://doi.org/10.1631/jzus.B2600366

Integrating pooled RT-qPCR and machine learning enables rapid field estimation of rice ragged stunt virus infection in rice

Jingjing LI[1,2*], Meng JIANG[1,2*], Wenzhuo ZENG[2], Xiaoyan ZHANG[3], Jianxiang WU[2,4✉], Qingyao SHU[1,2✉], Long WANG[1,2✉]

[1]*State Key Laboratory of Rice Biology & Breeding, and Zhejiang Provincial Key Laboratory of Crop Germplasm Innovation and Utilization, The Advanced Seed Institute, Zhejiang University, Hangzhou 310058, China*

[2]*Hainan Institute of Zhejiang University 572025, Sanya, China*

[3]*College of Modern Agriculture, Linyi Vocational University of Science and Technology, Linyi 276025, China*

[4]*State Key Laboratory of Rice Biology & Breeding, Institute of Biotechnology, Zhejiang University, Hangzhou 310058, China*

Abstract: Rice ragged stunt virus (RRSV) significantly impacts rice production across Asia and other major rice-growing regions. The virus infects rice during the vulnerable seedling stage, resulting in severe yield losses. Developing reliable assessment models for RRSV infection is therefore essential for early management and offers a cost-effective strategy for disease control. In this study, we constructed a machine learning model using linear regression, decision tree, and random forest algorithms based on reverse transcription quantitative polymerase chain reaction (RT-qPCR) data from pooled seedling samples of 10, 20, 30, and 50 individuals, respectively. Model performance was comprehensively evaluated based on coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), symmetric mean absolute percentage error (SMAPE), and mean absolute percentage error (MAPE), alongside comparisons of predicted and actual infection rates determined by PCR and RT-qPCR, as well as disease incidence observed in field validation assays. The model derived from 30-seedling bulks achieved the highest predictive accuracy, with estimates closely matching actual infection rates and

✉ Long WANG, longwangob@zju.edu.cn
✉ Qingyao SHU, qyshu@zju.edu.cn
✉ Jianxiang WU, wujx@zju.edu.cn
* These authors contributed equally to this work
Long WANG, https://orcid.org/0000-0002-3968-6658
Qingyao SHU, https://orcid.org/0000-0002-9201-0593
Jianxiang WU, https://orcid.org/0000-0002-7611-7833

Received May 28, 2026; Revision accepted Aug. 19, 2026;
Crosschecked xxx. xx, 20xx

observed field outcomes. A single RT-qPCR assay of 30 pooled seedlings thus provides a reliable and efficient means for RRSV infection assessment and disease incidence prediction under field conditions. This approach supports decision-making for seedling transplantation and enables monitoring of disease dynamics across developmental stages, ultimately facilitating timely interventions to mitigate yield losses and economic burdens on farmers.

Key words: Rice ragged stunt virus (RRSV); Assessment model; Machine learning; Disease incidence prediction

1 Introduction

Disease induced by the rice ragged stunt virus (RRSV) is one of the challenges faced by the rice industry in Asia and other rice-growing areas of the world. RRSV belongs to the genus *Oryzavirus* and the family *Spinareoviridae*. RRSV is transmitted by *Nilaparvata lugens* (the brown planthopper, BPH) in a persistent and proliferative manner (Ling et al., 1978; Boccardo, 1984). Its genome comprises ten double-stranded RNAs of 1.2 to 3.9 kb long (Omura et al., 1983). These segments, designated S1 to S10, have been fully sequenced. Segments S6, S7, and S10 encode non-structural proteins, while the remaining segments encode structural proteins, which are either integral components or predicted to assist in RRSV virion assembly (Wu et al., 2010; Jia et al., 2012) RRSV has been reported to induce various disease symptoms including plant stunting, excessive tillering, and dark green leaves with serrated edges and/or twisted tips (Fig. S1) (Hibino, 1996; Zhang et al., 2016), leading to poor grain quality and substantial yield losses. Importantly, these visible symptoms typically appear days to weeks after infection, which delays field recognition and intervention.

Since the first report of RRSV in the Philippines in the early 1970s, RRSV has now appeared in Thailand, India, China, and other countries, posing a severe threat to food security (Shikata et al., 1979; Wang et al., 2022). Significant outbreaks of RRSV in China and Vietnam between 2010-2012 led to severe yield and economic losses (Wang et al., 2014). In Hainan Province, China, the reported infection rate of dwarf rice plants was 34.85% in 2012, which increased to 46.6% during the period from 2013 to 2015 (Liu et al., 2014; Yang et al., 2017). Due to the emergence of insecticide-resistant BPHs and the lack of RRSV-resistant rice cultivars, this virus is continuing to exacerbate risks to rice production in many countries (Listihani et al., 2023).

Rice is a staple food for over half of the global population and a primary income source for farmers in many countries. Infections of rice seedlings with RRSV can lead to significant agricultural losses, often resulting in no harvestable crop. These losses are magnified by a latency period ranging from 9 to 30 d during which viral symptoms are not detectable (Ling, et al., 1978). Once RRSV is identified in the field, it cannot be eradicated, resulting in delayed management actions and substantial economic losses. Therefore, practical strategies that reduce sampling burden while retaining molecular accuracy are needed for timely estimation of field infection levels and to enable timely interventions to mitigate impacts on rice production. Currently, virus infection

assessment relies on labor-intensive methods, including field surveys based on visible disease symptoms and molecular techniques such as reverse transcription-polymerase chain reaction (RT-PCR), RT-quantitative PCR (RT-qPCR), RT-loop-mediated isothermal amplification (RT-LAMP), and dot-blot enzyme-linked immunosorbent assays (dot-blot ELISA) (Baehaki et al., 2017; Wang et al., 2017; Yang, et al., 2017; Lai et al., 2018). However, symptom-based surveys are limited by delays in symptom development (Ling, et al., 1978), while molecular techniques require complex laboratory analyses and extensive sample collection, making them unsuitable for routine monitoring and breeding programs.

In recent years, the development of mathematical assessment models for virus diseases using big data and machine learning has attracted considerable interest among the research community. For example, Ginsberg et al. (2009) reported an assessment model for influenza-like illnesses using factors correlated with disease incidence, and established a linear regression model in which accuracy and reliability were assessed through rigorous cross-validation. Compartmental, logistic, and Gaussian models have been applied to analyze the global spread patterns of COVID-19 (Ibrahim and Al-Najafi, 2020; Wibowo and Wihayati, 2021; Ojokoh et al., 2022). The advancement of machine learning has also enabled the development of robust assessment models for plant diseases (Singh et al., 2016; Munawar et al., 2024), addressing the limitations of early models such as linear and logistic regression, which were constrained by simplified assumptions and small datasets. For instance, Avelino et al. (2004) used traditional models to predict the spread of coffee leaf rust, but their accuracy was limited by the scope of available data. Modern machine learning models that integrate large datasets, like image databases and long-term weather records, have significantly improved predictive accuracy for disease incidence and severity. Owomugisha and Mwebaze (2016) and Benos et al. (2021) used machine learning techniques to enhance field disease forecasts, while Skelsey et al. (2018) used classification models to assess the dispersal of the potato late blight pathogen, improving the precision of disease risk predictions. More recently, De Cól et al. (2024) combined machine learning with a 24-year dataset of daily weather variables to successfully forecast wheat head blast outbreaks. In addition, Zhang et al. (2026) developed lightweight convolutional neural network models based on field color images to detect Fusarium head blight and quantitatively evaluate disease incidence and severity in wheat. Therefore, machine learning technologies have provided plant researchers with robust tools to predict disease outbreaks and progressions. However, these models frequently encounter challenges associated with nonlinear interactions in biological systems and complex transmission dynamics. To improve the accuracy of disease spread prediction, it is necessary to foster multidisciplinary collaborations to integrate molecular methods with machine learning techniques.

Given the substantial damage caused by RRSV, there is a clear need for practical tools that can estimate infection in the field before symptoms become widespread. Current approaches are often limited by delayed symptom observation or labor-intensive molecular testing, restricting their applicability for large-scale field monitoring. A major unresolved challenge in plant disease research is how to translate molecular-level measurements into quantitative population-level infection estimates under field conditions. In this study, we developed a framework integrating pooled RT-

qPCR detection with machine learning models to estimate RRSV infection rates in rice seedling populations. By linking pooled molecular measurements with data-driven modeling, this approach improves the efficiency and scalability of infection assessment at early growth stages (Fig. 1).

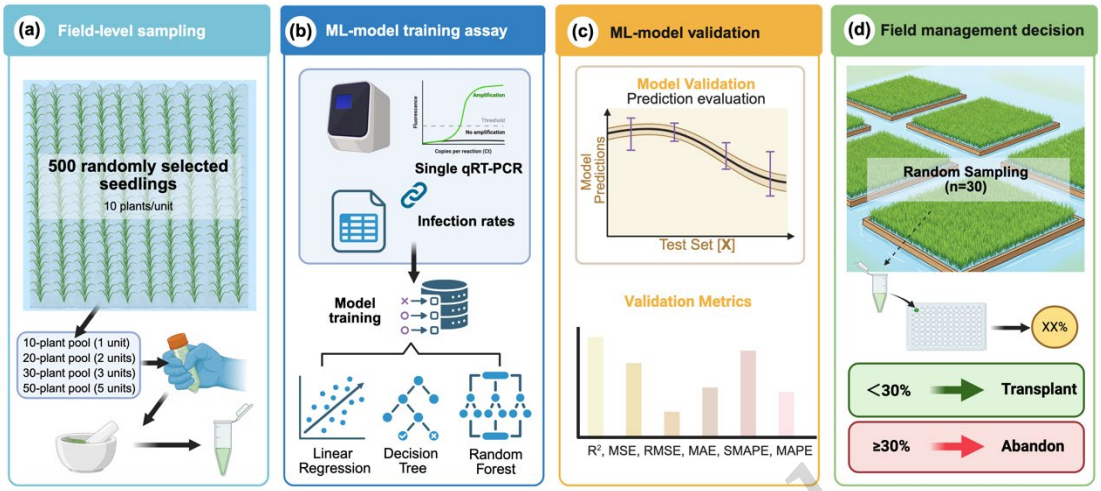

Fig. 1 Integrated framework for developing RRSV infection rate assessment models enabling precision decision-making in rice production. (a) Field-level sampling: strategy for collecting rice seedling units (10, 20, 30, and 50 plants per pool) from seedling beds for pooled analysis. (b) Machine learning (ML)-model training assay: development of the predictive pipeline by correlating RRSV S10 gene relative expression via single RT-qPCR with known infection rates across different pooling scales using linear regression, decision tree, and random forest algorithms. (c) ML-model validation: systematic evaluation of model performance on independent test sets, field datasets and statistical metrics, including R^2, MSE, RMSE, MAE, SMAPE, and MAPE. (d) Field management decision: a practical application workflow where a random leaf sampling of 30 seedlings provides a real-time estimate of infection rate to support transplanting decisions (transplant if <30%; abandon if ≥30%).

2 Materials and methods

2.1 Plants and insects

Seedlings of *Oryza sativa* L. cv. Jiaxi B (JXB) were used in model construction, validation and verification in this study. Seeds were soaked in water for 1 day and germinated at 32 °C for 2 days. The germinated seeds were then sown into seedling beds in the experimental farm of Hainan Institute of Zhejiang University, China, where there was a high incidence of BPHs and RRSV. After the seedlings had reached the four-leaf stage, 500 seedlings were randomly sampled. To confirm the RRSV infection of each seedling, total RNA was extracted individually for RT-PCR and RT-qPCR testing. To construct samples of different pool size, 500 seedlings were first randomly divided into 50 units, with 10 seedlings per unit. These 50 10-plant units were then randomly combined to generate 20-, 30- and 50-plant pools, by collecting the same amount of leaf tissue from each seedling (the second leaf), i.e., the 20-plant pools were randomly formed by mixing two units (e.g., units 1 + 2, 3 + 4, 1 + 3, 2 + 4, ⋯, 49 + 50), the 30-plant pools were randomly formed by mixing three units (e.g., units 1 + 2 + 3, 4 + 5 + 6, 1 + 2 + 4, 1 + 2 + 5, ⋯, 48 + 49 + 50), and the 50-plant pools were formed

by mixing any five units (e.g., unit 1 + 2 + 3 + 4 + 5, 6 + 7 + 8 + 9 + 10, 1 + 2 + 3 + 4 + 6, 5 + 7 + 8 + 9 + 10, ⋯, 46 + 47 + 48 + 49 + 50). The pooled samples were then homogenized in liquid nitrogen for subsequent RNA extraction and RT-qPCR testing.

To survey the carrying rate of RRSV in the insect vector, 500 adult BPHs were collected from 100 rice plants (JXB, tillering stage) in the experimental farm of Hainan Institute of Zhejiang University (spring, 2023). Of these, 180 individuals were randomly selected for subsequent RNA extraction and RT-PCR and RT-qPCR testing.

For validation and verification of the effectiveness of established models, rice samples were collected from different production sites within and outside Hainan province.

2.2 RT-PCR and RT-qPCR

For total RNA extraction, BPHs (individual) and rice leaf tissues (both individual and mixed-pools) were homogenized in liquid nitrogen. The extraction was performed using the Trizol reagent (Thermo Fisher Scientific Inc., Shanghai, China) following the manufacturer's protocol. The extracted total RNA (1 μg per 20-μL reaction) was reverse transcribed using ToloScript All-in-one RT EasyMix for qPCR, as instructed (Tolo Biotech., Shanghai, China).

For the detection of RRSV infection, the samples were assayed as described previously (Wang, et al., 2017). Initial RRSV testing was conducted using both RT-PCR and RT-qPCR on individual plants. The 10-, 20-, 30-, and 50-plant pooled samples were subjected to RT-qPCR analysis. Briefly, RT-PCR was conducted using the 2X Rapid Taq Master Mix (Vazyme, Nanjing, China), as described (Yang, et al., 2017), with a set of primers specific to the RRSV S8 gene (Table S1). The reaction conditions were 95 °C for 3 min; 30 cycles of 95 °C for 15 s, 52 °C for 15 s, and 72 °C for 15 s; 72 °C for 5 min. For each RT-qPCR, the reaction system consisted of 1 μL of cDNA (100 ng cDNA per μL), 0.5 μL of each primer (10 pmol each) specific RRSV S10 gene (Table S1) (Wang, et al., 2017), 3 μL of ddH_2O, and 5 μL of 2×Q3 SYBR qPCR Master Mix (Tool Biotech., Shanghai, China). The RT-qPCR conditions were 95 °C for 30 s and 40 cycles of 95 °C for 10 s and 60 °C for 30 s. To ensure the formation of single peaks and to prevent the formation of primer dimers and nonspecific products, a melt curve (15 s at 95 °C, 60 s at 60 °C, and 15 s at 95 °C) was generated at the end of each RT-qPCR reaction. The relative expression level of each individual sample was calculated using the $2^{-\triangle Ct}$ method (Livak and Schmittgen, 2001) followed by normalization against the expression level of *OsActin* in assayed rice plants, as described (Wang et al., 2023). The RRSV positive plants or BPHs were referred to as RRSV-infected, while those that tested negative were referred to as RRSV-uninfected.

2.3 Model construction

To establish a model suitable for estimating the occurrence of RRSV in rice fields using viral data of seedling samples, three assessment models—linear regression, decision tree, and random forest—previously used in plant disease recognition and prediction (Hatuwal et al., 2020; Patra et al., 2023) were tested. Six key performance metrics, the coefficient of determination (R^2), symmetric mean absolute percentage error (SMAPE), mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE), as reported (Li, 2017; Chicco et al., 2021; Fenu and Malloci, 2021), were used to assess the accuracy

and precision of the linear regression, decision tree, and random forest models. The R^2 quantifies the extent to which the variation in the dependent variable is explained by the independent variables in a regression model. The computation formula is $R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_{i-\bar{y}_i})^2}$, where y_i denotes the actual RRSV infection rate, $\hat{y}_i$ denotes the prediction RRSV infection rate, $\bar{y}_i$ denotes the mean of the actual RRSV infection rate, and n is the number of samples used in the test set. The SMAPE is an accuracy measurement based on the percentage error. Its formula is $SMAPE=\frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{(|y_i|+|\hat{y}_i|)/2}$. The MSE measures the average of the squares of errors, an average squared difference between the estimated value and the actual value. An analysis using $MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ provides an estimation of how well a regression model can predict the outcome with its predicted average distance squared from the actual data points (Plonsky and Ghanbar, 2018). The RMSE measures the variations from the actual values and the average magnitude of the errors, calculated using $RMSE=\sqrt{MSE}$. The MAE quantifies the average magnitude of errors from a set of predictions using $MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i|$, without considering their directions. The MAE result is the average of the absolute difference between the predicted and observed value, and the individual differences have equal weight. The MAPE is calculated using $MAPE=\frac{100\%}{n}\sum_{i=1}^{n}|\frac{y_i-\hat{y}_i}{y_i}|$ and describes the accuracy as a percentage of error (Bruce et al., 2020; Guo et al., 2024).

Inputs for establishing assessment models were derived from two primary data sources: a) the viral load data, measured through RT-qPCR for the relative expression levels of RRSV S10 gene and used to represent a quantifiable metric of viral transcripts in the assayed rice tissues; and b) the RRSV infection rate, calculated according to the result of RT-PCR. Thirty-five of these 50 composite datasets across the 10-, 20-, 30-, and 50-plant pools were randomly selected for training of the assessment models.

2.4 Model validation

For the linear regression model, the coefficients were adjusted to minimize the residuals between the observed and the predicted infection rates, and to enhance the model accuracy. In the decision tree model, the thresholds of class splits were refined according to the expression level of the RRSV S10 gene to effectively categorize the infection rates. For the random forest model, parameters (i.e., the number and depth of trees) were fine-tuned to optimize the bias-variance trade-off and ensure a robust model performance (Haq et al., 2022; Lee and Yun, 2024). The remaining 15 of these 50 composite datasets from various pool sizes were used to validate the accuracy of the model for each bulk size.

2.5 Model verification

To verify whether the prediction results of seedling stage are consistent with the disease incidence of heading-stage plants, rice seedlings were sampled from the same seedling beds. One hundred and fifty seedlings were randomly collected for five samples, 30 seedlings for each sample. The cycle threshold (Ct) values of the RRSV S10 gene of these five samples were tested by RT-qPCR, then used as input to generate the RRSV infection rates predicted by each of the three models. At the heading-stage, the RRSV disease symptoms of five samples (48 plants per sample) were surveyed to

calculate the RRSV disease incidence. The predicted values were compared with the actual incidences to verify whether the model could accurately predict the disease incidence rate at the heading-stage based on the seedling-stage RT-qPCR test.

To test the effectiveness of the predictive models in a broad range of field environments and growth stages, samples were also collected from five rice fields, i.e., Hongqi Village, 18.35°N, 109.20°E, winter 2023, six samples, tillering stage; Shengli Village, 18.37°N, 109.20°E, winter 2023, 12 samples, grain filling stage; Baihe Village, 18.38°N, 109.20°E, spring 2024, three samples, heading stage; Baichao Village, 18.36°N, 109.20°E, spring 2024, three samples, heading stage; and Gongbei Village, 18.38°N, 109.15°E, spring 2024, three samples, heading stage, in Sanya, Hainan Province, one each in Wuhan, 30.47°N, 114.35°E, Hubei Province, autumn 2024, maturity stage, and Guangzhou 23.16°N, 113.30°E, Guangdong Province, autumn 2024, maturity stage. Each sample consisted of 30 randomly selected plants. In total, 870 rice plants were collected. The same weight of flag-leaf tissue was harvested from each plant. The RRSV infection rate of each sample was detected by individually testing for the S8 gene using RT-PCR. The Ct value of each sample, detected by RT-qPCR, was used as input to generate the predicted infection rates under the three models. The predicted rates were then compared with the actual tested infection rates or field disease incidence to evaluate model accuracy.

2.6 Statistical analysis

The actual and predicted infection rates from five fields used for model evaluation were obtained from three or more independent testing groups. Statistical analyses were conducted using one-way analysis of variance (ANOVA) with Prism software (GraphPad Software, v8.1.2, San Diego, CA, USA). A p-value < 0.05 was regarded as a statistically significant difference.

3 Results

3.1 RRSV-carrying rate of BPHs in the paddy field

To confirm that the experimental field was suitable for constructing RRSV assessment models in rice seedlings, adult BPHs were collected from the experimental farm of Hainan Institute of Zhejiang University. Each rice plant in the field was naturally infested with about 20-100 BPHs (Fig. S2a). Total RNA was extracted from individual field-collected BPHs and the RRSV-carrying rate was determined using RT-PCR, with RT-qPCR used for confirmation. The results showed that 74 of the 180 (~41.11%) BPHs tested were infected with RRSV (Fig. S2b), indicating a high prevalence of viruliferous vectors in this field.

3.2 Analysis of RRSV infection in rice seedlings

To construct RRSV infection-rate assessment models, germinated rice seeds were sown in the seedbed at Hongqi Village. At the four-leaf stage, 500 seedlings were randomly sampled and individually tested for RRSV infection using RT-PCR and RT-qPCR, revealing a mean seedling infection rate of 29.40% (Fig. 2). Subsequently, the seedlings were pooled into mixed samples consisting of 10, 20, 30, or 50 plants per pool. Fifty biological replicates from each pooled sample were subjected to RT-qPCR testing to measure the relative expression levels of the RRSV S10 gene. The pooled RT-

qPCR measurements were then combined with the RT-PCR–derived seedling infection rates (ground truth) to develop and evaluate models that estimate seedling infection rate at early growth stages.

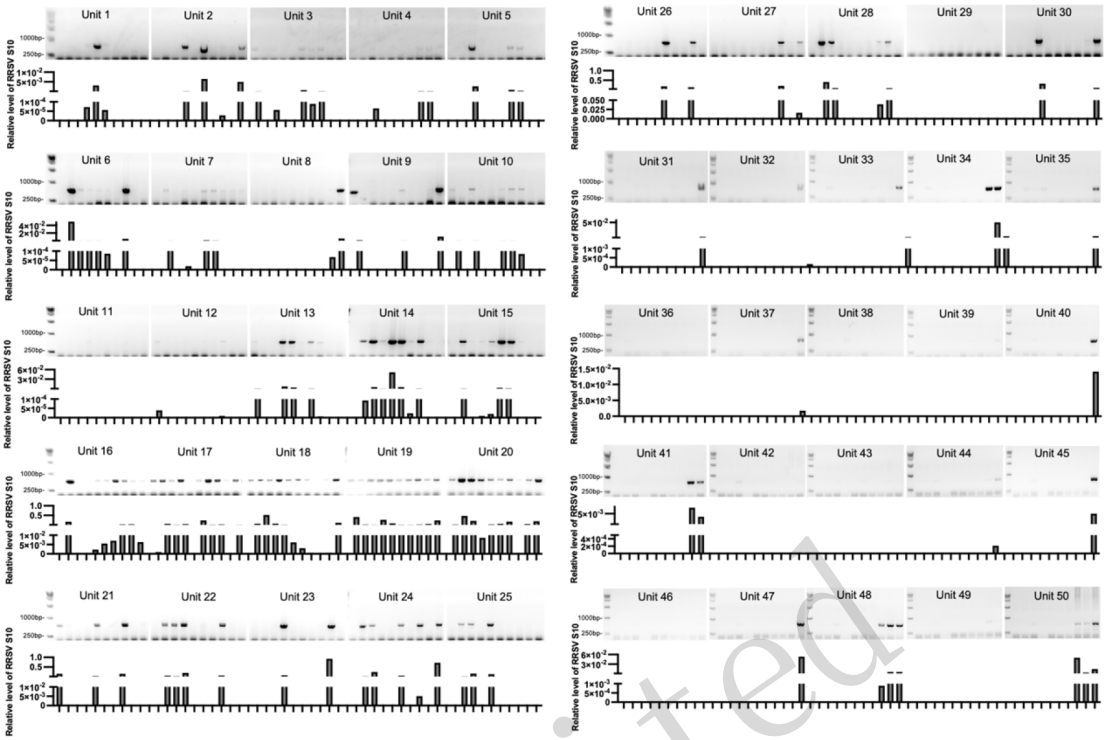

Fig. 2 Detection results of RRSV in 500 rice seedlings using RT-PCR and RT-qPCR. The upper panel shows the RT-PCR results, demonstrating the qualitative detection of RRSV. The lower panel displays RT-qPCR results, providing quantitative analysis of viral load within these rice seedlings.

3.3 Development and evaluation of assessment models using different pool sizes

A total of 35 pooled datasets were used to train the assessment models and the other 15 were reserved for testing model accuracy. Each assessment model was evaluated using multiple statistical parameters to determine the reliability of its infection rate predictions across different sample group sizes.

For the 10-plant pools**,** all three models showed poor predictive performance. The linear regression model provided a poor predictive ability with an equation $y = 172.82x+23.51$ and an $R^2 = -0.144$. The decision tree model also performed poorly, with a negative $R^2 = -0.643$, while the random forest model achieved an $R^2 = -0.061$. The MSE values ranged from 297.11 (random forest) to 460.00 (decision tree), and the RMSE values from 17.23 (random forest) to 21.77 (decision tree). The decision tree model had the highest MAE of 16.67. The SMAPE values ranged from 40.32% to 64.89%, while the MAPE values varied from 36.71% to 56.15%. (Figs. 3a, 4a, Table 1).

For the 20-plant pools, the linear regression model yielded $y = 90.60x +23.15$ with an $R^2 = 0.181$, indicating limited predictive capability. The R^2 obtained using the decision tree model and the random forest model were 0.353 and 0.459, respectively. The MSE values ranged from 99.01 (random forest) to 149.85 (linear regression). The RMSE values were 9.95 (random forest) to 12.24 (linear regression), respectively. The

MAE values ranged from 7.67 (decision tree) to 9.90 (linear regression) The highest SMAPE and MAPE values were 48.49% and 97.36%, respectively, of the linear regression model. The SMAPE and MAPE values were 31.39% and 39.99%, respectively, for the decision tree model, and 40.99% and 51.62%, respectively, for the random forest model (Fig. 3b, Table 1), indicating that 20-plant pools did not provide stable accuracy for infection-rate estimation under these conditions.

With the 30-plant pools, model performance improved substantially. The linear regression model yielded an equation of $y = 102.02x + 17.39$ with an $R^2 = 0.814$. The decision tree and random forest models achieved $R^2 = 0.847$ and 0.862, respectively. The MSE values of the 30-plant pools were 96.25 (linear regression), 79.27 (decision tree) and 71.57 (random forests), the lowest of the four sample groups. The RMSE values ranged from 8.45 (random forest) to 9.81 (linear regression), and the MAE values ranged from 5.56 (decision tree) to 8.06 (linear regression). The SMAPE and MAPE values for all three models in the 30-plant pool were also the lowest among the four sample groups, indicating that the 30-plant sample pooling provided the most accurate and stable infection-rate estimates (Fig. 3c, Table 1).

With the 50-plant pools, the linear regression model showed an equation of $y = 49.68x+18.29$ and an $R^2 = 0.476$. The decision tree and random forest models yielded $R^2 = 0.406$ and 0.528, respectively. The MSE values for the three models ranged from 228.05 to 286.93. The RMSE values ranged from 15.10 (random forest) to 16.94 (decision tree), and the MAE values from 10.15 (random forest) to 11.20 (decision tree). The SMAPE and MAPE values for the linear regression model were 35.66% and 49.98%, 31.30% and 30.10% for the decision tree model, and 29.39% and 28.65% for the random forest model (Fig. 3d, Table 1). These results indicated that pool size strongly influences prediction performance, and 30-plant pools provided the best balance of accuracy and stability.

Across the four sampling strategies (10-, 20-, 30-, and 50-plant pools), clear differences were observed in model behavior and prediction stability. In the 10- and 20-plant pools, all three models showed substantial variability between predicted and actual values, indicating high sensitivity to stochastic variation in viral load (Figs. 4a, b). In contrast, the 30-plant pool markedly improved prediction consistency, while the 50-plant pool provided only marginal additional gains despite increased sampling effort (Figs. 4c, d). Within each sampling group, the three models exhibited distinct prediction patterns. The linear regression model generally captured overall trends but tended to extrapolate beyond biologically meaningful boundaries under high infection conditions. The decision tree model effectively captured threshold-like responses but produced stepwise and discontinuous predictions across heterogeneous samples. In contrast, the random forest model consistently generated smoother, biologically constrained, and more stable predictions across all infection levels (Figs. 4a-4d).

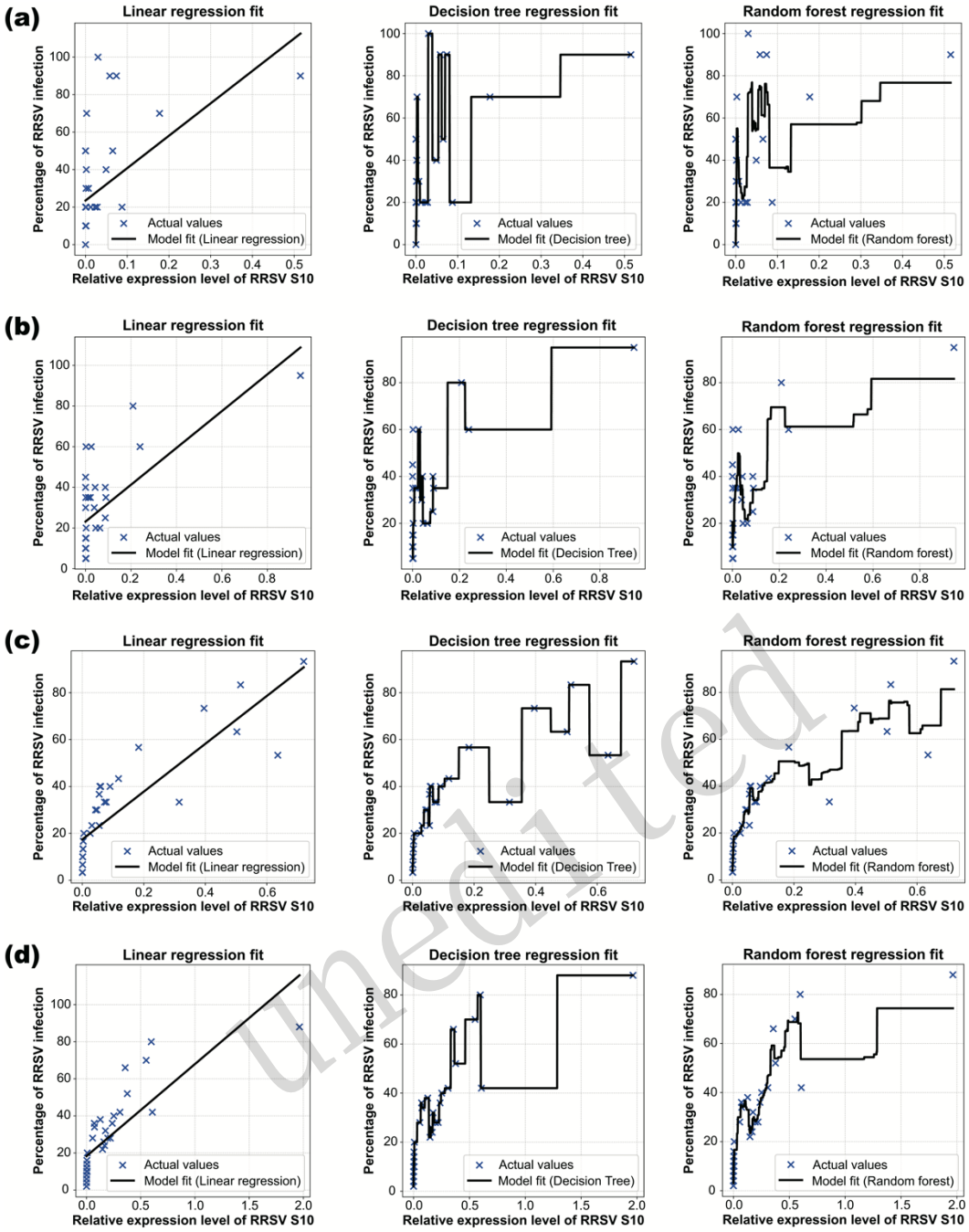

Fig. 3 Established linear regression, decision tree, and random forest models for predicting RRSV infection rate based on the relative expression level of the RRSV S10 gene in pooled rice seedling samples of different sizes: (a) 10-plant, (b) 20-plant, (c) 30-plant, and (d) 50-plant pooled samples.

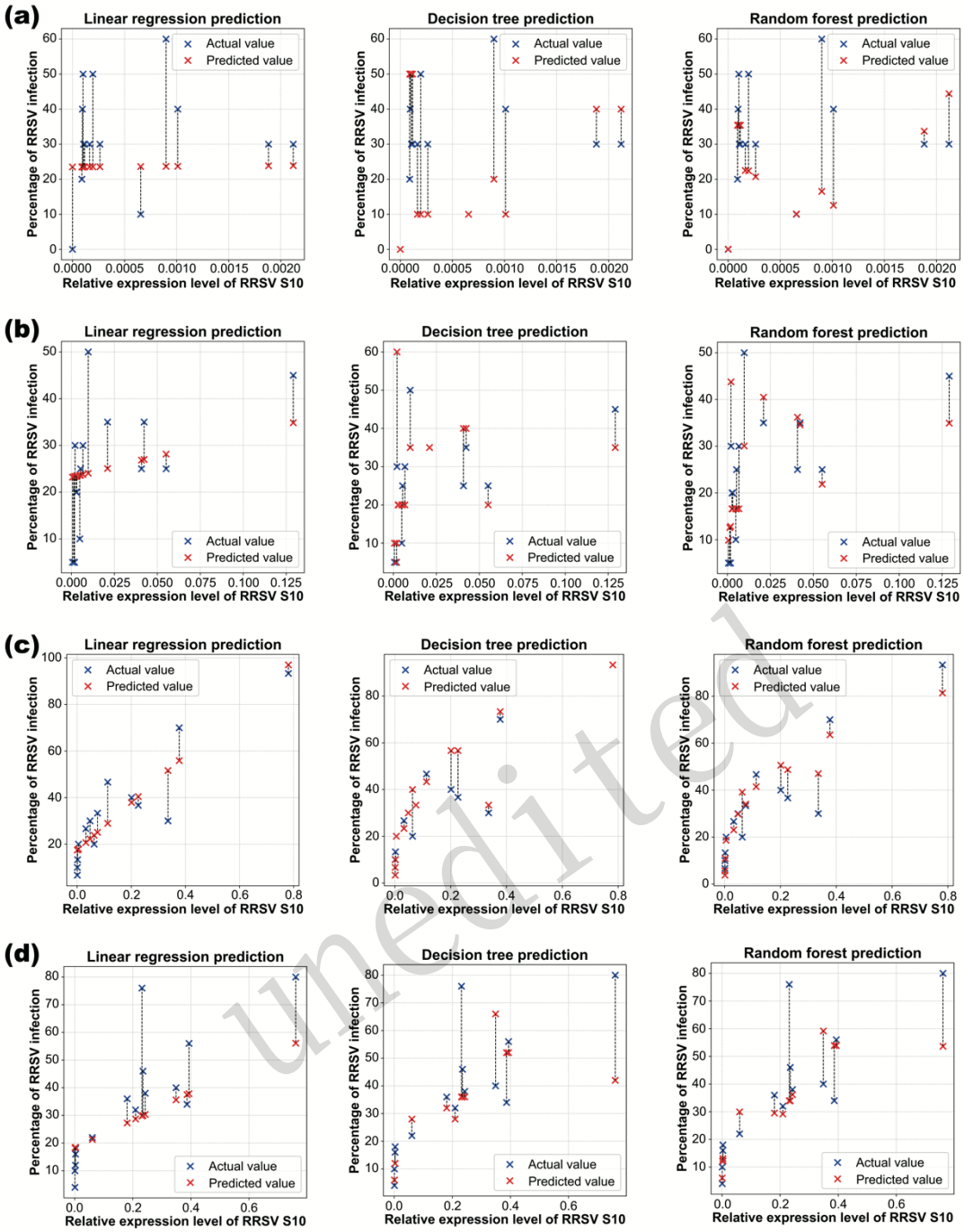

Fig. 4 Prediction performance of regression models for RRSV Infection. These figures show the effectiveness of linear regression, decision tree, and random forest models in predicting RRSV infection rates based on the relative expression levels of the RRSV S10 gene across various rice seedling pool sizes: (a) 10-plant, (b) 20-plant, (c) 30-plant, and (d) 50-plant pooled samples.

Table 1. Analysis results of the model performances

Model type	Pool size	Equation	R^2	MSE	RMSE	MAE	SMAPE (%)	MAPE (%)
Linear regression	10-plant	$y = 172.82x + 23.51$	-0.144	320.32	17.90	14.96	64.89	40.72
	20-plant	$y = 90.60x + 23.15$	0.181	149.85	12.24	9.90	48.49	97.36
	30-plant	$y = 102.02x + 17.39$	0.814	96.25	9.81	8.06	31.92	39.72
	50-plant	$y = 49.68x + 18.29$	0.476	252.99	15.62	10.97	35.66	49.98
Decision tree	10-plant	-	-0.643	460.00	21.77	16.67	40.32	56.15
	20-plant	-	0.353	118.33	10.88	7.67	31.39	39.99
	30-plant	-	0.847	79.27	8.90	5.56	23.24	23.78
	50-plant	-	0.406	286.93	16.94	11.20	31.30	30.10
Random forest	10-plant	-	-0.061	297.11	17.23	12.23	64.54	36.71
	20-plant	-	0.459	99.01	9.95	8.16	40.99	51.62
	30-plant	-	0.862	71.57	8.45	6.47	26.24	26.48
	50-plant	-	0.528	228.05	15.10	10.15	29.39	28.65

3.4 Validation of model performance using adjacent field seedlings at early growth stages

To evaluate the predictive accuracy of the assessment models, infection rates predicted using the 30-plant pooling strategy were compared against seedling infection rates measured by individual RT-PCR in five sampling sites within the same seedling experimental field of the Hainan Institute of Zhejiang University (Table 2). The results revealed that the linear regression model generated predicted infection levels that ranged from 17.91% to 22.43% across the five sampling sites. The decision tree model produced predictions between 20.00% and 30.00%, and the random forest model produced estimates from 18.68% to 29.25%. Symptom-based disease incidence recorded at heading stage ranged from 20.83% to 27.08%. In addition, sampling site #4 had a measured seedling infection rate of 20.00%, a value close to the total disease incidence of 22.92%, providing a direct reference for comparing model predictions with field observations.

Across the five plots, the decision tree and random forest predictions were numerically close to the investigated field incidences. In sampling site #4, the decision tree prediction of 20.00% was identical to the measured seedling infection rate, and the random forest estimate of 21.27% was also within the range of the observed field values. In other plots, both tree-based models frequently produced values that fell near the corresponding disease incidence investigated at heading stage.

Table 2. Predicted RRSV infection rates of rice seedlings and disease incidence of heading-stage rice plants in the experimental farm of Hainan Institute of Zhejiang University

Sampling	Linear regression predicted RRSV infection rate (%) [a]	Decision tree predicted RRSV infection rate (%) [a]	Random forest predicted RRSV infection rate	No. of RRSV-infected (RRSV disease incidence %) [b]

			(%) [a]	
#1	18.80	20.00	18.68	10 (20.83)
#2	20.97	23.33	24.73	12 (25.00)
#3	22.43	30.00	29.25	13 (27.08)
#4	19.85	20.00	21.27	10 (20.83)
#5	17.91	20.00	18.68	10 (20.83)
Total				22.92

a: Thirty seedlings were randomly collected for each sampling.

b: Fourty eight heading-stage plants were randomly surveyed for each sampling.

3.5 Independent field validation across geographic locations

To further validate the assessment models, predicted infection rates obtained from the 30-plant pooling strategy were compared with actual RRSV infection rates measured across seven independent rice fields, each comprising several areas. For the field in Baihe Village, prediction values from the linear regression model were 28.88% ± 5.08%, which matched well with the actual infection rates 34.44% ± 5.88% ($p = 0.515$, student's *t*-test). Predictions from the decision tree model were 40.00% ± 8.39%, occasionally approaching the actual infection rates ($p = 0.620$), while the random forest model generated estimates of 38.50% ± 6.11%, which were largely consistent with the actual values ($p = 0.658$, Figs. 5a, S3, Table 3).

For the field in Baichao Village, the actual infection rate was consistently high at 96.67%. The decision tree model produced the most accurate predictions, giving 93.33%. The random forest model underestimated the infection level at 81.00%, whereas the linear regression produced values of > 100% (128.80–149.41%), reflecting the fact that linear regression is unbounded and can exceed the biological maximum at high infection levels (Figs. 5b, S3, Table 3). For the field in Gongbei Village, the linear regression model predicted values of 43.17% ± 7.69%, which were close to the actual infection rates 48.89% ± 6.19% ($p = 0.595$). The decision tree model underestimated several areas, with values of 33.33% and 40.00%, while the random forest model produced moderately accurate estimates of 44.22% ± 2.05%, ($p = 0.536$, Figs. 5c, S3, Table 3). For the field in Hongqi Village, the linear regression model predicted values were generally aligned with the actual infection rates (45.00 ± 3.63% and 39.51% ± 3.88%, respectively, $p = 0.326$). The decision tree model produced highly variable predictions of 41.44% ± 4.92% ($p = 0.540$), while the random forest model generated relatively stable estimates close to the higher end of the observed values, ranging from 45.67% to 50.85% (Figs. 5d, S3, Table 3). For the field in Shengli Village, linear regression predicted values ranging from 19.23% to 74.63%, which closely followed the actual infection rates that ranged from 23.33% to 70.00% ($p = 0.808$). The decision tree model occasionally overestimated areas with moderate infection, predicting values such as 73.33% for an actual value of 63.33%. In contrast, the random forest model produced more stable and generally accurate predictions, with values from 20.05% to 76.05% ($p = 0.665$, Figs. 5e, S3, Table 3). For the field in Wuhan, Hubei, all three models aligned closely with the actual infection rate of 20.00%. The linear regression model predicted 18.92%, the decision tree model predicted 20.00%, and the random forest model predicted 19.68% (Figs. 5f, S4, Table 3). For the field in Guangzhou, Guangdong, predictions from the linear regression model, the decision tree model and the random forest model were 17.83%, 16.67% and 17.37% respectively, compared with the measured infection rate of 13.33% (Figs. 5g, S4, Table 3).

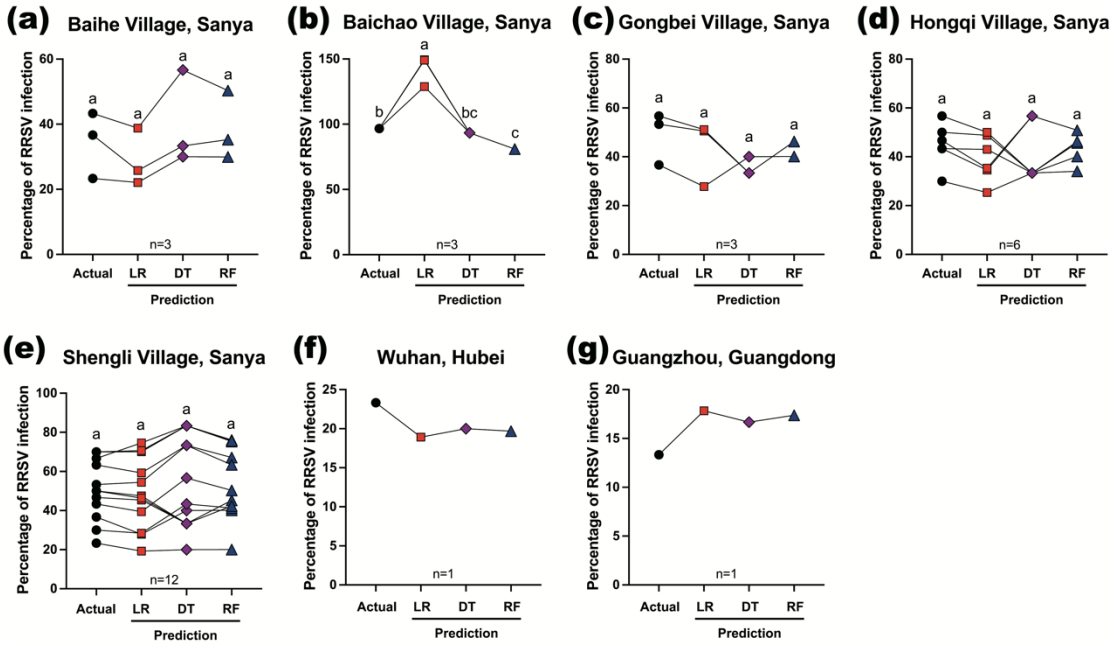

Fig. 5 Field validation of 30-plant pooled sample assessment models for forecasting RRSV infection rate. The performance of the linear regression (LR), decision tree (DT), and random forest (RF) models was evaluated across seven field sites: (a) Baihe Village, (b) Baichao Village, (c) Gongbei Village, (d) Hongqi Village, and (e) Shengli Village in Sanya, Hainan; (f) Wuhan, Hubei; and (g) Guangzhou, Guangdong. Values labeled with the same letter are not significantly different according to one-way ANOVA.

Table 3. RRSV infection rates of rice plants in different locations in Sanya, Wuhan and Guangzhou determined by RT-PCR

Sampling locations	Years	Sampling stage	No. of RRSV-positive/tested plants (Infection rate %)
Baihe Village, Sanya (18.38°N,109.20°E)	2024	Tillering	31/90 (34.44)
Baichao Village, Sanya (18.36°N,109.20°E)	2024	Grain filling	87/90 (96.67)
Gongbei Village, Sanya (18.38°N,109.15°E)	2024	Heading	44/90 (48.89)
Hongqi Village, Sanya (18.35°N,109.20°E)	2023	Heading	80/180 (44.44)
Shengli Village, Sanya (18.37°N,109.20°E)	2023	Heading	181/360 (50.28)
Wuhan, Hubei (30.47°N,114.35°E)	2024	Maturity	6/30 (20.00)
Guangzhou, Guangdong (23.16°N,113.30°E)	2024	Maturity	4/30 (13.33)

4 Discussion

Previous surveys of virus diseases in rice fields were done via large-scale sample collections, extraction of RNA from individual plants, and RT-PCR/RT-qPCR, which are all labor-intensive and time-consuming (Liu, et al., 2014; Yang, et al., 2017). In this study, we integrated machine learning with RT-qPCR data derived from pooled seedling samples to establish an efficient and scalable framework for estimating the RRSV seedling infection rate. Importantly, the model target was the molecular infection rate (ground-truthed by individual RT-PCR), whereas heading-stage symptom incidence was recorded only as a downstream field observation for contextual interpretation rather than a prediction endpoint. The field used for model development exhibited a high density of BPHs and significant virus transmission pressure, providing suitable conditions for robust model training and validation. The predictive framework was validated across multiple independent fields in Sanya, Wuhan, and Guangzhou, demonstrating that a pooled-assay strategy can support consistent infection-rate estimation across locations and growth stages. RT-qPCR is widely used to detect human viruses, including COVID-19, but its application has not been integrated with machine learning models to predict virus infection rates (Yelin et al., 2020). However, linking a single pooled RT-qPCR data to an infection-rate estimate in crop fields remains non-trivial, because a pooled signal reflects both the proportion of infected plants and the distribution of viral load among infected individuals. Although rice developmental stage may affect host physiology, RRSV accumulation is influenced more strongly by the infection duration and plant condition. The techniques described in this study offer novel perspectives and allow a better understanding of RRSV dynamics in different rice fields and at different rice growth stages. We consider that these novel techniques not only provide insights beyond the broader epidemiological patterns reported in previous studies, but also improve the accuracy of RRSV infection rate and disease incidence. The predictions generated with the three assessment models described above can help farmers to decide the transplantation schedule of seedlings to manage RRSV transmission. The results presented here also show the possibility of application of the random forest model to handle the complexities and heterogeneities of agricultural data. In addition, incorporation of three models can reduce variance and provide more reliable and robust predictions compared to the traditional methods using a single decision tree or a linear model reported in earlier studies.

To rigorously test the three assessment models for different mixed plant samples, we evaluated the individual characteristics of each dataset, effectiveness of model fit, and performance of statistical metrics. For the 10- and 20-plant pools, both the linear regression and tree-based models exhibited poor performance, likely due to significant variation in viral loads, data non-linearity and heterogeneous infection rates, which is consistent with recent findings by Bock et al. (2022), regarding the limitations of small sample sizes in plant disease estimation. For the 50-plant pool, all three models demonstrated slight performance improvements compared to the 10-plant and 20-plant pooling datasets, but remained suboptimal, likely reflecting a trade-off between the data aggregation and predictive accuracy recently reported by Ding et al. (2024). In contrast, the 30-plant pooling strategy consistently yielded the best overall performance, with higher R^2 and lower error metrics (MSE, RMSE, MAE, SMAPE, and MAPE). These

results suggest that 30-plant pools provide an effective balance between capturing population-level infection signals and minimizing sampling noise, making this pool size a practical unit for routine field monitoring.

Across the multiple field environments evaluated in this study, the three prediction models exhibited distinct strengths and limitations in estimating RRSV infection rates. The linear regression model performed reasonably well under low to moderate infection conditions but showed limited robustness under high infection pressure, occasionally producing biologically unrealistic estimates beyond the theoretical upper limit. This limitation highlights the restricted flexibility of linear models when applied to complex, non-linear disease systems in the field. The decision tree model showed improved performance under saturated infection conditions, indicating the advantage in handling threshold-like responses. Nevertheless, its strong stepwise structure resulted in abrupt fluctuations across heterogeneous field sites, reducing prediction continuity and stability. Such discontinuities limit the reliability of single-tree models under variable environmental and epidemiological conditions typical of large-scale agricultural fields. In contrast, the random forest model consistently produced stable, biologically constrained, and comparatively accurate predictions across low, moderate, and high infection regimes. Its ensemble learning framework effectively integrates multiple decision trees, thereby reducing sensitivity to local noise, outliers, and environmental heterogeneity. This robustness enables superior generalization across geographically separated fields and diverse disease pressure conditions. These comparative analyses indicate that, although all three models have predictive capacity, the random forest model provides the most robust and reliable performance for practical field application and was therefore selected as the optimal model for large-scale deployment in RRSV surveillance and resistance screening and practical field deployment in RRSV infection-rate estimation.

Regarding the optimal sample size for precise RRSV infection rate prediction, as shown in Section 3.4, predictions from the linear regression (19.85%), decision tree (20.00%), and random forest (21.27%) models using the 30-plant pools were closely aligned with the actual infection rate (20.00%) in rice seedlings and the overall disease incidence (22.92%) across the five separated sampling sites from the same seedling field. These results support the use of a 30-plant pooling strategy as an effective and practical unit for routine field-based RRSV incidence prediction and for informing seedling transplantation management decisions. Although larger sample sizes may marginally improve precision, the 30-plant pooling strategy offers an optimal balance between accuracy, labor efficiency, and operational feasibility.

Furthermore, the quantitative precision provided by the random forest model enables the establishment of a standardized decision-making threshold for field management. Based on our evaluations, we propose an infection rate of 30% as a critical threshold for seedling transplantation. Specifically, seedling beds with a predicted infection rate below 30% are considered suitable for transplantation. From an economic and epidemiological perspective, this threshold is intended as a management reference, not a strict biological limit. Most germplasm exhibits severe symptoms and poor grain filling after RRSV infection. In rice production, farmers usually transplant 2 to 4 seedlings per clump: here, we take 3 seedlings per clump as an example. Therefore, under a moderate compensation scenario (healthy plants compensate for 20% of the

yield), an infection rate of about 30% of seedlings is equivalent to a theoretical yield reduction of about 20%. From an economic perspective, this level of yield loss is sufficient justification in rice production for replacing or avoiding transplanting these highly infected seedlings. However, this threshold should be appropriately adjusted according to the cultivar, planting area, BPH population density, cultivation season, and production management requirements. Therefore, 30% is proposed as a preliminary threshold for which batch replacement or abandonment is a more economical choice, though further field and economic validation across different regions and cultivars is required. To facilitate the practical implementation of our findings, we developed a fully automated RRSV infection rate predictor (30-plant pool), provided in Table S2. This Excel-based tool allows end-users to input relative viral expression levels and instantaneously receive infection rate estimations derived from our optimized random forest, decision tree, and linear regression models.

This study also highlights the importance of selecting models with consideration of field-specific infection conditions. Because linear regression can yield unbounded outputs at high infection pressure, tree-based or ensemble approaches may be preferable when infection levels are high or heterogeneous. Despite the robust performance of the 30-plant pooling framework, several limitations should be acknowledged. First, the current predictive models were developed based on a relatively constrained training dataset. While sufficient for a proof-of-concept and demonstrating high predictive accuracy ($R^2 = 0.862$), the limited sample size may not fully encapsulate the extreme biological stochasticity of RRSV titers across all field scenarios. Consequently, the current random forest model may require further calibration when applied to diverse rice cultivars with variable levels of resistance or under drastically different agro-climatic conditions. Advanced machine learning approaches, including deep learning and crop-specific modeling platforms, should also be explored to enhance prediction performance from increasingly complex datasets. In addition, while the present framework focuses on the plant infection rate, incorporating data on vector abundance and dynamics (BPHs) may further improve interpretability and help link early infection estimates to later disease outcomes. Advanced approaches (e.g., constrained regression, probabilistic calibration, or deep learning when larger datasets become available) could also be explored, but their benefits should be weighed against interpretability and operational feasibility in routine surveillance. In addition, rice cultivars may differ in RRSV accumulation because of differences in resistance. Because pooled $2^{-\Delta Ct}$ values reflect viral load, the model may underestimate infection rates in resistant cultivars and overestimate them in susceptible cultivars. Since the present models were developed using JXB, they are most applicable to commercialized cultivars, which are vulnerable to RRSV. Further validation and recalibration across cultivars with different resistance levels are needed.

5 Conclusion

This study establishes a molecular-computational framework for estimating RRSV infection rates in rice populations by integrating pooled RT-qPCR with machine learning. This approach enables the transformation of molecular measurements into

quantitative population-level estimates, addressing a key limitation in current plant disease diagnostics. Compared with conventional methods, it significantly reduces sampling complexity while maintaining robust predictive performance across diverse field environments. Importantly, this framework provides a scalable strategy for linking molecular data with phenotypic outcomes, offering potential applications in disease surveillance, resistance screening, and crop management. More broadly, our results highlight the potential of integrating molecular diagnostics with data-driven modeling to advance quantitative plant pathology and precision agriculture.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

This work was supported by the National Key Research and Development Program (2022YFD1200702), the Postdoctoral Fellowship Program of China Postdoctoral Science Foundation (GZC20241529) and a fellowship from the China Postdoctoral Science Foundation (2024M762849), and the Hainan Provincial Sanya Yazhou Bay Science and Technology Innovation Joint Project (No: ZDYF2025GXJS129) and Sanya Technology "Fanxing" Special Research Project (No: 2024KJFX026).

Author contributions

Long WANG, Jianxiang WU, Qingyao SHU and Jingjing LI designed the study. Jingjing LI, Meng JIANG, Long WANG, Wenzhuo ZENG and Xiaoyan ZHANG performed the sample collection and molecular detection. Jingjing LI, Meng JIANG and Long WANG performed the model design and field surveys. Jingjing LI wrote the manuscript, Long WANG, Jianxiang WU and Qingyao SHU revised the manuscript.

Compliance with ethics guidelines

Jingjing LI, Meng JIANG, Wenzhuo ZENG, Xiaoyan ZHANG, Jianxiang WU, Qingyao SHU and Long WANG declare that they have no conflicts of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

Declaration on the use of generative AI Tools

During the preparation of this work, the authors used ChatGPT to improve language clarity and readability, and to check for grammatical errors. After using this tool, the authors reviewed and edited the content as needed. The authors take full responsibility for the content of the publication work.

References

Avelino J, Willocquet L, Savary S, 2004. Effects of crop management patterns on coffee rust epidemics. *Plant Pathol*, 53(5):541–547. https://doi.org/10.1111/j.1365-3059.2004.01067.x

Baehaki S, Iswanto EH, Munawar D, et al., 2017. Abilities of brown planthopper

immigrant transmits rice ragged stunt virus on rice of some district of java-indonesia. *Sch J Agric Vet Sci*, 4(8):300–310. https://doi.org/10.36347/sjavs.2017.v04i08.002

Benos L, Tagarakis AC, Dolias G, et al., 2021. Machine learning in agriculture: A comprehensive updated review. *Sensors*, 21(11):3758. https://doi.org/10.3390/s21113758

Boccardo G, 1984. Plant reovirus group. *GMI/AAB Descriptions of plant Viruses*, 294

Bock CH, Chiang K-S, Del Ponte EM, 2022. Plant disease severity estimated visually: A century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Trop Plant Pathol*, 47(1):25–42. https://doi.org/10.1007/s40858-021-00439-z

Bruce P, Bruce A, Gedeck P, 2020. Practical statistics for data scientists. translators, O'Reilly Media, Incorporated Sebastopol,

Chicco D, Warrens MJ, Jurman G, 2021. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Comput Sci*, 7:e623. https://doi.org/10.7717/peerj-cs.623

De Cól M, Coelho M, Del Ponte EM, 2024. Weather-based logistic regression models for predicting wheat head blast epidemics. *Plant Dis*, 108(7):2206–2213. https://doi.org/10.1094/PDIS-11-23-2513-RE

Ding W, Abdel-Basset M, Alrashdi I, et al., 2024. Next generation of computer vision for plant disease monitoring in precision agriculture: A contemporary survey, taxonomy, experiments, and future direction. *Inf Sci*, 665:120338. https://doi.org/10.1016/j.ins.2024.120338

Ginsberg J, Mohebbi MH, Patel RS, et al., 2009. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014. https://doi.org /10.1038/nature07634.

Haq IU, Khan NA, Sarwar MK, 2022. Predictive models for plant disease assessment. In: Ul Haq, I., Ijaz, S. Eds.), Trends in Plant Disease Asessment. Springer Nature Singapore, Singapore, p.225–239.

Hatuwal BK, Shakya A, Joshi B, 2020. Plant leaf disease recognition using random forest, knn, svm and cnn. *Polibits*, 62:13–19.

Hibino H, 1996. Biology and epidemiology of rice viruses. *Annu Rev Phytopathol*, 34:249–274. https://doi.org/10.1146/annurev.phyto.34.1.249

Ibrahim MA, Al-Najafi A, 2020. Modeling, control, and prediction of the spread of covid-19 using compartmental, logistic, and gauss models: A case study in iraq and egypt. *Processes*, 8(11):1400.

Jia D, Guo N, Chen H, et al., 2012. Assembly of the viroplasm by viral non-structural protein pns10 is essential for persistent infection of rice ragged stunt virus in its insect vector. *J Gen Virol*, 93(Pt 10):2299–2309. https://doi.org/10.1099/vir.0.042424-0

Lai D, Zhang Y, Huang Q, et al., 2018. Reverse transcription loop-mediated isothermal amplification to rapidly detect rice ragged stunt virus. *Saudi J Biol Sci*, 25(8):1577–1584. https://doi.org/10.1016/j.sjbs.2016.02.024

Lee S, Yun CM, 2024. Correction: A deep learning model for predicting risks of crop pests and diseases from sequential environmental data. *Plant Methods*. England, p.24.
https://doi.org/10.1186/s13007-024-01140-3

Li J, 2017. Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what? *PLoS One*, 12(8):e0183250. https://doi.org/10.1371/journal.pone.0183250

Ling K, Tiongco E, Aguiero V, 1978. Rice ragged stunt, a new virus disease. *Plant Dis Rep*, 62(8):701–705.

Listihani L, Ariati PEP, Wijaya I, et al., 2023. Relationship study between the brown planthopper population and the intensity of rice ragged stunt virus and rice grassy stunt virus, as well as the inoculum sources. *Int J Agric Technol*, 19(3):1055–1068.

Liu H, Song X-J, Ni Y-Q, et al., 2014. Highly sensitive and specific monoclonal antibody-based serological methods for rice ragged stunt virus detection in rice plants and rice brown planthopper vectors. *J Integr Agric*, 13(9):1943–1951.
https://doi.org/10.1016/S2095-3119(13)60533-X

Livak KJ, Schmittgen TD, 2001. Analysis of relative gene expression data using real-time quantitative pcr and the 2(-delta delta c(t)) method. *Methods*, 25(4):402–408.
https://doi.org/10.1006/meth.2001.1262

Munawar SM, Rajendiran D, Sabjan KB, 2024. Plant diseases diagnosis with artificial intelligence (ai). Microbial data intelligence and computational techniques for sustainable computing. Springer, p.187–193.

Ojokoh B, Sarumi O, Salako K, et al., 2022. Modeling and predicting the spread of covid-19: A continental analysis. Data science for covid-19. Elsevier, p.299–317.

Omura T, Minobe Y, Kimura I, et al., 1983. Improved purification procedure and rna segments of rice ragged stunt virus. *Jpn J Phytopathol*, 49(5):670–675.

Patra J, Chakraborty M, Gupta S, 2023. Random forest algorithm for plant disease prediction. Ai to improve e-governance and eminence of life: Kalyanathon 2020. Springer, p.89–101.

Plonsky L, Ghanbar H, 2018. Multiple regression in l2 research: A methodological synthesis and guide to interpreting r2 values. *Mod Lang J*, 102(4):713–731.
https://doi.org/10.1111/modl.12509

Shikata E, Senboku T, Kamjaipai K, et al., 1979. Rice ragged stunt virus, a new member of plant reovirus group. *Jpn J Phytopathol*, 45(4):436–443.

Singh A, Ganapathysubramanian B, Singh AK, et al., 2016. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci*, 21(2):110–124. https://doi.org/10.1016/j.tplants.2015.10.015

Skelsey P, Dancey SR, Preedy K, et al., 2018. Forecasting the spread of aerially transmitted crop diseases with a binary classifier for inoculum survival. *Plant Pathol*, 67(4):920–928.
https://doi.org/10.1111/ppa.12808

Wang H, Xu D, Pu L, et al., 2014. Southern rice black-streaked dwarf virus alters insect vectors' host orientation preferences to enhance spread and increase rice ragged

stunt virus co-infection. *Phytopathology*, 104(2):196–201. https://doi.org/10.1094/phyto-08-13-0227-r

Wang L, Lu H, Zhang X, et al., 2023. Disruption of serotonin biosynthesis increases resistance to striped stem borer without changing innate defense response in rice. *J Pineal Res*, 75(2):e12895. https://doi.org/10.1111/jpi.12895

Wang P, Liu J, Lyu Y, et al., 2022. A review of vector-borne rice viruses. *Viruses*, 14(10):2258. https://doi.org/10.3390/v14102258

Wang Z, Yu L, Jin L, et al., 2017. Evaluation of rice resistance to southern rice black-streaked dwarf virus and rice ragged stunt virus through combined field tests, quantitative real-time pcr, and proteome analysis. *Viruses*, 9(2) https://doi.org/10.3390/v9020037

Wu J, Du Z, Wang C, et al., 2010. Identification of pns6, a putative movement protein of rrsv, as a silencing suppressor. *Virol J*, 7:335. https://doi.org/10.1186/1743-422x-7-335

Yang X, Lv K, Wang M, et al., 2017. Investigation of viruses infecting rice in southern china using a multiplex rt-pcr assay. *Crop Prot*, 91:8–12. https://doi.org/10.1016/j.cropro.2016.09.012

Yelin I, Aharony N, Tamar ES, et al., 2020. Evaluation of covid-19 rt-qpcr test in multi sample pools. *Clin Infect Dis*, 71(16):2073–2078. https://doi.org/10.1093/cid/ciaa531

Zhang C, Ding Z, Wu K, et al., 2016. Suppression of jasmonic acid-mediated defense by viral-inducible microrna319 facilitates virus infection in rice. *Mol Plant*, 9(9):1302–1314. https://doi.org/10.1016/j.molp.2016.06.014

Zhang W, Ren Y, Guo Z, et al., 2026. Improved lightweight convolutional neural network models for the detection and evaluation of fusarium head blight in wheat. *J Zhejiang Univ Sci B*, 27(5):450–465. https://doi.org/10.1631/jzus.B2500225

Supplementary information

Table S1; Figs. S1-S5