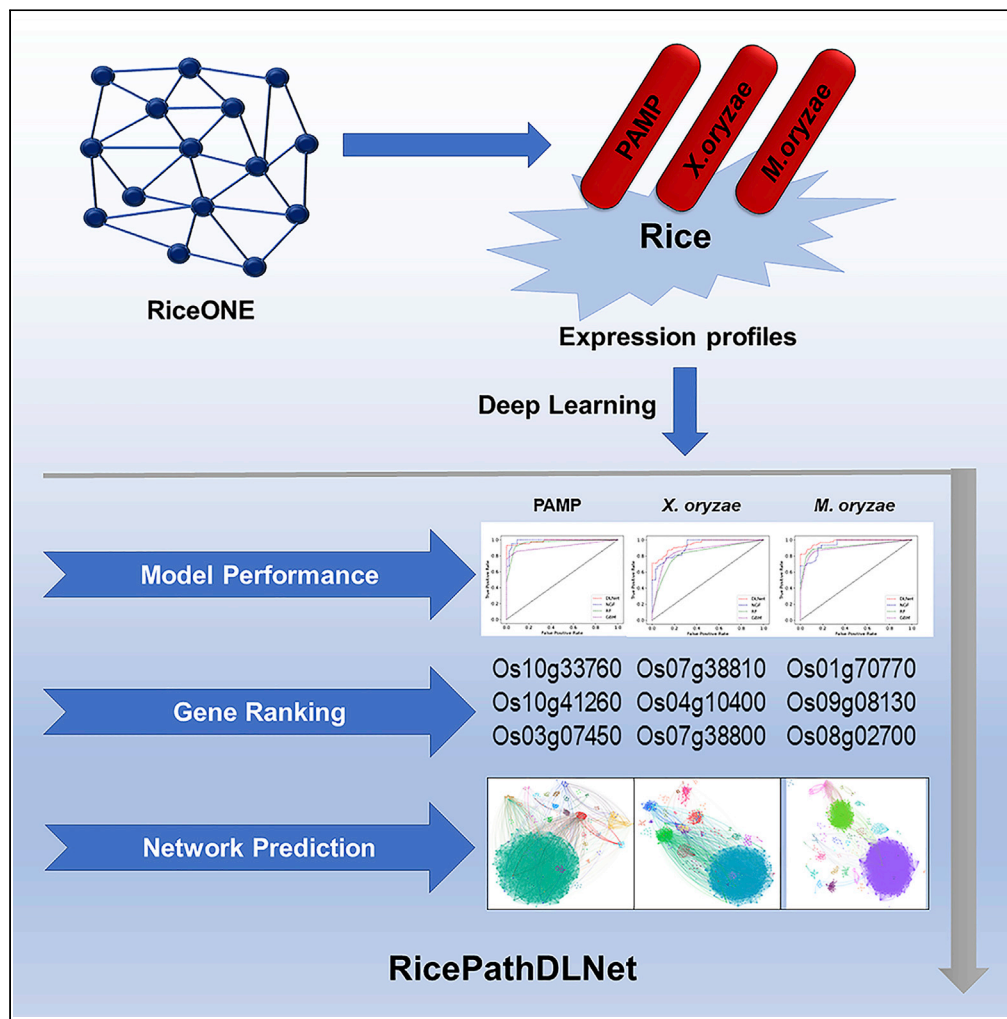**Article**

Deep learning uncovers distinct behavior of rice network to pathogens response

Ravi Kumar,
Abhishek Khatri,
Vishal Acharya

vishal@ihbt.res.in

Highlights

Limited understanding of rice gene network in response to biotic stress

DLNet combines expression profile and network data to predict defense to pathogens

Protein disorder of rice network hub genes varies with different biotic stress

RicePathDLNet reveals distinct gene network architecture of rice to pathogens

Article

Deep learning uncovers distinct behavior of rice network to pathogens response

Ravi Kumar,[1,2] Abhishek Khatri,[1] and Vishal Acharya[1,2,3,*]

SUMMARY

Rice, apart from abiotic stress, is prone to attack from multiple pathogens. Predominantly, the two rice pathogens, bacterial *Xanthomonas oryzae* (*Xoo*) and hemibiotrophic fungus, *Magnaporthe oryzae*, are extensively well explored for more than the last decade. However, because of lack of holistic studies, we design a deep learning-based rice network model (DLNet) that has explored the quantitative differences resulting in the distinct rice network architecture. Validation studies on rice in response to biotic stresses show that DLNet outperforms other machine learning methods. The current finding indicates the compactness of the rice PTI network and the rise of independent modules in the rice ETI network, resulting in similar patterns of the plant immune response. The results also show more independent network modules and minimum structural disorderness in rice-*M. oryzae* as compared to the rice-*Xoo* model revealing the different adaptation strategies of the rice plant to evade pathogen effectors.

INTRODUCTION

The rice plants are constantly exposed to potential pathogens, which cause diseases, compromising crop yield and quality. The major rice pathogens are the hemibiotrophic fungus *Magnaporthe oryzae* (syn. *Pyricularia oryzae*), which causes rice blast and affects all parts of the rice plant (Duan et al., 2014; Liu and Wang, 2016), and the biotrophic bacterium *Xanthomonas oryzae* pv. *Oryzae* (*Xoo*) causes leaf blight (Niño-Liu et al., 2006). These two pathogens account for substantial losses in rice production that impact 60 million people worldwide (Ma et al., 2019; Parker et al., 2008).

To counter pathogen invasion, the plant immune system employs a two-tiered defense mechanism governed by pattern recognition receptors (PRRs) and cytoplasmic immunoreceptors (Han, 2019; Jones and Dangl, 2006), resulting in pattern-triggered immunity (PTI) and the recognition of avirulence (Avr) pathogen effector, activates the resistance (R) genes of a plant, resulting in effector-triggered immunity (ETI) (Meisrimler et al., 2021). The PTI/ETI (PRRs/*R* gene) network fine-tunes the plant response to achieve robust immunity against multiple pathogens (Kim et al., 2014). Based on the hypothesis underlying PTI/ETI discrimination, the entire PTI/ETI network of *Arabidopsis thaliana* was reconstructed, revealing distinct network modules encompassing different immune-related genes (Dong et al., 2015). However, in addition to the PTI/ETI discrimination for a single pathogen, rice is infected by multiple pathogens, such as biotrophic pathogen *Xoo* and hemibiotrophic fungus *M. oryzae* (Dodds, 2010; Kämper et al., 2006), which trigger different ETI response and complicates the study of the desired network. Besides, high yield and broad-spectrum resistance (BSR) are the two most important goals of rice breeding. However, the genes encoding BSR with higher yields are difficult to achieve because of the diverse rice responses to major pathogens (Ke et al., 2017). Together with the limitations of the currently available datasets, the above complications impede understanding the immune-response compendium of genes activated in response to multiple pathogens. So, a complete systematic conception of how the intertwined networks of genes related to the immune system contribute to typical PTI/ETI and within-ETI differences for two contrasting pathogenic lifestyles in rice is urgently needed.

Machine learning (ML) methods are highly effective for automatic information mining from vast datasets. To date, ML approaches have been used in *Arabidopsis* to identify abiotic stress-responsive genes (Ma et al., 2014); compute developmental stage-specific gene-association networks (Bassel et al., 2011); and to differentiate between stress-responsive genes (Dong et al., 2015; Zhang et al., 2020). In rice, ML has been used to distinguish abiotic and biotic stress-responsive genes (Shaik and Ramakrishna, 2014) and
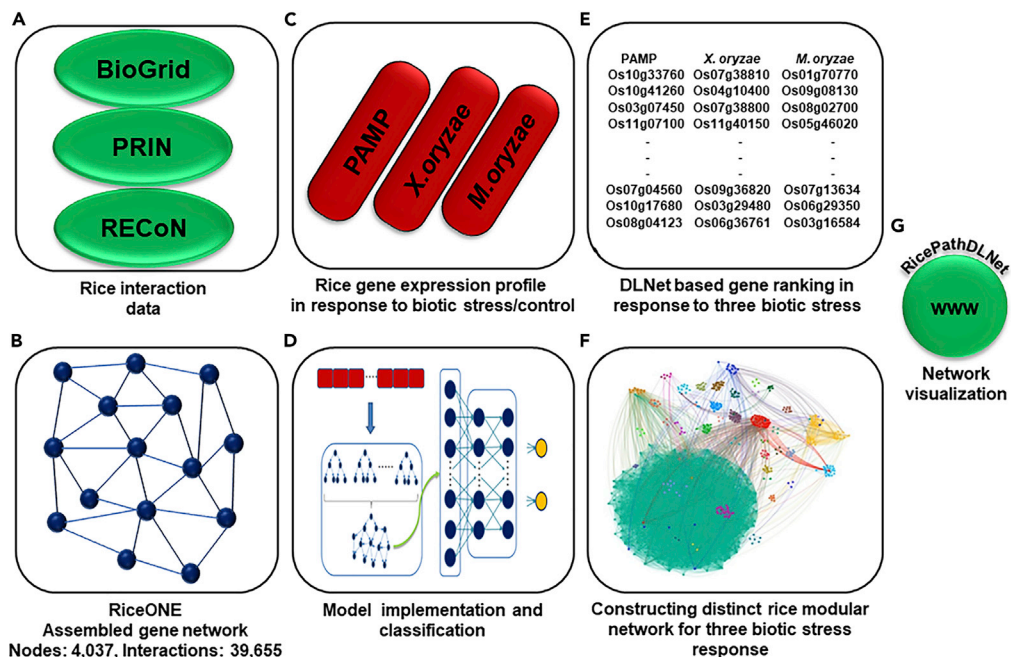
Figure 1. Schematic diagram of the study methodology

(A) Protein-protein interaction and abiotic stress response gene co-expression data were obtained from three databases.

(B) Unreliable gene association data was filtered out (see STAR Methods). The remaining high-quality data was assembled into an integrative gene network with 4,037 genes and 39,655 interactions (RiceONE).

(C) the mRNA expression profile of rice for three biotic stress responses was extracted for 4,037 genes identified in the previous step.

(D) the model was trained and tested for the classification task using 10-fold cross-validation.

(E) the DLNet model is applied to expression data to calculate the feature score for each gene.

(F) DLNet-based modular networks for the three defense responses were constructed using expression profile and integrative interaction data. Permutation, 100 null hypothesis models, and Student's *t*-test were performed to construct the modular network.

(G) network data visualization web browser (herein, named as RicePathDLNet) is developed.

for the development of RicePPINet, a database of predicted protein-protein interactions of rice associated with multiple agronomic traits (Liu et al., 2017). Advances in deep learning (DL), an ML subtype, have enormous potential in this field of research in terms of enhanced precision and have been successfully used in plant phenotypic imaging (Taghavi Namin et al., 2018; Ziamtsov and Navlakha, 2019). In DL applications, to date, the number of samples used exceeds the number of features (i.e., n > p) for predictive model construction. However, in the case of expression profiling data, DL classification is impeded where n < p. Two kinds of DL approaches (based on n < p)—forgeNet and a graph-embedded deep feed-forward network—were recently used successfully as a core for the classification of human tumor samples with higher accuracy (Kong and Yu, 2018, 2020).

In the present study, forgeNet based deep neural network on a rice model was developed that implements the integration of supervised DL algorithm and gene network analysis (here referred to as DLNet in the present study) in response to PAMP/MAMP, *Xoo*, and *M. oryzae* (Figure 1). To apply this framework to analyze the network behavior against different biotic stress responses, we downloaded the raw gene association data from various databases. After filtration of the raw data, we created an interconnected network by integrating high-quality 4,037 genes. Expression profiles of the 4,037 genes in response to PAMP/*Xoo*/*M. oryzae* and their control were downloaded from the NCBI database. DLNet uses the assembled gene network and mRNA expression profile of network genes in response to PAMP/*Xoo*/*M. oryzae* to further recognize candidate plant genes/interactions that participate in the defense response against biotrophic and hemibiotrophic pathogens. DLNet-based gene set enrichment analysis was performed to provide function annotation to each module and reveal the rice gene network's distinct organization to multiple pathogens. Intrinsic disorder proteins (IDPs) or intrinsic disorder regions (IDRs) are an important attribute

of the network hub genes (Ceulemans et al., 2021). We additionally examine the rice hub's disorder property of three different responses.

In comparison, we found that DLNet clearly outperforms state-of-the-art machine learning algorithms; network-guided forest (NGF), gradient boosting machine (GBM), and random forest (RF) on three different treatments. In contrast to NGF, GBM, and RF, DLNet reveals numerous critical genes with known roles in the plant defense response among the top 30 genes for each treatment group. Finally, we made our findings available to the scientific community by creating a user-friendly visualization tool named Rice Pathogen Deep Learning Network (RicePathDLNet). These observations and tools will facilitate the understanding of the rice plant to reveal its adaptational strategy under stress conditions and its engineering that are more resistant to pathogens.

RESULTS

Rice gene network assembly (RiceONE)

We first created an interconnected network of genes by integrating two types of gene-association data: (i) experimental and predicted protein-protein interaction (PPI) data that supported the interologs of six model organisms (large-scale PPI experiments), and could be used for prospective analysis of rice functional and systems biology approaches; and (ii) interaction data for abiotic stress-related transcription factors (TFs) and co-expression relationships. These concise datasets will be highly relevant to analyzing the relationships between gene regulatory mechanisms at the center of the current study (McWhite et al., 2020). For the PPI, interaction authenticity was evaluated using the method developed by Patil and Nakamura (2005). Gene associations in the abiotic stress response were filtered based on the Pearson correlation coefficient and retained if at least one gene in a pair encoded a TF.

The integrated gene network comprises PPI and abiotic stress response data combined with immune response expression data. We excluded the genes from the network that were not accounted for in the Affymetrix Rice Genome Array. An integrated network (RiceONE) incorporates 4,037 genes and their 39,655 interactions. The bench-filtering scheme and increased genome coverage of RiceONE compared to the rice PPI network indicated its suitability for further downstream analysis.

Model prediction performance and comparison

DL-based rice model (DLNet) uses both the forest model and graph-embedded deep-forward network (GEDFN) simultaneously to provide a useful classification model for biotic stress discrimination (Figure 2). We trained the model on rice gene expression (mRNA) data for three distinct immune responses: (i) rice response to pathogen-associated molecular patterns, microbe-associated molecular patterns, and host-derived damage-associated molecular patterns that trigger PTI; (ii) response to *Xoo*, i.e., defense against a bacterial pathogen; and (iii) response to *M. oryzae*, i.e., defense against a fungal pathogen. The classification was achieved via a 10-fold cross-validation process and predicted (a) the final accuracy, representing correctly predicted events; and (b) the area under the receiver operating characteristics curve (AUC), which signified the lack of bias of the classification models used. DL-based rice model was highly reliable in all groups, with an accuracy of approximately 0.92 for PTI, about 0.87 for *Xoo*, 0.94 for *M. oryzae* responses, and an AUC of roughly 0.96 for *Xoo*, and around 0.98 for the PAMP/MAMP and *M. oryzae* response (Table 1; Figures 3A–3C). Previously, one other algorithm (Dong et al., 2015), the network guided forest (NGF), was applied to *Arabidopsis* datasets, yielding an accuracy of 0.85 and AUC of 0.94. Therefore, in order to evaluate the predictive performance and generalization of our developed DLNet rice model, we also attempted NGF, GBM, and RF in the current study. Table 1 reveals that DLNet outperformed NGF, GBM, and RF on all the above datasets.

To gain a detailed review of the performance of the DLNet model, we computed the true positive rate (TPR) for the given false positive rates (FPR) and carried out a parallel comparison with NGF, GBM, and RF on all three datasets. In Table S1, it can be observed that the DLNet rice model shows greater TPR for the given FPR than any other classification method. For example, in response to *M. oryzae*, at an FPR of 0.1 (10%), DLNet gives a TPR of 0.898 (~90%), which is about 0.157 (~16%), 0.08 (8%), and 0.03 (3%) higher than NGF, GBM, and RF, respectively. Similarly, in response to *Xoo*, DLNet gives a TPR of 0.789 (~79%), that is about that is about 0.067 (~7%), 0.098 (~10%), and 0.254 (25%) higher than NGF, GBM, and RF, respectively. On the other hand, in response to PAMP/MAMP, the DLNet model performed almost as good as the other methods (Table S1).
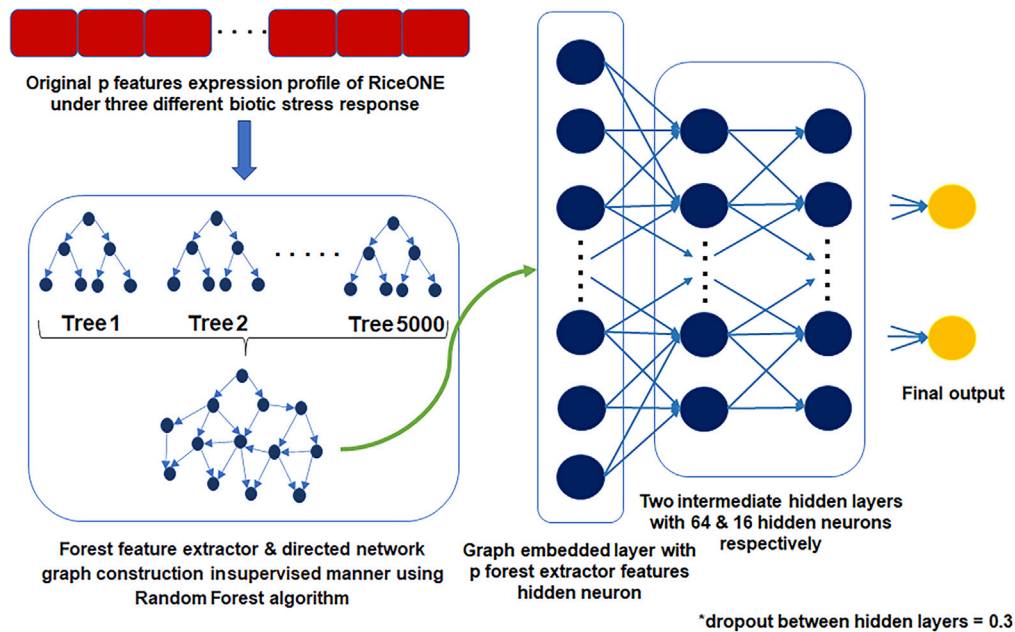
Figure 2. Graphical representation of the DLNet rice model

The feature extractor portion was first used to select the effective features from the original *p* features. The outcome directly involves the construction of a directed feature graph in a supervised manner. The resulted features graph is more informative to the specific classification task. The deep-learning portion uses the features graph as input and comprises four hidden layers with corresponding neurons. In each hidden layer, the number of neurons is: features derived from the graph, 64, 16, and 2, respectively.

DLNet-based ranking of rice features

We averaged the significant score of all ten cross-validation tests for feature selection to obtain gene ranking for the three groups (Tables S2–S4). Among the top 30 genes related to all models, the DLNet model successfully described 14, 15, and 10 immune-related genes for the PAMP/MAMP, *Xoo*, and *M. oryzae* classes, respectively. The top 30 genes with the highest score, complete functional annotation, and the respective reference list for the three categories of responses are provided in Tables S5–S7. We examine the top-ranked 30 genes from each class predicted by DLNet with differentially expressed genes (DEGs). Of note, some of the top-ranked genes predicted by the DLNet were featured in the list of high-ranked DEGs, but present models clearly differentiate the top-ranked gene from the noise of other genes compared to the traditional DEGs method (Figures 4 and S1A–S1F). To test the effectiveness of the developed DLNet rice model, we have attempted to validate it on the three different treatments (PAMP/MAMP, *Xoo*, and *M. oryzae*).

Validation of DLNet on rice response to PAMPs

The DLNet rice model predicted many immune-related genes involved in basal immunity (Table S5). One of the top-ranked lectin receptor-type protein kinases (LOC_OS07g38810) recognized the bacterial pathogen-associated patterns (LPS) and mounted a defense response. It is also involved in the phytohormones (brassinosteroid) signaling pathways against pathogens infection (Girija et al., 2017; Zhang et al., 2016). OsWRKY107 (LOC_Os01g09080), the type II pathogen recognition receptor (PRR), negatively regulated the plant immune response and up-regulated only in response to the compatible interaction with the *Xoo* (Choi et al., 2017). Another ranked gene, receptor-type protein kinase (LOC_Os11g40970; LRR-XII subfamily, non-RD class), homologous of the EF-Tu receptor (EFR), functioned as PRR and was found to regulate pattern trigger immune response in rice (Zhang et al., 2016). A complete list of the top 30 genes in response to PAMP with their role in plant immunity is provided in Table S5.

Validation of DLNet on rice response to *X. oryzae* (*Xoo*)

Similarly, the DLNet rice model also predicted some specific genes in response to a bacterial pathogen, *Xoo* (Table S6). OsRad51A1 (LOC_Os11g40150) is conserved across the species involved in the

Table 1. Classification performance of DLNet model for rice in response to pathogens with other machine learning algorithms

Method	DLNet		Network guided Forest (NGF)		Gradient Boosting Machine (GBM)		Random Forest (RF)	
	ACC[a]	AUC[b]	ACC[a]	AUC[b]	ACC[a]	AUC[b]	ACC[a]	AUC[b]
M. oryzae vs Control	0.94	0.98	0.89	0.97	0.92	0.97	0.91	0.97
SD[c]	0.04	0.03	0.05	0.03	0.05	0.03	0.05	0.03
Xoo vs Control	0.87	0.96	0.84	0.92	0.86	0.95	0.83	0.91
SD[c]	0.04	0.02	0.04	0.06	0.04	0.04	0.08	0.06
PAMP vs Control	0.92	0.98	0.94	0.98	0.92	0.97	0.92	0.98
SD[c]	0.02	0.03	0.02	0.04	0.03	0.04	0.05	0.04

[a]represents accuracy.
[b]represents the area under the receiver operating characteristic curve.
[c]represents standard deviation.

SA-mediated signaling pathways in plant defense response. Overexpressed OsRad51A1 gene in plants prompts a hypersensitive reaction, resulting in cell death and enhanced the disease resistance in rice and *Arabidopsis* by changing the expression level of immune-related genes and SA-dependent signaling genes during inoculation of rice pathogens. Among the top genes, NB-ARC domain-containing disease-resistant gene aquaporin NIP1-1 (LOC_Os02g13870), a plasma membrane aquaporin, was found to have higher expression in response to hydrogen peroxide H_2O_2 which disrupts cell signaling and promotes programmed cell death in plants as a defense mechanism (Sadhukhan et al., 2017). The aquaporin NIP1-1 has been discovered to interact with the bacterial type III effector HopD1 and is crucial to conferring plant defense response (Meisrimler et al., 2021; Weßling et al., 2014). MYB51 (LOC_Os08g33150) is one of the top-ranking genes demonstrated elevated in *Xoo* contaminated rice leaves of vulnerable rice lines (Tariq et al., 2019). A recent study revealed that the Arabidopsis MYB51 gene modulates the indolic glucosinolate synthesis in the leaves and interacts with wound response signaling pathways after inoculation with a bacterial pathogen (Lakshmanan et al., 2012). A complete list of the top 30 genes in response to *Xoo* with their role in plant immunity is provided in Table S6.

Validation of DLNet on rice response to *M. oryzae*

The DLNet rice model predicted particular genes found specific in response to the devastating pathogen *M. oryzae* (Table S7). OsCPK4 (LOC_Os02g03410) confers resistance to fungal pathogen *M. oryzae* by preventing it from penetrating the plant. This resistance is associated with the accumulation of glycosylated salicylic acid hormone, production of ROS, and gene expression of defense-related genes without affecting rice production. Overexpressed OsCPK4-mediated resistance interferes with fungal penetration rather than colonization (Bundó and Coca, 2016). Another essential gene, OsERF83 (LOC_Os03g64260), confirms the disease resistance to the hemibiotrophic pathogen *M. oryzae*. Numerous phytohormones stimulate the OsERF83 expression in the biotic stress response. Its overexpression in transgenic rice lines suppresses the formation of the lesion. It regulates numerous pathogenesis-related genes in response to rice blast disease, indicating that it positively regulates the rice immune response and is involved in several different signaling pathways (Tezuka et al., 2019). Another top-ranked specific gene, OsWAK25 (LOC_Os03g12470), was overexpressed in the early phase of rice blast disease, with a 60-fold increase in expression compared to wild-type and resistance to the biotrophic pathogen *Xoo*. Overexpressed OsWAK25 rice line modulates the NH1 (OsNPR1), which confers disease resistance to biotrophic and hemibiotrophic pathogens and susceptibility to necrotrophic pathogens (Brutus et al., 2010; Harkenrider et al., 2016; Seo et al., 2011). A complete list of the top 30 genes in response to *M. oryzae* with their role in plant immunity is provided in Table S7.

Exploring rice network response to PAMP/MAMP, *Xoo*, and *M. oryzae*

Plants are an excellent example of how nature uses a modular design to organize information, as interconnecting modules play significant roles in specific biological processes (Bassel et al., 2011). To identify densely connected network modules, RiceONE was clustered using the Markov cluster (MCL) algorithm (Enright et al., 2002) and eliminated the modules with less than three genes, resulting in 314 modules (Table S8).
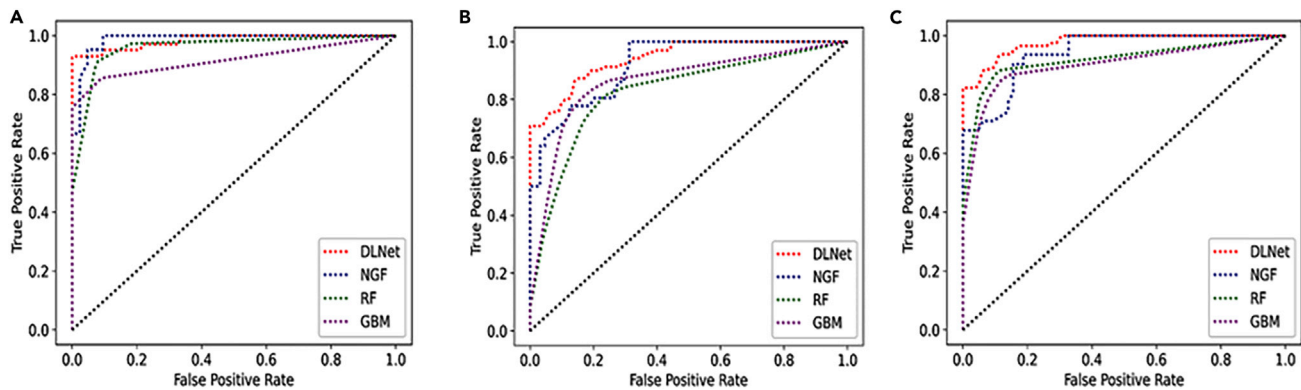
**Figure 3. Classification performance of the model**

(A–C) ROC-AUC plot (A) in response to PAMP/MAMP, (B) in response to *Xoo,* and (C) in response to *M. oryzae.* In response to *Xoo* and *M. oryzae,* DLNet revealed better performance than the other three methods; in response to PAMP/MAMP, DLNet performed comparably to the other three methods.

To gain insight into the overall rice gene networks enabled by PAMP/MAMP, *Xoo*, and *M. oryzae* responses, Gene Set Enrichment Analysis (GSEA) and Student's *t*-test were used to define immune response-related network modules. It led to the construction of three distinct modular network models. First, we performed DLNet-based GSEA by interchanging gene expression data with the genes score. We identified significant modules by using 100 null hypothesis models in the background by randomly changing the phenotypic levels of expression data used by the DLNet model (Figure S2). To detect biological modules of interest, DLNet rice models were used to identify 105, 113, and 104 modules in response to PAMP, *Xoo*, and *M. oryzae* groups, respectively (permutation test, p < 0.05). The generated modules were assigned into five functional groups using the GO biological term; defense response, signal transduction, transcription regulation, growth or development, and hormones. Modules that did not contain these five GO enrichment terms were omitted from further analysis, yielding 39, 39, and 40 modules in the rice network related to PAMP/MAMP, *Xoo*, and *M. oryzae* groups, respectively. These corresponded to 708 genes and 7,788 interactions, 886 genes and 9,431 interactions, and 882 genes and 15,886 interactions, respectively. Interestingly, 11 modules were shared between PAMP/MAMP and the *Xoo* group; 10 modules were shared between PAMP/MAMP and the *M. oryzae* group; 08 modules were shared between *Xoo* and the *M. oryzae* group, and three modules were found to be common to all groups. (Tables S9–S11).

Rice modular network models in response to PAMP/MAMP, *Xoo*, and *M. oryzae* were distinct with higher compactness in the rice's PTI networks, consistent with the *Arabidopsis* PTI network. The generated rice network in response to *Xoo and M. oryzae* was also observed in line with the analysis of *Arabidopsis* (Dong et al., 2015) that encompassed independent modules in response to *M. oryzae* and *Xoo*. Close inspection of the rice immune response to *Xoo* and *M. oryzae* reveals that the modules were more densely connected in the *Xoo* network model than those in the *M. oryzae* network model (Figure 5). The *M. oryzae* network model encompassed numerous independent modules, with only a few detected links to the central module. Moreover, *M. oryzae* is one of rice's most devastating pathogens; our finding of more independent signaling pathways in the rice defense modules model in response to *M. oryzae* also confers its tremendous advantages over the network models in response to PAMP and *Xoo*.

Intrinsic structural disorder prediction of hub genes in the rice network induced by PAMP/MAMP, *Xoo*, and *M. oryzae*

Hub proteins frequently include the intrinsic disorder domain, most often found in regions with intrinsically disordered protein structures. Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) strictly lack defined secondary and tertiary protein structures and possess flexible regions. Upon binding to their particular targets, IDPs or IDRs can bring a transition from disorder to order, resulting in a more stable fold (Hamann et al., 2020). This flexibility of IDPs/IDRs enables the hub protein to undergo various protein-DNA interactions, protein-protein interactions, or posttranslational modification sites. In plant-pathogen interactions, numerous proteins involved in biotic stress stimulus are known to harbor IDRs/IDPs, thus relying on the interaction partners to alter the gene expression and activate the signaling transduction cascade (Covarrubias et al., 2020). A recent review on the disorder (IDRs/IDPs) in hub protein
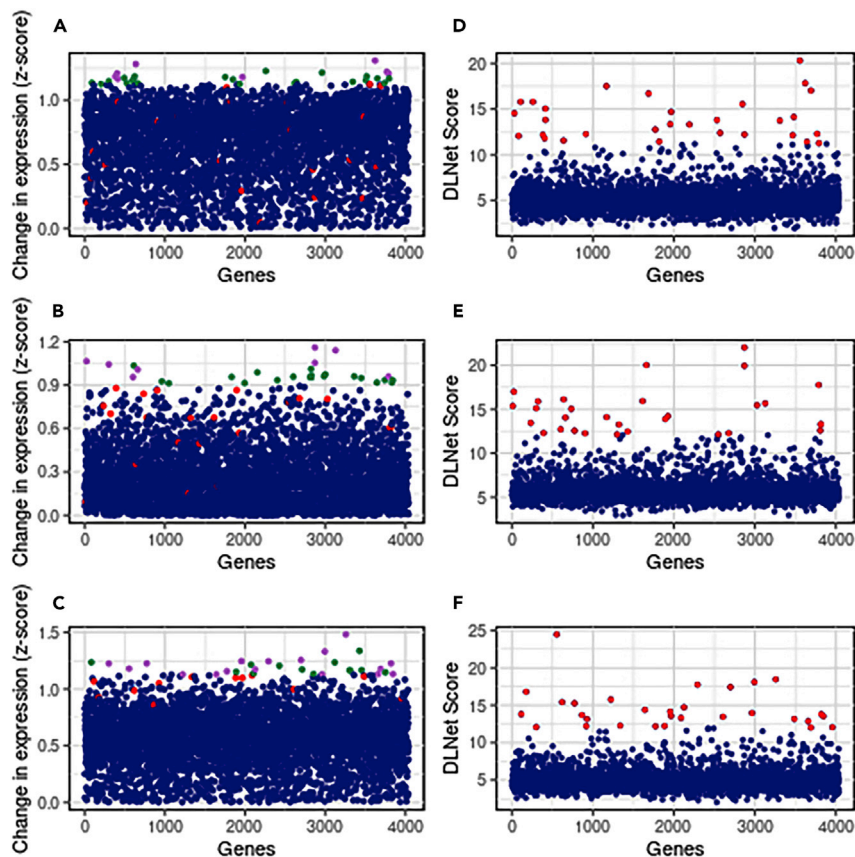
**Figure 4. Differences between gene expression analysis and the DLNet model**

(A–C) Changes in gene expression of top 30 genes (green) (A) in response to PAMP/MAMP associated data, (B) in response to *Xoo*, and (C) in response to *M. oryzae*. Violet color genes are among the 30 top genes discovered by the traditional gene expression method and the DLNet method.

(D and E) DLNet model score of each of the top 30 genes (red) (D) in response to PAMP/MAMP, (E) in response to *Xoo*, and (F) in response to *M. oryzae* ("see also Figure S1").

has explored the nature of immune-related genes during plant-pathogen interactions (Ceulemans et al., 2021). It is still not apparent if hub genes associated with IDPs/IDRs prove to be beneficial to either pathogen or plant itself. Our study has investigated the IDPs/IDRs behavior of rice immune-related hub genes in response to PAMP/MAMP, *Xoo*, and *M. oryzae*.

IDPs/IDRs of the respective hub genes in the three distinct rice networks are computed using an online application called SoftBerry (www.softberry.com). We have selected those hub genes having more than 50 interaction partners as an input to compute the disorder. At the level of more than 50% disorder, we observed 57% (59/104) rice hub genes in response to PAMP, 51% (62/122), and 33% (83/251) rice hub genes in response to *Xoo* and *M. oryzae*, respectively.

Top-ranked genes with more than 50 interactions were thoroughly investigated for disorder prediction using the flDPnn (Hu et al., 2021) (http://biomine.cs.vcu.edu/servers/flDPnn/) in response to PAMP/MAMP, *Xoo*, and *M. oryzae*. Accordingly, OsGRF8 (LOC_Os11g35030) with 65% disorder (Figure 6A), MYB family transcription factor (LOC_Os08g06240) with 59% disorder (Figure 6B) were observed among the top-ranked genes in PAMP/MAMP, and *Xoo* mediated network response. However, in the case of *M. oryzae* response, we observed four of the genes with a significantly lower percentage of disorder level, including leucine-rich repeat family protein (LOC_Os04g55420); 15% disorder (Figure 6C), and GRAS family protein (LOC_Os11g47870); 12% disorder (Figure 6D). The lower percentage of disorder (IDPs/IDRs) hub genes in rice in response to a devastating pathogen (*M. oryzae*) reflects the need to avoid maximum disorderliness in rice hub genes, hence likely probable escape recognition from pathogen effectors. Therefore, it can be
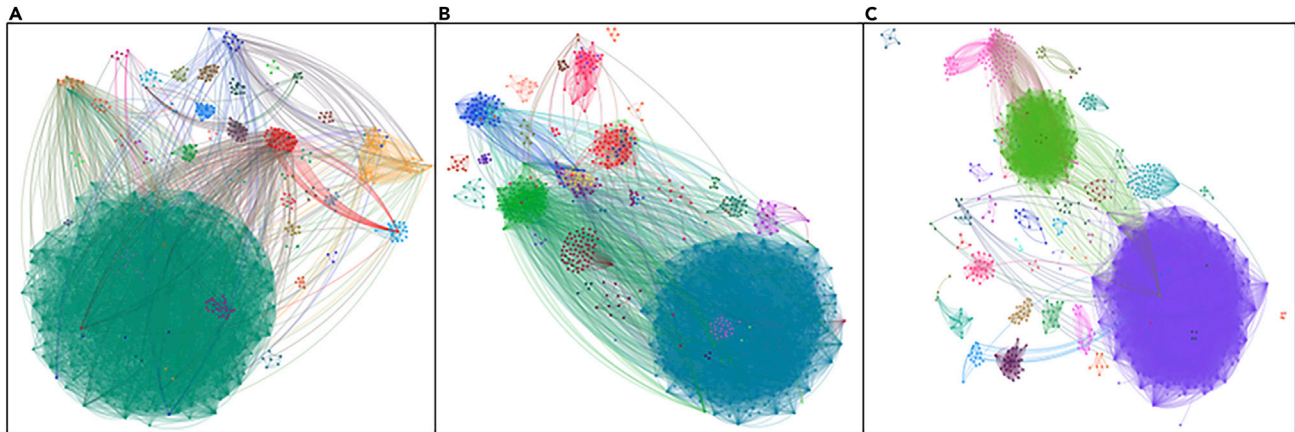
Figure 5. Network architectures of three different immune responses
(A) PTI interactome consists of 39 shared and distinct modules for 708 genes and 7,788 interactions.
(B) ETI-*Xoo* interactome comprises 39 shared and specific modules for 886 genes and 9,431 interactions.
(C) ETI-*M. oryzae* interactome, consisting of 40 shared and distinct modules for 882 genes and 15,886 interactions.

postulated that depending on the stress condition, IDPs/IDRs can play a significant role in comprehending how the signaling cascade contributes from two aspects to the phenotypic plasticity: the pathogenic microbes and the host plant to outcompete one another.

RicePathDLNet: An online visualization repository for exploring network modules

To assist experimental rice researchers, a user-friendly interactive web browser, RicePathDLNet (https://fgcsl.ihbt.res.in/RicePathDLNet), is built using Linux-Apache-MySQL-PHP (LAMP), a package built on the Linux operating system. This repository will provide detailed information on the gene composition for each module from three different network models. Different module architectures in response to biotic stresses are mentioned in detail. It also describes the modules assigned to five GO terms and their responses in plant defense based on the rice genes and their interactions. For further analysis, the user can look through the entire list of genes, their interactions, the list of modules, and the protein sequences of the network hub genes. It also contains the protein sequences of hub genes in response to the above networks. Thus, the RicePathDLNet repository will help in the section that includes comprehensive information on accessing the repository. The model predictions will aid the rice-focused researchers in understanding the distinct network levels in PTI- and ETI-based models and within ETI-*M. oryzae* and ETI-*Xoo* models.

DISCUSSION

Plants are continually competing with pathogens and vice-versa for survival and development. Such interactions in agriculturally important species can directly impact crop yield. One of the major obstacles to attaining maximum rice yield is the severity of diseases caused by multiple pathogenic organisms. To overcome this, plants display dynamic, carefully controlled transcriptome alterations in response to pathogen challenges. Both positive and negative regulators regulate plant immune signaling. It is critical to balance growth and defense responses to respond appropriately to environmental changes.

Interactome mappings in diverse plant species during the past 20 years have led to the expansion of network biology premises. However, it was not enough to gain insights from a single molecular system description in plant-pathogen interactions (Ahmed et al., 2018). Therefore, to address this, we developed such networks based on expression profile data in response to PAMP/*Xoo/M. oryzae*, leading to the construction of DL-based supervised RicePathDLNet. The findings revealed distinct behavior of rice defense architecture and intrinsic disorder structure, providing profound insights into how rice immune networks closely regulate and coordinate with each other depending on stress conditions. The distinct network architecture of the rice in response to pathogens shown in the study will be helpful for further understanding of rice immunity evolution, and it can provide valuable information for breeding the next generation of disease-resistant crops.
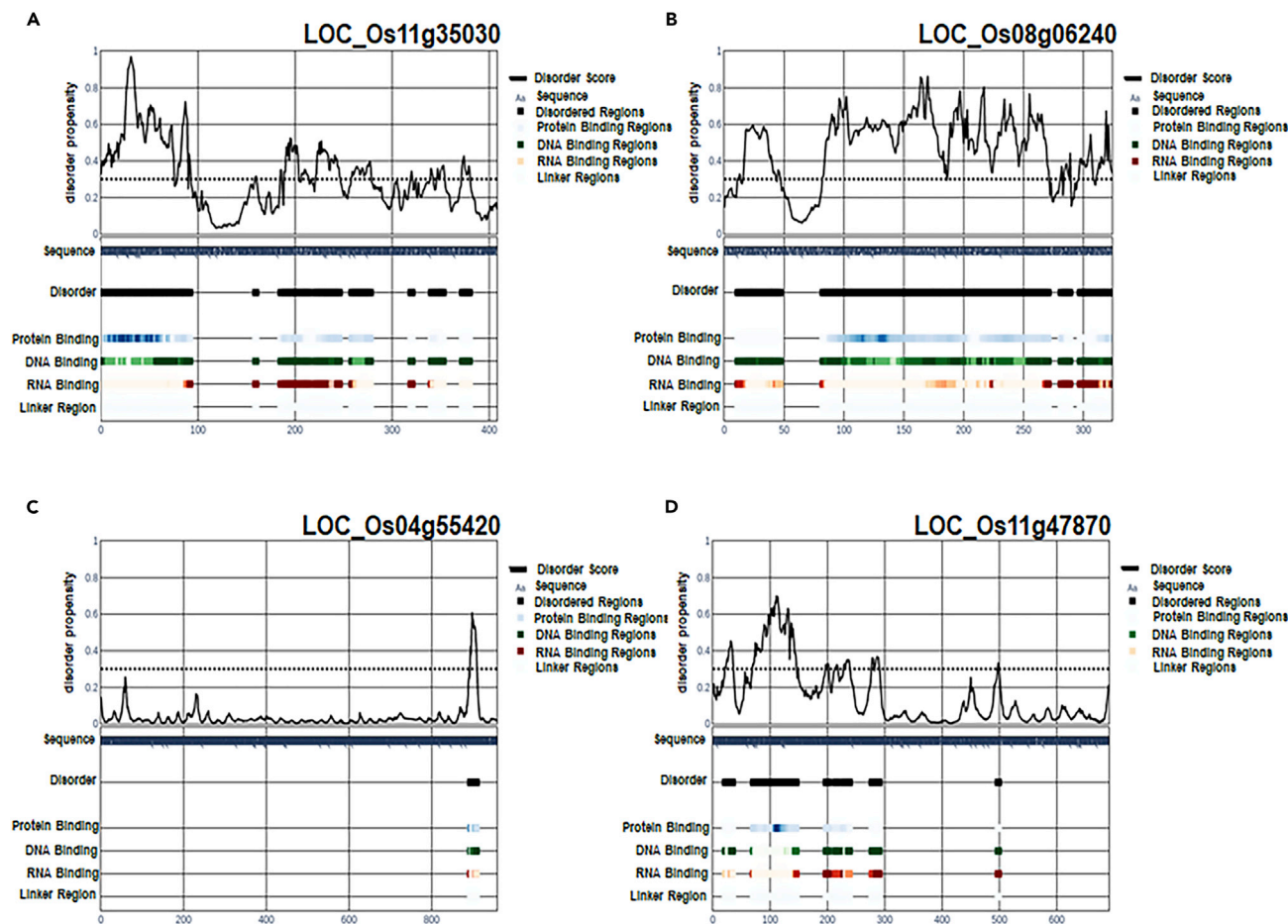
Figure 6. Prediction of intrinsically disordered protein region in network hub genes
(A–D) Intrinsically disordered protein region prediction using flDPnn web server of (A) LOC_Os11g35030, one of the top-ranked and network hub genes in response to PAMP/MAMP. (B) LOC_Os08g06240 is the top-ranked and also a network hub gene in response to *Xoo*. (C and D) LOC_Os04g55420 and LOC_Os11g47870 are the top-ranked and network hub genes in response to *M. oryzae*.

We observed higher discrimination accuracy of the DLNet model compared to other ML (NGF/GBM/RF) methods in all treatment datasets used. The rice protein-protein interaction dataset is still far from complete, which might result in misspecification of the feature graph, hence detrimental to proper classification. The better advantages of the DLNet model in coping with a limited interaction dataset lies in the proper construction of an appropriate graph of features within the feature space from the known expression dataset. Moreover, the other advanced ML algorithm, NGF, incorporates the feature network information from a well-complete feature graph; therefore, it is likely that there might be less accuracy because of the incomplete feature graph obtained.

We generated and explored the list of the top 30 predicted genes and found most of the genes specific to the particular response in rice. Another advantage of the DLNet model reveals a higher probability of rice genes with known defense functions among the top 30 genes for each specific treatment group, in contrast to NGF, GBM, and RF. Some of the top-ranked genes previously unknown to plant immune response found in this study can be considered for further experimental validation. It also observed that several top-ranked immune-related genes identified by DLNet in each group show low or moderate expression in the traditional gene expression method (Figure 4). For example, OsWRKY51 (LOC_04g21950) in the PTI group (Figures S1A and S1B), aquaporin NIP1-1 (LOC Os02g13870) in response to *Xoo* (Figures S1C and S1D), and OsCPK4 (LOC_Os02g03410) in response to *M. oryzae* (Figures S1E and S1F) all are the positive regulators in plant immune response to the pathogen, and neglected by traditional gene expression methods but assign higher ranked in present model classification. These top candidate genes can be used as a

biomarker for specific pathogen responses, whereas shared genes equip rice with broad-spectrum resistance to numerous biotic stresses.

By reconstructing refined rice PPI (RiceONE), we lay out an integrable, step-by-step network analysis structure, RicePathDLNet, which incorporates network assembly as per prior information (a supervised step) to construct the DLNet in response to multiple pathogens. This approach generates different group models, such as PTI, ETI-*Xoo*, and ETI-*M. oryzae*. Analysis of the models of the immune responses of rice and *Arabidopsis* suggested the operation of synergistic signaling pathways in response to PTI and compensatory responses to ETI, in agreement with the conclusions of Tsuda et al. (2009). The complete network analysis of ETI-*Xoo* and ETI-*M. oryzae* generates the hypothesis that the more destructive the pathogen, the network module immune response will be profoundly independent with one module that can affect the encompassing modules that the defense modules in the related model are not affected at all. Elaboration of the PTI and ETI models revealed a distinct network response of the immune-related genes in *Arabidopsis* and rice. In *Arabidopsis*, protein phosphorylation plays a central role in connecting the modules in PTI and ETI groups (Dong et al., 2015).

Interestingly, transcription regulation plays a central module-linking role in all the generated models (Figure S3). Response to a pathogen leads to a dramatic reprogramming of transcription to favor stress/defense responses over routine cellular function. The activity of these transcriptional regulators is orchestrated by a blend of signaling hormones, such as SA, JA, and ethylene (Moore et al., 2011). Previous reports suggested a difference in the core network of rice and *Arabidopsis*, leading to different mechanisms triggered in response to the same pathogen, *M. oryzae* (Park et al., 2009). Hence, the reconstructed network of the immune response of rice to the pathogen(s) appears to be distinct from that in *Arabidopsis*.

For the last decade, IDPs/IDRs assumed much significance, and reportedly, more than 30% of the protein has been predicted as disordered in the eukaryotes. Moreover, proteins with IDPs/IDRs are associated with the pathogenesis of human diseases (Hu et al., 2021). The finding of IDPs/IDRs has challenged the dogma of structural biology, stating that the tertiary structure (3D) determines the protein's function. The lack of 3D folds in IDPs/IDRs resulting in the highly variable regions with some conserved motifs can allow the flexibility of these proteins to mediate interaction for more than one partner(s). The extensive presence of hub proteins harboring IDPs/IDRs under abiotic/biotic stress conditions in plants is well reported (Ceulemans et al., 2021; Marin and Ott, 2014), generating different hypotheses. One hypothesis was whether these plant hub proteins are simultaneously getting any advantage of displaying more percentage of disordered structure (IDPs/IDRs). So, we have thoroughly investigated the hub genes in the PTI, ETI-*Xoo*, and ETI-*M. oryzae* networks that provide insights to display their nature as likely disorder states (IDPs/IDRs). And the proportion of these hub genes surprisingly varies among these three distinct immune-related networks leading to the speculation about how plants have adjusted accordingly depending on the stress conditions.

The lower proportion of rice network IDPs/IDRs hub genes in response to a devastating pathogen reflects the need to avoid maximum disorderliness in rice hub genes, hence likely probable escape recognition from pathogen effectors. This observation led to the hypothesis that the disorder level of the genes might be correlated with the biotic stress condition wherein the percentage of hub genes having IDPs/IDRs is drastically decreasing with the increasing destructiveness of the pathogen. It can thus be postulated that our finding is in line with Carella's zigzag model of immune responses (Carella et al., 2018), where plants and pathogens are in the coevolutionary arms race to outdo each other; more studies will be required to substantiate these hypotheses.

Though the current work is impeded by the unavailability of the large-scale datasets for rice treatments in response to different pathogens, we believe that the problem will be diminished eventfully with the incorporation of more treatment data and advancement in deep learning to cope with small scale datasets. Furthermore, integrated expression profiling of mRNA and their target genes associated with the treatments can enhance and advance our study. Moreover, the role of effectors will be more required to ascertain the complete picture of rice-pathogen interactions. So far, to our knowledge, this is the first-ever implementation of deep learning-based analysis not only for rice but also for the plant genomics datasets.

LIMITATIONS OF THE STUDY

Plant immune network generation is a highly complex process, and this study attempted to focus on distinguished three different immune-related rice network responses under two conditions (Treatment/Control). This method tried to filter the interactions with the gold-standard process; however, because of the limited availability of high-quality omics and molecular network data and thus should be interpreted with caution. Another limitation of current work is that only highly correlated protein-protein interactions are considered; those gene(s) without interaction partner(s) that might play a role in plant immune response will be largely missed from our study. The approach used in this study could be applied to other genomics data to identify actionable gene/module sets in other plant-pathogen genomics data. Still, the main issue is low to moderate ground truth multi-omics plant immune dataset for other models/non-models plant species, which may lead to class imbalance and overcome of negatively labeled samples than positive labeled samples. Plant diseases are always a significant concern in agriculture; if the number of samples increases, this developed method can help identify the defense-related genes at different time phases of the disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Data collection and processing
  - Filtering the PPI and abiotic response network
  - DLNet rice model construction
  - Significant network module identification
  - Gene set enrichment and functional classification
  - Model evaluation and performance metrics
  - DLNet-based rice model implementation, architecture, training and testing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis

AUTHOR CONTRIBUTIONS

V.A. conceived the project. R.K. and V.A. designed the experiments. R.K. carried out the experiments. R.K. and V.A. analyzed the data. A.K. designed and implemented the data visualization server. R.K. and V.A. wrote the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (USENIX Association), pp. 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

Ahmed, H., Howton, T.C., Sun, Y., Weinberger, N., Belkhadir, Y., and Mukhtar, M.S. (2018). Network biology discovers pathogen contact points in host protein-protein interactomes. Nat. Commun. 9, 2312. https://doi.org/10.1038/s41467-018-04632-8.

Bassel, G.W., Glaab, E., Marquez, J., Holdsworth, M.J., and Bacardit, J. (2011). Functional network construction in Arabidopsis using rule-based machine learning on large-scale data sets. Plant Cell 23, 3101–3116. https://doi.org/10.1105/tpc.111.088153.

Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brohée, S., and van Helden, J. (2006). Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformatics 7, 488. https://doi.org/10.1186/1471-2105-7-488.

Brutus, A., Sicilia, F., Macone, A., Cervone, F., and De Lorenzo, G. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. Proc. Natl. Acad. Sci. USA 107, 9452–9457. https://doi.org/10.1073/pnas.1000675107.

Bundó, M., and Coca, M. (2016). Enhancing blast disease resistance by overexpression of the calcium-dependent protein kinase OsCPK4 in rice. Plant Biotechnol. J. 14, 1357–1367. https://doi.org/10.1111/pbi.12500.

Carella, P., Evangelisti, E., and Schornack, S. (2018). Sticking to it: phytopathogen effector molecules may converge on evolutionarily conserved host targets in green plants. Curr. Opin. Plant Biol. 44, 175–180. https://doi.org/10.1016/j.pbi.2018.04.019.

Ceulemans, E., Ibrahim, H.M.M., De Coninck, B., Coninck, B.D., and Goossens, A. (2021). Pathogen Effectors: exploiting the promiscuity of plant signaling hubs. Trends Plant Sci. 26, 780–795. https://doi.org/10.1016/j.tplants.2021.01.005.

Choi, N.Y., Lee, E., Lee, S.G., Choi, C.H., Park, S.R., Ahn, I., Bae, S.C., Hwang, C.H., and Hwang, D.-J. (2017). Genome-wide expression profiling of OsWRKY superfamily genes during infection with Xanthomonas oryzae pv. oryzae using real-time PCR. Front. Plant Sci. 8, 1628. https://doi.org/10.3389/fpls.2017.01628.

Covarrubias, A.A., Romero-Pérez, P.S., Cuevas-Velazquez, C.L., and Rendón-Luna, D.F. (2020). The functional diversity of structural disorder in plant proteins. Arch. Biochem. Biophys. 680, 108229. https://doi.org/10.1016/j.abb.2019.108229.

Dodds, P.N. (2010). Genome evolution in plant pathogens. Science 330, 1486–1487. https://doi.org/10.1126/science.1200245.

Dong, X., Jiang, Z., Peng, Y.-L., and Zhang, Z. (2015). Revealing shared and distinct gene network organization in arabidopsis immune responses by integrative analysis. Plant Physiol. 167, 1186–1203. https://doi.org/10.1104/pp.114.254292.

Duan, L., Liu, H., Li, X., Xiao, J., and Wang, S. (2014). Multiple phytohormones and phytoalexins are involved in disease resistance to Magnaporthe oryzae invaded from roots in rice. Physiol. Plantarum 152, 486–500. https://doi.org/10.1111/ppl.12192.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584. https://doi.org/10.1093/nar/30.7.1575.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307–315. https://doi.org/10.1093/bioinformatics/btg405.

Girija, A.M., Kinathi, B.K., Madhavi, M.B., Ramesh, P., Vungarala, S., Patel, H.K., and Sonti, R.V. (2017). Rice leaf transcriptional profiling suggests a functional interplay between Xanthomonas oryzae pv. oryzae lipopolysaccharide and extracellular polysaccharide in modulation of defense responses during infection. Mol. Plant Microbe Interact. 30, 16–27. https://doi.org/10.1094/MPMI-08-16-0157-R.

Gu, H., Zhu, P., Jiao, Y., Meng, Y., and Chen, M. (2011). PRIN: a predicted rice interactome network. BMC Bioinformatics 12, 161. https://doi.org/10.1186/1471-2105-12-161.

Hamann, F., Schmitt, A., Favretto, F., Hofele, R., Neumann, P., Xiang, S., Urlaub, H., Zweckstetter, M., and Ficner, R. (2020). Structural analysis of the intrinsically disordered splicing factor Spp2 and its binding to the DEAH-box ATPase Prp2. Proc. Natl. Acad. Sci. USA 117, 2948–2956. https://doi.org/10.1073/pnas.1907960117.

Han, G.Z. (2019). Origin and evolution of the plant immune system. New Phytol. 222, 70–83. https://doi.org/10.1111/nph.15596.

Harkenrider, M., Sharma, R., De Vleesschauwer, D., Tsao, L., Zhang, X., Chern, M., Canlas, P., Zuo, S., and Ronald, P.C. (2016). Overexpression of rice wall-associated kinase 25 (OsWAK25) alters resistance to bacterial and fungal pathogens. PLoS One 11, e0147310. https://doi.org/10.1371/journal.pone.0147310.

Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J., and Kurgan, L. (2021). flDPnn: accurate intrinsic disorder prediction with putative propensities of disorder functions. Nat. Commun. 12, 4438. https://doi.org/10.1038/s41467-021-24773-7.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning (MIT Press).

Jones, J.D.G., and Dangl, J.L. (2006). The plant immune system. Nature 444, 323–329. https://doi.org/10.1038/nature05286.

Kämper, J., Kahmann, R., Bölker, M., Ma, L.-J., Brefort, T., Saville, B.J., Banuett, F., Kronstad, J.W., Gold, S.E., Müller, O., et al. (2006). Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis. Nature 444, 97–101. https://doi.org/10.1038/nature05248.

Ke, Y., Deng, H., and Wang, S. (2017). Advances in understanding broad-spectrum resistance to pathogens in rice. Plant J. 90, 738–748. https://doi.org/10.1111/tpj.13438.

Kim, Y., Tsuda, K., Igarashi, D., Hillmer, R.A., Sakakibara, H., Myers, C.L., and Katagiri, F. (2014). Mechanisms underlying robustness and tunability in a plant immune signaling network. Cell Host Microbe 15, 84–94. https://doi.org/10.1016/j.chom.2013.12.002.

Kingma, D.P., and Ba, J. (2014). Adam: a method for stochastic optimization. Preprint at arXiv. https://arxiv.org/abs/1412.6980v9.

Kong, Y., and Yu, T. (2020). forgeNet: a graph deep neural network model using tree-based ensemble classifiers for feature graph construction. Bioinformatics 36, 3507–3515. https://doi.org/10.1093/bioinformatics/btaa164.

Kong, Y., and Yu, T. (2018). A graph-embedded deep feed-forward network for disease outcome classification and feature selection using gene expression data. Bioinformatics 34, 3727–3737. https://doi.org/10.1093/bioinformatics/bty429.

Krishnan, A., Gupta, C., Ambavaram, M.M.R., and Pereira, A. (2017). RECoN: rice environment coexpression network for systems level analysis of abiotic-stress response. Front. Plant Sci. 8, 1640. https://doi.org/10.3389/fpls.2017.01640.

Lakshmanan, V., Kitto, S.L., Caplan, J.L., Hsueh, Y.-H., Kearns, D.B., Wu, Y.-S., and Bais, H.P. (2012). Microbe-associated molecular patterns-triggered root responses mediate beneficial rhizobacterial recruitment in arabidopsis. Plant Physiol. 160, 1642–1661. https://doi.org/10.1104/pp.112.200386.

Lazar, C., Meganck, S., Taminau, J., Steenhoff, D., Coletta, A., Molter, C., Weiss-Solís, D.Y., Duque, R., Bersini, H., and Nowé, A. (2013). Batch effect removal methods for microarray gene expression data integration: a survey. Brief. Bioinform. 14, 469–490. https://doi.org/10.1093/bib/bbs037.

Liu, S., Liu, Y., Zhao, J., Cai, S., Qian, H., Zuo, K., Zhao, L., and Zhang, L. (2017). A computational interactome for prioritizing genes associated with complex agronomic traits in rice (Oryza sativa). Plant J. 90, 177–188. https://doi.org/10.1111/tpj.13475.

Liu, W., and Wang, G.L. (2016). Plant innate immunity in rice: a defense against pathogen infection. Natl. Sci. Rev. 3, 295–308. https://doi.org/10.1093/nsr/nww015.

Ma, C., Xin, M., Feldmann, K.A., and Wang, X. (2014). Machine learning–based differential network analysis: a study of stress-responsive transcriptomes in Arabidopsis. Plant Cell 26, 520–537. https://doi.org/10.1105/tpc.113.121913.

Ma, S., Song, Q., Tao, H., Harrison, A., Wang, S., Liu, W., Lin, S., Zhang, Z., Ai, Y., and He, H. (2019). Prediction of protein-protein interactions between fungus (Magnaporthe grisea) and rice

(Oryza sativa L.). Brief. Bioinformation *20*, 448–456. https://doi.org/10.1093/bib/bbx132.

Marín, M., and Ott, T. (2014). Intrinsic disorder in plant proteins and phytopathogenic bacterial effectors. Chem. Rev. *114*, 6912–6932. https://doi.org/10.1021/cr400488d.

McWhite, C.D., Papoulas, O., Drew, K., Cox, R.M., June, V., Dong, O.X., Kwon, T., Wan, C., Salmi, M.L., Roux, S.J., et al. (2020). A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell *181*, 460–474.e14. https://doi.org/10.1016/j.cell.2020.02.049.

Meisrimler, C.N., Allan, C., Eccersall, S., and Morris, R.J. (2021). Interior design: how plant pathogens optimize their living conditions. New Phytol. *229*, 2514–2524. https://doi.org/10.1111/nph.17024.

Moore, J.W., Loake, G.J., and Spoel, S.H. (2011). Transcription dynamics in plant immunity. Plant Cell *23*, 2809–2820. https://doi.org/10.1105/tpc.111.087346.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273. https://doi.org/10.1038/ng1180.

Niño-Liu, D.O., Ronald, P.C., and Bogdanove, A.J. (2006). *Xanthomonas oryzae* pathovars: model pathogens of a model crop. Mol. Plant Pathol. *7*, 303–324. https://doi.org/10.1111/j.1364-3703.2006.00344.x.

Oughtred, R., Stark, C., Breitkreutz, B.-J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., et al. (2019). The BioGRID interaction database: 2019 update. Nucleic Acids Res. *47*, D529–D541. https://doi.org/10.1093/nar/gky1079.

Park, J.Y., Jin, J., Lee, Y.W., Kang, S., and Lee, Y.H. (2009). Rice blast fungus (*Magnaporthe oryzae*) infects arabidopsis via a mechanism distinct from that required for the infection of rice. Plant Physiol. *149*, 474–486. https://doi.org/10.1104/pp.108.129536.

Parker, D., Beckmann, M., Enot, D.P., Overy, D.P., Rios, Z.C., Gilbert, M., Talbot, N., and Draper, J.

(2008). Rice blast infection of Brachypodium distachyon as a model system to study dynamic host/pathogen interactions. Nat. Protoc. *3*, 435–445. https://doi.org/10.1038/nprot.2007.499.

Patil, A., and Nakamura, H. (2005). Filtering high-throughput protein-protein interaction data using a combination of genomic features. BMC Bioinformatics *6*, 100. https://doi.org/10.1186/1471-2105-6-100.

Sadhukhan, A., Kobayashi, Y., Nakano, Y., Iuchi, S., Kobayashi, M., Sahoo, L., and Koyama, H. (2017). Genome-wide association study reveals that the aquaporin NIP1;1 contributes to variation in hydrogen peroxide sensitivity in *Arabidopsis thaliana*. Mol. Plant *10*, 1082–1094. https://doi.org/10.1016/j.molp.2017.07.003.

Seo, Y.S., Chern, M., Bartley, L.E., Han, M., Jung, K.H., Lee, I., Walia, H., Richter, T., Xu, X., Cao, P., et al. (2011). Towards establishment of a rice stress response interactome. PLoS Genet. *7*, e1002020. https://doi.org/10.1371/journal.pgen.1002020.

Shaik, R., and Ramakrishna, W. (2014). Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. Plant Physiol. *164*, 481–495. https://doi.org/10.1104/pp.113.225862.

Taghavi Namin, S., Esmaeilzadeh, M., Najafi, M., Brown, T.B., and Borevitz, J.O. (2018). Deep phenotyping: deep learning for temporal phenotype/genotype classification. Plant Methods *14*, 66. https://doi.org/10.1186/s13007-018-0333-4.

Tariq, R., Ji, Z., Wang, C., Tang, Y., Zou, L., Sun, H., Chen, G., and Zhao, K. (2019). RNA-Seq analysis of gene expression changes triggered by *Xanthomonas oryzae pv. oryzae* in a susceptible rice genotype. Rice *12*, 44. https://doi.org/10.1186/s12284-019-0301-2.

Tezuka, D., Kawamata, A., Kato, H., Saburi, W., Mori, H., and Imai, R. (2019). The rice ethylene response factor OsERF83 positively regulates disease resistance to *Magnaporthe oryzae*. Plant Physiol. Biochem. *135*, 263–271. https://doi.org/10.1016/j.plaphy.2018.12.017.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. *45*, W122–W129. https://doi.org/10.1093/nar/gkx382.

Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J., and Katagiri, F. (2009). Network properties of robust immunity in plants. PLoS Genet. *5*, e1000772. https://doi.org/10.1371/journal.pgen.1000772.

Vlasblom, J., and Wodak, S.J. (2009). Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. BMC Bioinformatics *10*, 99. https://doi.org/10.1186/1471-2105-10-99.

Weßling, R., Epple, P., Altmann, S., He, Y., Yang, L., Henz, S.R., McDonald, N., Wiley, K., Bader, K.C., Gläßer, C., et al. (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. Cell Host Microbe *16*, 364–375. https://doi.org/10.1016/j.chom.2014.08.004.

Yasrebi, H. (2016). Comparative study of joint analysis of microarray gene expression data in survival prediction and risk assessment of breast cancer patients. Brief. Bioinform. *17*, 771–785. https://doi.org/10.1093/bib/bbv092.

Yi, X., Du, Z., and Su, Z. (2013). PlantGSEA: a gene set enrichment analysis toolkit for plant community. Nucleic Acids Res. *41*, W98–W103. https://doi.org/10.1093/NAR/GKT281.

Zhang, N., Zhou, S., Yang, D., and Fan, Z. (2020). Revealing shared and distinct genes responding to JA and SA signaling in Arabidopsis by meta-analysis. Front. Plant Sci. *11*, 908. https://doi.org/10.3389/fpls.2020.00908.

Zhang, Y., Zhao, J., Li, Y., Yuan, Z., He, H., Yang, H., Qu, H., Ma, C., and Qu, S. (2016). Transcriptome analysis highlights defense and signaling pathways mediated by rice pi21 gene with partial resistance to *Magnaporthe oryzae*. Front. Plant Sci. *7*, 1834. https://doi.org/10.3389/fpls.2016.01834.

Ziamtsov, I., and Navlakha, S. (2019). Machine learning approaches to improve three basic plant phenotyping tasks using three-dimensional point clouds. Plant Physiol. *181*, 1425–1440. https://doi.org/10.1104/pp.19.00524.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw expression data	NCBI	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL2025/ and Table S12
Experimental and predicted protein-protein interactions	BioGRID	https://thebiogrid.org/
Predicted protein-protein interactions	PRIN	http://bis.zju.edu.cn/prin/
Co-expression interactions data under abiotic stress	RECoN	https://plantstress-pereira.uark.edu/RECoN/
Code availability	Present study	https://gitlab.com/ravisaroch/dlnet-host-pathogen-interaction/
RicePathDLNet	Present study	https://fgcsl.ihbt.res.in/RicePathDLNet/
Software and algorithms		
Python3	Python	https://www.python.org/
NumPy	NumPy	https://numpy.org/
Pandas	pandas	https://pandas.pydata.org/
Sklearn	sklearn	https://scikit-learn.org/stable/
Random Forest	sklearn	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier/
XGBoost	sklearn	https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier/
R	R Project	https://cran.r-project.org/
Tensorflow 2	Tensorflow	https://www.tensorflow.org/
SciPy	SciPy	https://scipy.org/
matplotlib	Python Library	https://matplotlib.org/
mcl	Enright et al. (2002)	https://micans.org/mcl/

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources should be directed to the lead contact, Vishal Acharya (vishal@ihbt.res.in).

Material availability

This study did not generate new unique reagents or materials.

Data and code availability

- The raw expression data used in the study were publicly available and provided as Table S12. Additionally, datasets resources link is provided in key resource table.
- The source code used in the study are available at GitLab link https://gitlab.com/ravisaroch/dlnet-host-pathogen-interaction. RicePathDLNet: an online visualization repository for exploring network modules and other data is available at https://fgcsl.ihbt.res.in/RicePathDLNet.

● Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon reasonable request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No experimental model system used for this study.

METHOD DETAILS

Data collection and processing

Experimental and multispecies predicted PPI network was compiled from BioGRID (Oughtred et al., 2019) and the Predicted Rice Interactome Network (PRIN) databases (Gu et al., 2011). The abiotic stress response co-expression network dataset was downloaded from the Rice Environment Coexpression Network (RECoN) (Krishnan et al., 2017).

Rice gene expression data for four different immune responses (PTI, *Xoo*, *M. oryzae*, and control) were downloaded from the Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/geo) of the National Center for Biotechnology Information. For the PTI response, 21 gene expression profiles were extracted for rice treated with PAMPs, MAMPs, and DAMPs. For the ETI response, 36 and 31 gene expression profiles were extracted in response to avirulent *Xoo* and *M. oryzae* strains, respectively. The control dataset comprised 64 gene expression profiles that reflected control experiments for the above treatments. Complete information on the gene expression profiles is provided in Table S12. Raw expression data (Affymetrix platform) were pre-processed using a robust multichip average (RMA) process of affy package (Gautier et al., 2004), and z-score transformed to eliminate the batch effect as z-score overcome the other well-known methods like ComBat and the inverse normal method (meta-analysis) (Lazar et al., 2013; Yasrebi, 2016).

Filtering the PPI and abiotic response network

The accuracy of PPI network was evaluated by adopting the filtering process of Patil and Nakamura (2005). The interaction between two proteins was deemed valid, providing at least one annotation with the same GO. agriGo v2.0 (Tian et al., 2017) was used to compute the Gene Ontology of the interacting pairs.

For the abiotic response network data, only interaction pairs with a Pearson correlation coefficient (PCC) $\geq$ 0.7 and with at least one TF gene in a pair were retained as PCC value equal to or larger than 0.7 for interaction data is considered as the best probable candidates for protein-protein interactions. The immune-related gene co-expression network profile was avoided to reduce the risk of introducing bias in the algorithm (Table S13).

DLNet rice model construction

DLNet model was constructed to calculate features importance, model classification, and identification of significance modules rice data under three distinct biotic stress conditions. The DLNet framework comprises three components: the features extractor part, deep feed-forward neural network (DFN) part (see detail in Kong and Yu, 2020) and significant modules identification. First, the extractor part selects the useful features from the expression profile of rice treated with PAMPs, *X. oryzae,* and *M. oryzae* respectively, using the forest model of rice genes with the supervision of a training label for each sample (treatment + control) of rice. It constructs a directed feature graph based on individual decision tree splitting order for each expression profile.

$$F(\Theta) = \{T_m(\Theta_m)\}, \ m = 1, \ldots\ldots, M,$$

where *F* represents the assembly of decision trees by extracting the useful subset of rice features (genes); *M* corresponds to the total number of forest trees; and $\Theta = \{\Theta = \Theta_1, \ldots\ldots, \Theta_m\}$ describes the splitting values and splitting variables, that is, the splitting order parameters. *F* is equipped with the rice expression profile (training) data and rice expression (training) levels (rice features) as $X_{train} \in \Re^n{}_{train}{}^{\times P}$ and $Y_{train} \in \Re^n{}_{train,}$ respectively, to build decision trees with a subset of features and directed connections as splitting order. Then, the set of graphs with directed edges can be built as

$$G = \{G_m(V_m, E_m)\}, m = 1, \ldots\ldots, M,$$

where V_m corresponds to vertices (features) and E_m corresponds to edges (features connections) in G_m. The aggregate rice feature graph G obtained after the assembly of all graphs is represented as

$$G(V, E) = \bigcup_{m=1}^{M} G_m(V_m, E_m)$$

A directed rice feature graph contains only those features that at least once participate in the graph to split the given samples in the rice expression profile. The input feature graph as a matrix is then transformed as $X_{train} \in \Re^{n \times |v|}$ because the number of vertices denoted in the feature graph (G) is $|v|$ and $|v| < p$ (Kong and Yu, 2020).

The second part of the model uses the deep feed-forward neural network (DFN) to incorporate the graph layer. DFN can be described as follows:

$$Pr(y|X, O) = f(Z_{out}W_{out} + b_{out})$$

Where $X = X_{train} \in \Re^{n \times |v|}$ represents the data matrix of rice features with n number of rice samples and $|v|$ are extracted features from the original rice expression profile; $y_{train} \in \Re^n$ is the outcome label for binary ($b_{out} = 2$) classification task (control or treated with PAMPs/*Xoo/M. oryzae*); O represents the possible parameter of the model and $f(\cdot)$ represents the softmax function, which converts the output layer into the predicted probability.

$$Z_{out} = \sigma(Z_l W_l + b_l)$$
$$\ldots\ldots$$
$$Z_{k+1} = \sigma(Z_k W_k + b_k)$$
$$\ldots\ldots$$
$$Z_1 = \sigma(X W_{in} + b_{in})$$

Where $Z_k, (k=1,\ldots\ldots, l\text{-}1, out)$ are neurons in the hidden layer; W_k corresponds to the weight matrix; b_k represents the biasness, and $\sigma(\cdot)$ represents the rectified linear unit (ReLu) activation function. A stochastic gradient descent (SGD) based algorithm (Adam optimizer) is used to train the model by a cross-entropy loss function to update the parameter O. To change the weight parameters, the back-propagation algorithm calculates from the cost function parameter and flows back over the network to measure the gradient (Goodfellow et al., 2016)

$$H(O) = -\frac{1}{n} \sum_{i=1}^{n} \{ y_i \log(\widehat{p}_i) + (1 - y_i)\log(1 - \widehat{p}_i) \}$$

$H(O)$ is the scalar cost function, and $\widehat{p}_i$ is the fitted value of p_i (probability prediction). A feature graph information incorporated in the DFN to construct GEDFN with the dimensions of the first hidden layer (Z_1) is the same as the input layer (containing the same subset of rice features). A sparse connection can be achieved between the input layer and the first hidden layer, and $Z_1 = \sigma(X W_{in} + b_{in})$ is changed to

$$Z_1 = \sigma(X(W_{in} \odot A) + b_{in})$$

Here $\odot$ is the hamadard product, and A is the asymmetric adjacency matrix (in the current study), which filters the connections between the input and the first hidden layer. More details about GEDFN can be found in Kong and Yu (2018).

The PTI and control expression profiles were trained to generate the PTI-interlinked model; similarly, the Xoo-ETI–interlinked model generated from gene expressions of *Xoo*, and the control; and, expression profiles of *M. oryzae* and the control were implemented to train the *M. oryzae*-ETI interlined model. The classification performance was conducted via 10-fold cross-validation; the final AUC and accuracy of the model were assessed by average testing AUC and accuracy across the 10 datasets.

The significant contribution of the features to the classification imparted the primary biological mechanism. To calculate the importance of the features in the classification, the model uses the information of nodes in the input layer and the hidden neurons in the first hidden layer. For each feature, the significance score is determined as the sum of absolute weight values directly associated with the characteristic node and its corresponding hidden neuron in the first hidden layer:

$$s_i = \sum_{u=1}^{P} \left| w_{iu}^{(in)} I(A_{iu} = 1) \right| + \sum_{vi}^{P} \left| w_{iu}^{(in)} I(A_{vi} = 1) \right| + \sum_{m=1}^{h_1} \left| w_{in}^{(1)} \right|, \ i = 1, \ldots, p,$$

Here, s_i is the significance score for features i; $w^{(in)}$ and $w^{(1)}$ denote the weights between the input and first hidden layers and weights between the first and the second hidden layers, respectively. The score

comprises the importance of rice (in rice-pathogen interaction) features according to the directed vertex connection in the feature graph and the feature's contribution according to the connection to the second hidden layer (Kong and Yu, 2018, 2020).

Significant network module identification

Markov cluster algorithm (MCL) (Enright et al., 2002) is very successful and can be used to discern quality functional modules in a biological network based on the transition probability manipulation or stochastic flows between graph nodes (Brohée and van Helden, 2006; Vlasblom and Wodak, 2009). The MCL algorithm with default parameters was used to cluster RiceONE in the present study, and network modules of more than two genes were retained for further analysis.

The GSEA algorithm (Mootha et al., 2003) with the custom scripts was used to classify network modules that were substantially enriched with high-scoring genes in the three classes (PAMPs, *Xanthomonas*, and *Magnaporthe*) (Figure S1). DLNet-based GSEA corresponding to the three categories was performed as follows: First, according to their DLNet-based gene ranking score, all (N) genes in RiceONE were ordered in decreasing order. Next, the summation from the gene with the maximal average score (g_1) to the gene with the minimal average score (g_N) was determined for a module with M genes (Figure S2). The summation across all genes is described as follows:

$$\sum_{i=1}^{j} X_i (i \leq j \leq N)$$

if g_i is a member of module M, then

$$X_i = \sqrt{\frac{N - M}{M}}$$

if g_i is not a member of module M, then

$$X_i = -\sqrt{\frac{M}{N - M}}$$

The module enrichment score was then calculated as:

$$\max_{i \leq j \leq N} \sum_{i=1}^{j} X_i$$

Gene set enrichment and functional classification

It is essential to know the statistical significance of the enrichment modules for each immune group. Accordingly, the class phenotype of the expression profile was randomly permuted, and 100 random DL-based models were generated. The above procedures for DL-based GSEA were replicated in each random model. Enrichment scores of all modules from the random DL models were amalgamated to form the background scores. Then, Student's *t*-test was performed to analyze the real enrichment scores as the proportion of background enrichment and modules with $p < 0.05$ were considered significant and retained for further study.

PlantGSEA (Yi et al., 2013) was used for detailed essential network module annotation. The modules were divided into five functional groups: defense response, signal transduction, transcriptional control, hormone response, and growth or development. The modules enriched with other functional elements were not considered for further analysis. Since more than one functional term may be enriched within a module, terms with the lowest p value were considered significant GO terms. A thorough detail of the related modules is given in Tables S9–S11.

Model evaluation and performance metrics

To evaluate the DLNet classification model, accuracy, sensitivity, and specificity parameters are considered. The area under the ROC curve (AUC) is also considered the evaluation parameter based on the true positive rate (TPR) and false positive rate (FPR).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Sensitivity } = \text{ True Positive Rate (TPR)} = \frac{TP}{TP + FN}$$

$$\text{Specificity } = \frac{TN}{TN + FP}$$

$$\text{False Positive Rate (FPR) } = \frac{FP}{TN + FP} = 1 - \text{Specificity}$$

where true positive (TP) refers to the model correctly predicting the positive class, false positive (FP) refers to the model incorrectly predicting the positive class, false negative (FN) refers to the model incorrectly predicting the negative class, and true negative (TN) refer to the model correctly predict the negative class.

DLNet-based rice model implementation, architecture, training and testing

The framework uses the expression profile of genes from two different conditions (treatment and control) to train a classification model to predict a class with a specific expression profile. The details information about the algorithm is provided in the supporting method. The DLNet rice model comprises a supervised random forest (RF) component to select useful features from raw data and a neural network component consisting of hidden layers (containing graph-embedded p features, 64, 16, and 2 neurons in each hidden layer, respectively) takes the generated feature graph and raw input to predict final outcomes. A dropout rate of 0.3 was used between the hidden layers to overcome overfitting. The directed network model was trained for 50 epochs using a batch size of 8 and a learning rate of 1×10^{-4}. The DL models for three different datasets were built and trained using the RF in the Scikit-learn (Breiman, 2001) package for supervised learning and a Tensorflow (Abadi et al., 2016) version 2.0 with adaptive moment estimation (Adam optimizer) algorithm for gradient descent that updates network weight iterative based in training data (Kingma and Ba, 2014). In the present study, 65% of each dataset (treatment + control) were randomly selected to train the model, and the remaining 35% were used to validate the model performance.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

To construct the immune related modular network, Student's t-test was performed between the original enrichment score and 100 randomly background enrichment scores. The background score was calculated by permutating class level of the expression data profiles. The modules having p value ≤ 0.05 were considered significant for further analysis.