Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A quick appraisal of Machine learning techniques in different aspects of rice cultivation

Abhishek Singh
asbhu2006@gmail.com

Banaras Hindu University

Nikhil Kumar Singh
Banaras Hindu University

Systematic Review

Keywords: machine learning, rice cultivation, comparative assessment

Posted Date: July 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9948244/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Rice is the most important cereal crop in the world, consumed by more than half of the global population. With the projected world population reaching around 9 billion by 2050, the cultivation of the rice crop becomes important for the food security of the world. Machine learning methods are playing a crucial role in optimising technologies related to rice cultivation. This paper is a quick appraisal of 343 research articles for the comparative utilisation of Machine Learning methods pertaining to different aspects of rice cultivation viz crop improvement, crop monitoring, crop mapping, irrigation management, weed management, disease management, etc. The study indicated that ML methods are used unevenly in different aspects of rice cultivation, where the aspects like disease management, yield forecasting, post-harvest, crop mapping, crop monitoring lead in the utilization of ML methods in comparison to the other aspects like weed management, irrigation management, fertilizer management, etc. both in terms of the quantity and quality of methods utilized.

1. Introduction

Rice (*Oryza sativa L.)* is one of the most important cereal crops as it is widely consumed by more than half of the world's population, utilising more than 12% of the world's arable land. It is one of the major staple foods that supplies almost 20% of the calories consumed globally (Kubo & Purevdorj, 2004). It is a water-intensive crop utilising between 24% to 30% of global fresh water resources (Bouman et al., 2007). In 2023/24, rice crop production was approximately 518 million tons (Childs & LeBeau, 2023). This is a staple crop that plays a crucial role in the food security of the world. It is an important commodity that provides financial security to the farmers as well. The role of rice cultivation in achieving United Nations Sustainable Development Goal 2 (SDG2) (Desa, 2016), which aims to achieve zero hunger, ensuring food security, improving nutrition and promoting sustainable agriculture, becomes very important. Rice cultivation has a significant role in reducing greenhouse gas emissions, controlling climate change and increasing water use efficiency. Asia is responsible for about 89.9% of world rice production (Faostat, n.d. (2020)). Top ten leading rice producers are located in Asia with China, India, Indonesia, Bangladesh, Vietnam and Japan accounting for 80% of the global production and consumption (Muthayya et al., 2014). It is estimated that by 2050 the world population will reach 9 billion and therefore there is a pressing need for increasing yields of major food crops amidst global population growth, decreasing agricultural land due to urbanization and industrialization and increasing uncertainties due to climate change.

Machine learning has emerged as the most prominent technology that is used to analyse big data generated in fields like agriculture. Leveraging the application of these techniques in rice cultivation can mitigate the effects of crucial challenges in rice cultivation. The utilization of advanced machine learning methods in paddy cultivation is very important because the conventional methods used in paddy cultivation are time consuming, less precise and prone to human errors. Moreover, due to climate change and other challenges which are dynamic and enormous, prompt, timely and appropriate remedial measures can effectively counter their adverse effect. Due to advancement in the computing and storage technologies the machine learning method can analyze the big data generated in smart paddy cultivation. The use of satellites, unmanned aerial vehicles (UAV), multispectral instruments, drones, IOT has increased the velocity, veracity, volume and variety of the data generated which can be analysed only with the use of sophisticated machine learning methods.

Recent advancements in the field of machine learning are being used, but their application is not distributed uniformly over all aspects of rice cultivation. The major divisions in which rice cultivation can be divided are crop improvement, weed management, insect and disease management, irrigation management, fertiliser management, crop monitoring, crop mapping, checking greenhouse gas emission, etc. Even a slight modification in the techniques and technologies used in these aspects can have a considerable effect on rice cultivation. There are many reviews in journals which deals with certain specific aspects of rice cultivation like (De Clercq & Mahdi, 2025) for yield prediction, (Sharma et al., 2024,) (Ismail et al., 2025) (Simhadri et al., 2025) for detection and identification of diseases, (Islam et al., 2025b) for rice classification,(Fang et al., 2024) for rice mapping,(Padhiary et al., 2024) for enhancing precision agriculture, but there is no review for comparative assessment of utilization of machine learning methods for different aspects of rice cultivation. The main aim of this paper is to compare the utilisation of machine learning techniques with respect to different aspects of rice cultivation. It has been found that research in some aspects involves sophisticated use of ML methods, but some aspects are lagging in the use of ML and Artificial Intelligence. The prime purpose of this paper is to present important aspects of rice cultivation with respect to the utilisation of machine learning methods to researchers and different stakeholders. This review will further apprise the researchers working in different fields related to rice cultivation with the recent developments, making a comparative assessment and identifying the gaps in their particular field with respect to the novel methods of machine learning. It will encourage researcher to explore new dimensions in their particular area of specialisation with respect to rice cultivation.

2. Methodology

A search of articles was conducted in the Science Direct and Google Scholar databases using precise and detailed search queries. The queries were concentrated on the titles containing the terms "(rice or paddy) AND (Machine Learning OR Artificial Intelligence). The queries were not very specific in order to select all the papers involving the use of machine learning techniques on all aspects of rice cultivation. The search was restricted to articles published from the years 2020 to 2025 in order to involve the recent advancements in the application of machine learning techniques in rice cultivation. This was done in order to conduct a quick appraisal as the field is developing at a fast pace, and also to investigate the recent advances. These articles were sorted by relevance, and the top 600 articles were included in the study. This was done in order to avoid an unmanageable number of articles and also to focus on the most recent developments and applications of machine learning in rice cultivation, as well as to maintain relevance and rigour. After removing the duplicates in the list 411 articles were selected for the study.

These 411 articles were then classified in one of the 17 defined categories after studying the papers deeply with respect to the abstracts, conclusion and methodology rejecting book chapters, papers in conference proceedings, and selecting papers dealing mainly with rice crops, finally 343 articles were included in the study (Fig. 1.)

These categories were based on the different aspects of rice cultivation like Crop improvement (CI), Crop mapping (CM), Crop Monitoring, Contamination, Disease Management (DM), Data Set (DS), Insects Management, Fertilizer Management (FM), Farm Machinery (FMa), Green House Gas (GHG), Irrigation Management (IM), Post-Harvest (PH), Reviews (R), Seed Science (SS), Weed Management (WM),Yield Forecasting (YF) and rejected papers.

For the category of Crop improvement those papers were selected which specifically discussed the selection of genotypes/accessions for various characters like flowering time, nitrogen use efficiency, phenological dates, genome wide scanning, genotypic and phenotypic selections, heading dates, identification of saline alkali tolerant varieties etc. For Crop mapping those papers were selected which dealt with the mapping of rice crop over geographical regions which included terms like rice planting area mapping, rice agricultural system mapping, irrigated rice field mapping etc. Under the category of contamination there were papers which predicted and estimated the contamination of rice crop with various metals like cadmium, arsenic, selenium and other heavy metals. Crop monitoring dealt with methods which monitored the rice crop over a period of time like estimating the rice canopy, rice lodging assessment, heading date prediction, Dry biomass estimation, Simulating rice growth, rice phenological stage recognition, rice panicle detection etc. Disease management included papers which used ML methods in disease prediction like rice blast disease identification, brown spot disease, pre-symptomatic diagnosis, rice disease classification etc. Papers which dealt with the management of rice pests like rice plant hopper, Hispa pest, counting of rice plant hopper, life cycles of rice pests, prediction and monitoring of rice pest populations etc. were included in the category of insect management. Under the fertilizer management category those papers were included which predicted, estimated the level of nitrogen, nitrogen deficiency, nutrient deficiency diagnosis, predicting soil organic content, NPK stress, rice nutrient level at the panicle initiation stage, nutrition diagnosis etc. Farm machinery category included paper which dealt with the use of ML methods in operating different machines in rice cultivation like that of rice trans planter, control of unmanned rice harvester, seeding accuracy of pneumatic seed metering device etc. Under the category of the greenhouse gases those papers were included that dealt with the estimation, prediction of paddy rice methane emissions, detection of rice straw burning, mapping of daily methane concentrations in rice paddies etc. Irrigation Management included papers dealing with accurate water level monitoring, estimating water status of rice, water stress evaluation, water productivity, estimating rice water footprint, soil moisture simulation, evapo-transpiration etc. The post-harvest category included papers dealing with the application of the ML in determining the origin and variety of individual rice grains, precision estimation of rice grain moisture content, identification of broken rice grains, predicting content of gluten, fatty acid, amylose, amylopectin etc. Seed Science category included papers methods for identifying rice seed purity, rice variety classification and identification, viability and vigor of naturally aged rice seeds etc. Weed management category included papers which applies ML methods for distinguishing between weeds and rice seedlings, rice field weed sensing image segmentation etc.

These articles were then analysed deeply, and data were extracted on different variables like study information (publication year, Journal, title) and methodology (model type used) for all the 411 articles, which were classified into 17 categories. The information generated was then consolidated in a master sheet and was further analysed and visualised using Excel. An analysis was then done to identify common patterns among different categories about modelling approaches, their current state and gaps in the existing research and future scopes. The titles of all the papers were downloaded in an Excel sheet along with their year of publication, journal name, software used and model used. Then the papers were classified on the basis of year of publication and journals. All the models were segregated and written in different columns, and then the final Excel sheet was prepared, having three columns: one the name of the model and a 2nd column was the categories, and a third column was the subcategory. Then, by using pivot tables, different 2 by 2 tables were produced to investigate the extent of use of different kinds of machine learning models in different categories, in which rice cultivation was segregated.

3. Results and discussion

The total 411 articles were categorized into 17 categories (Fig. 2). Maximum number of articles were of the category yield forecasting (YF-56) followed by Disease management (DM–49), Post-Harvest (PH-42), Crop Monitoring – 34, Fertilizer management (FM-28), Crop Improvement (CI-27), Contamination (26), Crop Mapping (CM-18), Irrigation Management (IM-13), Farm Machinery (FMa-12), Green House Gases (GHG-11), Seed Science- 11, Insect Management – 8, Weed Management-8. Apart from these 9 datasets, 20 Reviews and 39 rejected were not included in the study. At this point, it is important to note that out of 20 Review papers, 9 review papers were for Disease Management, 6 for Crop Improvement, 1 for crop mapping, 1 for yield forecasting, 1 for post-harvest,1 for greenhouse gases and 1 for farm machinery. This indicates the imbalance in the application and usage of machine learning methods in different aspects of rice cultivation. Finally, 343 articles were considered for this study.

These 343 articles were published in various journals pertaining to different aspects of rice cultivation. The journals that have published more than 4 papers are shown in Fig. 3

The journal having the maximum publications was Computers and Electronics in Agriculture having 69 articles, followed by Smart Agriculture Technology (16), then Field Crop Research (13), Science of The Total Environment (13), European Journal of Agronomy (12), Remote Sensing Applications: Society and Environment (9), Plant Phenomics (7), ISPRS Journal of Photogrammetry and Remote Sensing(7), Ecological Informatics (7), Agricultural and Forest Meteorology (6), Journal of Hazardous Material (5), Journal of Agricultural and Food Research (5), Infrared Physics and Technology (5). Maximum number of papers were from Asia (China, India, Taiwan, Bangladesh, Korea, Indonesia, Thailand, Japan, Malaysia, Vietnam), Very few from Middle East (Iran, Turkey, Saudi Arabia,), from Europe (Germany, Netherland, Belgium Spain.), from Africa (Tanzania, Sierra lone), South America (Brazil)), Australia and USA. Of all the papers majority of the papers were from China (90%) followed by India (5%), followed by other south east countries and very few from Europe, Australia and USA.

The machine learning methods used in these 343 studies were then segregated, isolated and classified into 11 classes namely CNN, Decision Tree (DT), Deep Neural Network (DNN), Ensemble, K-Nearest Neighbours (KNN), Neural Networks (NN), Regression, Random Forest (RF), Recurrent Neural Network (RNN), Support Vector Machine (SVM) and Others. A two-way cross-classification table (Fig. 4) was prepared to show the category-wise distribution of these models. The term CNN was used for Convolutional Neural Networks, which use specialised convolutional and pooling layers to classify images and objects. Deep Neural Network (DNN) is a broader category that encompasses CNN, but here they were used for those deep neural networks that were not used for

image or object classification. The term Neural Network (NN) was used for networks that were not deep, and the Recurrent Neural Network was used for networks that specialise in time series data. Similarly, Random Forest (RF) was indicated differently from ensemble because of the widespread application of this technique in the research articles. Overall distribution of machine learning methods over all the categories is depicted in Fig. 5.

3.1 Convolutional Neural Networks

CNN was the most prominent tool among all the others used for analysing the data in almost all the categories except in the case of Yield forecasting, Irrigation management, fertiliser management and Contamination (Table 1). This indicates the uneven use of this technique in research related to different aspects of the rice cultivation. This could be due to the nature of the input variables being generated or may be due to the lack of computational as well as domain knowledge among the researchers. Overall, 303 times the term CNN was used in all the papers. Among different categories, CNN was used maximum in disease management, followed by post-harvest and then crop monitoring and crop improvement. The minimum use of CNN was in contamination, followed by irrigation management and greenhouse gases.

Table 1
Types of CNN used in different categories related to rice cultivation

CI	**CNN,CNN-1D, Inception V3, Res2Net50 ResNet (34,50)**
CM	DeepLab (V3,V3+), ETUnet, FAPNET, Fr-Net, HR.Net, PLANET, RAUNetpan, ResUnet, RRCNet, XM-UNet, UNet
Contamination	ERT
Crop Monitoring	AAUConvNeXt, ConvNeXt, ConvNeXtBase,ConvUNeXt, CO-ResNetRS50-SSL, DenseNet121, EfficientDet, EfficientNet, EfficientNet v2, Faster R-CNN, FasterNet-T1, HRNet, IDeepLabV3, LW-(Segnet,Unet), MI-UConvNeX ResNet50, ResNet-50 + Yolov5, ResNetv2-LSTM algorithm, RiceRes2Net, ShuffleNetV2, U + LSTM-F, U2net,UNet, VGG16, YOLOv5
DM	EfficientNet(B0,B3,B6,B7,V2B3,V2S),ResNET(101,152V2,18,34,50),VGG(16,19),MoblieNet(V2,V3),Inception(ResnetV2,V3),DenseNet(201,12 AlexNet, CNN with (BiGRU),FSLNet,resNext101,RiceNet,YOLOv5,YoloX,CNN-1D
FM	AlexNet, CNN-(1D,3D), CNN-LSTM, DenseNet, DS-CNN, GoogleNet, Inception V3, ND-CNN, ResNet101 + LSTM, VGG16
FMa	DeepLabv3 + with ResNet, Deeplabv3+, ERT, ESPNetv2, Fast-SCNN, FCN, IERT, improved YOLOv5n, MobileNetV3, PIDNet, PSPNet, RSHRNe SCTNet, ShuffleNetV2,UNet, YOLOv5 7.0 , YOLOv5-Lite neural network, YOLOv5n (RP-YOLO
GHG	CNN-1D, Faster R-CNN, FR-Net, Mask R-CNN, YOLOv5, YOLOv8
IM	ConvNeXtBase, EfficientNetB0, Inception V3, ResNet50, VGG16, CBAM
Insect Management	CBAM, DC-GhostNet, Inception V3, ResNet50, SAFPN,VGG16, VGG19, YOLOv8
PH	AlexNet, CBAM-ParNeXt (V1,V2), CNN-AdaBoost, ConvNeXt, DenseNet (201,121), EfficientNet (B0,V2L),ELSSR, GhostNet, GMMNet, Googl HDNet, HINet, Inception V3, LTRN, MIRNet, MobileNet(V2,V3), MPRNet, NAM-EfficientNetV2, NASNetLarge, Raman-CNN (RC), RepVggNet_ ResNet-(152,152V2,50,50V2), ResNeXt50, ShuffleNetV2_1.0,VGG (16,19), Xception
Seed Science	DCGAN,DenseNet, EfficientNet, GAN, Inception, PCA-CNN, ResNet, ResNet50, SAM-GAN, VGG, VGG 16, WGAN-GP, ClusterRiceNet
WM	DeepLabV3, MobileNetV2, SegNet, UNet, YOLOv5, YOLOV5 CS-
YF	1D-CNN + LSTM, AlexNet, CNN-(1D,2D,3D), Faster R-CNN,ResNet50, TCNN, VGG16, Xception, YOLOv3, YOLOv5, YOLOv8

While using the conventional ML methods for classification purposes, the investigator has to follow certain steps in series, like cleaning of the data, extracting important input variables, selecting relevant input variables, learning the pattern from the data and then classification. The advantage of CNN over the conventional methods is that the investigator doesn't have to worry about the various sequential steps needed before classification, as CNN can detect significant input variables without any human intervention. Utilising this important advantage of the CNN, this technique was comprehensively used in almost all aspects of rice cultivation, though the appropriate application varied from category to category. Starting from Alexnet (Krizhevsky et al., 2017) arranged in their evolutionary order, ZefNet (Zeiler & Fergus, 2014) ,Visual geometry group (VGG) (Simonyan & Zisserman, 2015), GoogleNet (also called Inception-V1) ,Highway network (Srivastava et al., 2015), ResNet, Inception: ResNet and Inception-V3/4, ResNext, WideResNet, (Zagoruyko & Komodakis, 2017) Xception ,DenseNet, MobileNet, ShuffleNet, EfficientNet, Residual attention neural network, Convolutional block attention module, High-resolution network (HRNet) and YOLO were used. The use of the most prominent and advanced CNN models was in the field of DM, Crop monitoring, FMa and PH while the other categories like crop improvement, crop mapping, contamination, FM, IM and Insect management utilizing less advanced models on average. One of the hindrances and blocks encountered in the use of CNN models is that they require huge data; this problem can be neutralised by using transfer learning and data augmentation techniques. Other problems are Imbalanced data, Interpretability of data, Uncertainty scaling, Catastrophic forgetting, Model compression, Overfitting, vanishing gradient problem, exploding gradient problem and under-specification. An excellent review of CNN architecture, concepts, etc. is given by (Alzubaidi et al., 2021). Some of the novel application of CNN was observed in crop improvement category very it was fused with GNN (Cai et al., 2022a), in disease management (Lu et al., 2023), in post-harvest (He et al., 2025) etc.

3.2 Random Forest

The RF(Breiman, 2001) was used 144 times in the papers and was the second most used Machine learning technique in the papers and most of the time it was used as a benchmark for comparison with other models. The use of RF was almost uniformly distributed over all the categories with the minimum use in Insect Management, followed by Farm machinery and weed management. The category of yield forecasting utilised the technique of Random Forest the most, followed by fertiliser management and then post-harvest, crop monitoring and contamination. Some of the novel application of Random Forest was

observed in the crop improvement category, where Random Forest plus kinship was used to increase the classification ability of the model (Liu et al., 2024). Similarly, network-guided forest where the network information is summarized into a sampling probability of predictor variables which is further used in the construction of the RF was used in (Kumar et al., 2022), conditional random forest, which is built upon the implementation of conditional inference trees, and is used for variable selection was used by (Garnaik et al., 2022). Random forest (RF) model combined with recursive feature elimination (RFE) was used in(Ding et al., 2025), The Hybrid RF-XGB model performed better than the individual model

3.3 Support Vector Machine

Support Vector Machines, though discovered in 1960's by Vapnik and Chervonenkisin came into prominence in the 1990s after the Kernel trick was discovered and the use of SVM for regression problems apart from classification was initiated. This contributed to the widespread use of this technique in all fields of science. There are two main problems associated with the application of SVM: one is the choice of the kernel for each application, and the second is the availability of huge data sets. It becomes extremely difficult to train an SVM using a dataset having a very large number of vectors. This may restrict the use of SVM in studies having massive datasets. Refining training datasets is again a research problem that needs comprehensive research. (Nalepa & Kawulok, 2019) .The SVM models were used 138 times in the papers. Their maximum use was in PH, followed by YF, Crop Monitoring, DM, FM and Contamination. The categories of WM, Insect Management, FMa, GHG, IM, and CM have less utilised this technique. Some of the novel application of this methods are discussed here- (W. Wang et al., 2023) used LS-SVM which is a modified version of SVM which reduces the computational costs of traditional SVM,(Asiri, 2023) used HOG-SVM, **t**he feature representation method Histogram of Oriented Gradients (HOG) for feature representation and refining training data set for proper implementation of SVM technique. (S. Wang et al., 2025) used the criteria importance through inter criteria correlation (CRITIC) weighting method, along with the Gaussian membership function (GMF), to process features The processed features were utilized to train, calibrate, and validate a model based on Support Vector Machine (SVM) technology, culminating in the development of the CRITIC-GMF-SVM framework for the precise detection of rice blast during its symptomless stage.(Tian,L. *et al*., 2021) also used feature selection strategy which is machine learning based sequential floating forward selection (ML-SFFS) methodology for determining the optimal feature combination (OFC) and overall accuracy (OA) in the detection of RLB at various infection stages. Compared to the use of all DSSFs with a SVM classifier, the SVM-based SFFS (SVM-SFFS) algorithm prevailed in the classification accuracy up to 10% over the sampling period. (Zhao et al., 2023) compared several common meta-heuristic optimization algorithms including the Genetic Algorithm (GA), Particle Swarm Algorithm (PSA), Sparrow Search Algorithm (SSA) Slime Mold algorithm (SMA), Sine Cosine Algorithm (SCA) and Artificial Bee Colony algorithm to optimize the penalty kernel parameters of SVMs for the prediction of unknown datasets, SSA-optimized SVMs have better performance than others.(Liu,R. *et al*., 2022) used Dimension Reduction by Recursive Feature Elimination (RFE) to develop a SVM + RFE, which was found to be the best among the candidate models.

3.4 Ensembling- A prominent problem in machine learning is the overfitting in the training set and the lack of generalisation ability for out-of-set data. Ensemble learning emerged as a solution to this, as it utilises multiple base learners to increase the generalisation ability of the system. On the other hand, costs for training and testing with multiple base learners act as a major hindrance to its application. An important review of ensembling methods is available in (Yang,Y. *et al*., 2023), they divided the ensemble methods into three categories: Traditional ensemble learning (TEL), Usual ensemble deep learning (UEDL) and faster ensemble deep learning(FEDL).TEL consists of feature extraction, generating base learners and integrating base learners. Well known learning strategies for base learners at data level were Bagging (Breiman, 1996), Dagging (Ting & Witten, 1997), Weighting data points (Boosting) (Freund & Schapire, 1997), Wagging (Bauer & Kohavi, 1999), Multiboosting (Webb, 2000), Manipulating at feature level - Random Subspace (Ho, 1998), Manipulating at both data and feature levels - Random Forests (Breiman 2001), Rotation Forests (Rodriguez et al., 2006) and others like Negative correlation learning (Liu, Y. *et al.*, 1999), ANN-based random initialization (Hansen and Salamon 1990). Various well-known ensembling criteria (EC) for forming ensemble learner were Weighting methods (Majority voting / averaged summarization) (Mendes-Moreira et al., 2012), Minimizing the objective (Mao et al., 2015), Meta-learning methods (Stacking (Wolpert, 1992), Mixture of expert (Masoudnia & Ebrahimpour, 2014), Weighting-based learning (Omari & Figueiras-Vidal, 2015) and Ensemble selection (Zhou et al., 2002) methods (many could be better than all), (Clustering based criteria (Bakker & Heskes, 2003) Ranking-based criteria (Martinez-Munoz et al., 2009), Selection based criteria (Hernández-Lobato, 2009). Base learners can be generated in two ways: using different types of learning algorithms to generate 'heterogeneous' base learners, and the other is to use the same learning algorithm to generate 'homogeneous' base learners. In usual ensemble deep learning, all three steps, extracting features, generating base learners and forming learners are based on deep learning, which results in an end-to-end framework usually involving massive training data. Some of the drawbacks of these approaches are that with the increase in the number of base learners, the training and testing of these base learners become expensive, especially when deep learning is used at every stage. Fast ensemble deep learning reduces the time, costs and space overheads of UEDL. These approaches were used in rice cultivation for different aspects, but their distribution in different categories was not uniform. Ensembling was used 115 times, but the maximum use of the ensemble methods was in YF, Crop Monitoring, PH, DM, Contamination and Seed Science. The rest of the categories underutilised this technique (Table 2). Some of the novel application of ensembling was in crop monitoring ((P. Zhang et al., 2024),Yield foracasting (Sarkar et al., 2024a) and post harvest (Azmi et al., 2024) etc.

Table 2
Types of ensembling methods used in different categories related to rice cultivation

CI	**AdaBoost, BRT(Boosted Regression Technique), LightGBM, XGBOOST**
CM	Bagging, GBT
Contamination	CatBoost, GBM, GBRT, XGBOOST
Crop Monitoring	Extra Trees (ET),GB, GBDT, GBM, LightGBM, XGBOOST
DM	AdaBoost, Extra Trees (ET), GBM,GBR, HGB classifier, LightGBM, XGBOOST
FM	AdaBoost, cubist, Gradient Boosting Decision Tree (GBDT), XGBOOST
FMa	XGBOOST
GHG	AdaBoost, Gradient Boosting Machine (GBM), LightGBM, XGBOOST
IM	
Insect Management	AdaBoost
PH	AdaBoost, Ensemble-based Rice Moisture Content and Localisation (eRMCL) algorithm, Extra Trees (ET),GB, GBC, LightGBM, XGBOOST
Seed Science	CatBoost, Extra Trees (ET), Gradient Boosting Decision Tree (GBDT), XGB-MLP, XGBOOST
WM	Ensemble Classifier (EC), Extra Trees (ET)
YF	AdaBoost, Bagging Tree (Treebag), cubist, Ensemble bagging, ensemble techniques based on Bayesian model averaging (BMA),Extra Trees (ET), GB,GBM,gradient boosting regressor (GBR),LightGBM, XGBOOST

3.5 Regression- Regression was used 112 times, and the maximum use of it was in YF followed by Crop Monitoring and then FM and PH, CI, Seed Science and very sparsely in other categories (Table 3). Some of the techniques, like multivariate adaptive regression splines(MARS), which is non non-parametric technique to model nonlinear relationship, were sparsely used. kernel partial least squares (Lee et al., 2025), Elastic Net (Chen et al., 2025) (enet) (Gaussian process regression (Wickramasinghe et al., 2021), Huber Regression (De Clercq & Mahdi, 2024)

Table 3
Types of regression used in different categories related to rice cultivation

CI	**LASSO, Linear Regression, Logistic Regression, LSR, MLR, PLSR**
CM	
Contamination	LASSO, Linear Regression, Logistic Regression, PLSR, Ridge
Crop Monitoring	Elastic Net (enet), GLM, LASSO, Logistic Regression, MLR, multivariate adaptive regression splines (MARS), PLSR, Polynomial regression, Ridge Regression
DM	Bagging Regressor, kernel partial least squares, KPLS
FM	LASSO, Linear Regression, MLR, multi-period simple regression (SR), PLSR, Quadratic Curve Regression (QCR), Simple Linear Regression, stepwise multiple linear regression (SMLR)
FMa	Elastic Net (enet)
GHG	
IM	Additive regression (AR)
Insect Management	Logistic Regression
PH	Gaussian process regression (GPR, Williams, 1998), Linear Regression, Logistic Regression, MLR, PLSR, Ridge Classifier
Seed Science	Linear Regression, Logistic Regression, PCA-LR, PLS-DA
WM	Linear Regression, Logistic Regression
YF	Elastic Net (enet), gaussian process regression, Generalized Additive Models, (GAM), Generalized Linear Models (GLM), Huber regression, LASSO ,Linear Regression, Linear regression with stepwise selection (LRSS) Logistic Regression ,MLER algorithm, MLR ,MLR –STEPWISE, multivariate adaptive regression splines (MARS), Ordinary regression with LASSO penalty (rqlasso), PLSR, power regression (PR), Ridge regression algorithm, robust regression (RR

3.6 Neural Network

Some of the other neural networks, aside from deep neural networks, were used in some of the studies. NN were utilised 82 times with the maximal use in YF followed by PH, contamination and others (Table 4). Some of the novel applications of NN were – GNN (Cai et al., 2022b),, ELM (W. Wang et al., 2023), MTOC-NN, (Yamaguchi & Katsura, 2024), probabilistic neural network (PNN) (Liu et al., 2021), ANFIS (Sarkar et al., 2024b) .

Table 4
Types of NN used in different categories related to rice cultivation

CI	**ANN, BPNN, GNN, MLP**
CM	ANN,
Contamination	ANN, extreme learning machine (ELM), GA-BP, PSO-BP
Crop Monitoring	BPNN, extreme learning machine (ELM), MLP, MTOC-NN, MTO-NN, NNR
DM	NNET, MLP, MS-STNet model, NN, probabilistic neural network (PNN)
FM	ANN, BPNN, extreme learning machine (ELM), Multilayer Perceptron (MLP), NN, TFPN
FMa	
GHG	ANN, BPNN, FFNN
IM	
Insect Management	ANNX, FFNN
PH	ANN, ANNX, BPNN, extreme learning machine (ELM), hybrid principal component analysis-neural network (PC-NN), MLP, Multi-Layer Perceptron Classifier, PC-NN
Seed Science	MLP
WM	hybrid neural network-particle swarm optimization algorithm (NN-PSO), hybrid NN-ICA, MLP
YF	adaptive neuro FIS (ANFIS), ANN, BPNN, FFNN, MLP, NN, single-layer perceptron feedforward neural networks (NNET)

3.7 K Nearest Neighbours (KNN)- was used 40 times and PH was utilizing at the most number of times and the categories of GHG and IM not using these techniques even once. Some of the novel applications were stabilized nearest neighbour, SNN (Lee et al., 2025)

3.8 Decision tree was used 33 times and almost all categories used it uniformly with PH utilizing it for the most number of times and FMa, GHG not utilizing them even once. Some of the novel applications were random tree (RT) and reduced error pruning tree (REPT) (Mokhtar et al., 2021), decision tree algorithms Quinlan's algorithm (M5P) and REPTree (ReepT) (de Oliveira Carneiro et al., 2023), CART(Muradi et al., 2024) and M5 model tree (M5 Tree) (Sarkar et al., 2024b).

3.9 Recurrent Neural Network (RNN) - were used 25 times and in only four categories YF, DM, CM and FM. This could be due to the scarce utilization of time series data in different aspects of rice cultivation. Some of the novel application of RNN were- Bidirectional LSTM (BiLSTM) (Tian,S. *et al*., 2024) BiLSTM DB(Hegde et al., 2025), Elman recurrent neural network (Elman RNN) (L.-W. Liu et al., 2021), DL–LSTM (Cao et al., 2021), IndRNN (Chu & Yu, 2020)

3.10 Others -Some of the novel applications which were used 24 times in different aspects of rice cultivation are DNABert Algorithm (Sheng et al., 2025) ,BLS (Broad Learning Systems) ((Jin et al., 2024), BSKD (Ramachandran & KS, 2024) ,GP-Radial (Elangovan et al., 2023) Prospect- ((Yu et al., 2024)) GMF model (Wang,S. *et al*., 2025), Maxent model (Sathiyaseelan et al., 2025). memory-based learning (MBL) (Dai et al., 2024), Gene Expression Programming (GEP) ((Shiri et al., 2024), PROSAIL model (Y. Zhang et al., 2025), Passive Aggressive Classifier - (Islam et al., 2025a), SHapley Additive exPlanations (SHAP) (Zhu et al., 2025), hybrid 'RaNN' model (Lingwal et al., 2024), Explainable AI (XAI) + LIME (Local Interpretable Model-Agnostic Explanations) (Yenkikar et al., 2025), Knowledge-informed machine learning (KIML)(Wang,J. *et al.*, 2025)

3.11 DNN apart from CNN and others NN was used 17 times. Especially in DM, YF and Crop Monitoring in contamination, FMa, GHG, IM and Insect management it was not used even once.

The use of various ML methods in a particular category of rice cultivation are shown in Fig. 6.

Whereas, the earlier discussion was focussed on the results of different types of ML methods used in different categories of rice cultivation it was evident that the number of techniques will vary on the basis of the papers attributed to that category. Therefore, in order to identify the usage and impact of each ML method on a particular category, the number of time the method was used was divided by the number of papers in that category to emphasise the method used per paper in that category. The results are shown in Fig. 7.

Crop Improvement

In the 27 studies pertaining to Crop improvement the most frequently used technique was **CNN (1.11) followed by RF and SVM (0.41)**, Regression (0.30), Ensemble (0.26), KNN and NN (0.15), **b** (0.07), DNN and Other (0.04). RNN was not used even once in the crop improvement studies.

Crop Mapping (CM)

In the 18 studies pertaining to crop mapping the most frequently used technique was **CNN (0.89) followed by RF (0.33) and SVM (0.28)** and then RNN (0.17), DT, Ensemble, NN, DNN all (0.11) and regression not used at all.

Contamination

In the 26 studies pertaining to contamination the most frequently used technique was **RF (0.58) followed by NN (0.42), SVM (0.38**), Ensemble (0.35), Regression (0.27), DT (0.12), KNN (0.08), CNN (0.04) and DNN, Others and RNN not used at all

Crop Monitoring

In the 34 studies pertaining to crop monitoring the most frequently used technique was **CNN (0.88) followed by Regression (0.56), RF (0.44)**, SVM (0.41), Ensemble (0.38), Other (0.21), NN (0.21), KNN (0.09), DNN (0.09), DT (0.03) and then RNN not used at all.

Disease Management (DM)

In the 49 studies pertaining to Disease management the most frequently used technique was **CNN (1.73) followed by SVM (0.24), Ensemble (0.20)**, RF (0.18), NN (0.14), KNN (0.08), DNN (0.08), RNN (0.08) Other (0.04), DT (0.04), Regression (0.04).

Fertilizer Management (FM)

In the 28 studies pertaining to fertilizer management the most frequently used technique was **RF (0.68), Regression (0.64), CNN (0.46)**, SVM (0.39), Ensemble (0.25), NN (0.21), KNN (0.07), DT (0.04), DNN (0.04), Other (0.04) and RNN (0.04).

Farm Machinery (FMa)

In the 12 studies pertaining to farm machinery the most frequently used technique was **CNN (1.75), SVM (0.25), KNN (0.17)**, Ensemble (0.08), Regression (0.08), RF (0.08) and RNN, Other, DT, DNN, NN not used at all.

Green House Gases (GHG)

In the 11 studies pertaining to greenhouse gases the most frequently used technique was **RF (0.73), CNN (0.55), Ensemble (0.45)**, SVM (0.27), NN (0.27) and DT, DNN, KNN, Other, Regression and RNN not used at all.

Irrigation Management (IM)

In the 13 studies pertaining to irrigation management the most frequently used technique was **CNN (0.46), SVM (0.38), RF (0.31)**, Other (0.23), DT (0.15), Regression (0.08) and DNN, Ensemble, KNN, NN, RNN not used at all.

Insect Management

In the 08 studies pertaining to insect management the most frequently used technique was **CNN (1.00), NN (0.38), SVM (0.25)**, Ensemble (0.25), DT (0.13), Regression (0.13), KNN (0.13) and DNN, Other, RNN, RF not used at all.

Post-Harvest (PH)

In the 42 studies pertaining to Post Harvest the most frequently used technique was **CNN (1.07), SVM (0.69), RF (0.36)**, NN (0.36), Regression (0.31), Ensemble (0.29), KNN (0.29), DT (0.19), Other (0.07), DNN (0.05) and RNN not used at all.

Seed Science

In the 11 studies pertaining to seed science the most frequently used technique was **CNN (1.45), SVM (0.82), Regression (0.73)**, Ensemble (0.64), RF (0.55), KNN (0.45), DT (0.27), DNN (0.09), NN (0.09), Other (0.09) and RNN was not used at all.

Weed Management (WM)

In the 08 studies pertaining to weed management the most frequently used technique was **CNN (0.88), NN (0.38), Ensemble (0.25)**, KNN (0.25), Regression (0.25), SVM (0.25), DT (0.13), RF (0.13) and other, DNN, RNN not used at all.

Yield Forecasting (YF)

In the 56 studies pertaining to yield forecasting the most frequently used technique was **Ensemble (0.68), RF (0.61), Regression (0.57)**, SVM (0.39), NN (0.36), CNN (0.34), RNN (0.30), DT (0.13), Other (0.11), DNN (0.05), KNN (0.04),

The type of models used for a study depends upon the type of data used in that particular study. The data used in the reviewed papers can be divided into two broader categories measured/ sensed data and remotely sensed data. The remotely sensed data can further be divided into the categories the data generated through satellites or through unmanned aerial vehicle. Majority of the data sets used in the different categories were huge and utilized remotely sensed data most of the times followed by the synthesis of remotely sensed and measured/sensed data. There were very few studies which utilized the time series data for their analysis. Therefore, it is evident from the results that CNN was the most prevalent ML method used which was used in 10 of the 14 categories most number of times per paper, RF in 3 categories and Ensemble in only one category. CNN was used most prominently in farm machinery (1.75) followed by disease management (1.73), seed science (1.45), crop improvement (1.11), post-harvest (1.07), insect management (1.00), crop mapping (0.89), crop monitoring (0.88), weed management (0.88) greenhouse gases (0.55), fertilizer management (0.46), irrigation management (0.46), yield forecasting (0.34) and contamination (0.04). when we analyse the number of papers related to AI and machine learning in rice cultivation we find that yield forecasting-56, disease management-49, post-harvest – 42, crop monitoring – 34 were the leading areas where ML methods were quantitatively better than those of the other fields. In field like fertilizer management – 28, crop improvement-27, contamination-26 and crop mapping – 18 lesser use of these methods was visible. Some of the aspects of rice cultivation, like irrigation management-13, farm machinery-12, greenhouse gases-11, seed science-11, insect management-8 and weed management-8 utilised these methods less in terms of quantity. Though the category of yield forecasting has maximum number of

papers (56) they utilized the CNN second least number of times (0.34) with respect to other categories which indicates at the scope of utilizing this method for better forecasting of yields of rice. Similarly, ensembling has been used most number of times (0.68) in yield forecasting but this technique which enhances the generalizing ability of a model is used comparatively less in all the categories, though hybrid models and other combinations has been used in some studies. This review indicated that the SVM and RF are the most utilized models of all the studies.

4. Conclusions

The study was planned to assess the usage of sophisticated machine learning methods in different aspects of rice cultivation. 411 research articles were classified into 17 categories and after removing 68 articles due to various reasons, the remaining 343 articles related to different aspects of rice cultivation (14 categories) were analysed both qualitatively and quantitatively for the usage of ML methods. After the assessment, it was found that the application of ML methods is not distributed uniformly over all the categories, both in terms of the quality as well as quantity of ML methods used. This pointed towards a huge gap in the application of ML methods in different aspects of rice cultivation. yield forecasting-56, disease management-49, post-harvest – 42, crop monitoring – 34 were the leading areas where ML methods were quantitatively better than those of the other fields. In field like fertilizer management – 28, crop improvement-27, contamination-26 and crop mapping – 18 lesser use of these methods was visible. Some of the aspects of rice cultivation, like irrigation management-13, farm machinery-12, greenhouse gases-11, seed science-11, insect management-8 and weed management-8 utilised these methods less in terms of quantity. It is implicated by the study that there is sub optimal use of these methods in categories related to the management of the rice crop like irrigation, weed (Effective weed management can enhance rice productivity under aerobic conditions by 27–300%(Hussain et al., 2008)), insect (pre harvest losses due to insect ranges from 25–43%.), fertilizer and use of farm machinery. All these aspects of rice cultivation are hugely important and enhancing the use of machine learning and Artificial intelligence in these categories will optimize the rice production leading to sustainable and increased production. If the quality of ML techniques is concerned, though the category of Yield forecasting has maximum number of papers (56) they utilized the CNN second least number of times (0.34) with respect to other categories which indicates at the scope of utilizing this method for better forecasting of yields of rice. This indicates, scope of utilization of this method in yield forecasting because CNN have advantage over other machine learning methods in advanced image analysis which is due to the ability of the CNN for feature extraction and its unique mapped layer structure,(Dey et al., 2022). Similarly, ensembling has been used most number of times (0.68) in yield forecasting but this technique which enhances the generalizing ability of a model is used comparatively less in all the categories, though hybrid models and other combinations has been used in some studies. The irregular use of these techniques in research related to different aspects of the rice cultivation could be due to the nature of the input variables being generated or may be due to the lack of computational facilities as well as domain knowledge among the researchers. One of the hindrances and blocks encountered in the use of these models is that they require huge data; this problem can be neutralised by using transfer learning and data augmentation techniques. This review also indicated that the SVM and RF are the most utilized models across all the studies. Moreover, this study utilised only research articles from Science Direct and Google Scholar; the inclusion of articles from other databases like Scopus and Web of Science will definitely increase the validity and scope of this kind of study. Also, in future studies inclusion of more statistical information like that of the kinds of dependent and independent variables used in the research articles, the software used, and comparative assessment of ML methods used in terms of their accuracy and validity (meta-analysis), will further add to the application of ML methods in rice cultivation. Overall all this quick appraisal of ML methods used in 343 recent studies does provide a rapid assessment of ML methods used in different aspects of rice cultivation and points out the research gaps prevailing in this field.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The author declares no conflict of interest.

Competing interests

The authors declare no competing interests

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Chat GPT in order to improve the readability and language of the manuscript. After using this tool/ service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data Availability statement

Not Applicable

Author Contributions

Abhishek Singh: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data Collection, Data curation, Formal analysis Conceptualization. **Nikhil Kumar Singh**: Writing – review & editing, Data curation.

Author Information

Abhishek Singh: Professor (Agricultural Statistics); Department of Agricultural Engineering, Institute of Agricultural Sciences, Banaras Hindu University

Nikhil Kumar Singh: Assistant Professor (Agronomy), Department of Agronomy, Institute of Agricultural Sciences, Banaras Hindu University

Corresponding author:

Correspondence to Abhishek Singh (asbhu2006@gmail.com)

References

1. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Farhan L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8(1):53. https://doi.org/10.1186/s40537-021-00444-8.
2. Asiri Y. Unmanned aerial vehicles assisted rice seedling detection using shark smell optimization with deep learning model. Phys Communication. 2023;59:102079.
3. Azmi N, Kamarudin LM, Yeon A, Zakaria AS, Syed Zakaria A, Nishizaki SMM, Mohamed H, Mao L, X., Rahiman F, M. H. Integration of dual band radio waves and ensemble-based approach for rice moisture content determination and localisation. J Stored Prod Res. 2024;108:102399. https://doi.org/10.1016/j.jspr.2024.102399.
4. Bakker B, Heskes T. Clustering ensembles of neural network models. Neural Netw. 2003;16(2):261–9.
5. Bauer E, Kohavi R. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. Mach Learn. 1999;36(1–2):105–39. https://doi.org/10.1023/A:1007515423169.
6. Bouman BA, Humphreys E, Tuong TP, Barker R. Rice and water. Adv Agron. 2007;92:187–237.
7. Breiman L. Bagging predictors. Mach Learn. 1996;24(2):123–40. https://doi.org/10.1007/BF00058655.
8. Breiman L. Random Forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.
9. Cai J, Xiao G, Su R. GC6mA-Pred: A deep learning approach to identify DNA N6-methyladenine sites in the rice genome. Methods. 2022a;204:14–21. https://doi.org/10.1016/j.ymeth.2022.02.001.
10. Cai J, Xiao G, Su R. GC6mA-Pred: A deep learning approach to identify DNA N6-methyladenine sites in the rice genome. Methods. 2022b;204:14–21.
11. Cao J, Zhang Z, Tao F, Zhang L, Luo Y, Zhang J, Han J, Xie J. Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. Agric For Meteorol. 2021;297:108275.
12. Chen Z, Cai Y, Liu Y, Liang Z, Chen H, Ma R, Qi L. Towards end-to-end rice row detection in paddy fields exploiting two-pathway instance segmentation. Comput Electron Agric. 2025;231:109963.
13. Childs N, LeBeau B. (2023). *Rice Outlook: October 2023 (Report No. RCS-23I): Technical Report*.
14. Chu Z, Yu J. An end-to-end model for rice yield prediction using deep learning fusion. Comput Electron Agric. 2020;174:105471.
15. Dai L, Xue J, Lu R, Wang Z, Chen Z, Yu Q, Shi Z, Chen S. In-situ prediction of soil organic carbon contents in wheat-rice rotation fields via visible near-infrared spectroscopy. Soil Environ Health. 2024;2(4):100113.
16. De Clercq D, Mahdi A. Feasibility of machine learning-based rice yield prediction in India at the district level using climate reanalysis and remote sensing data. Agric Syst. 2024;220:104099.
17. De Clercq D, Mahdi A. Modern computational approaches for rice yield prediction: A systematic review of statistical and machine learning-based methods. Comput Electron Agric. 2025;231:109852. https://doi.org/10.1016/j.compag.2024.109852.
18. de Oliveira Carneiro L, Coradi PC, Rodrigues DM, Lima RE, Teodoro LPR, de Moraes S, Teodoro R, Nunes PE, Leal MT, M. M., Lopes LR. Characterizing and predicting the quality of milled rice grains using machine learning models. AgriEngineering. 2023;5(3):1196–215.
19. Desa UN. (2016). *Transforming our world: The 2030 agenda for sustainable development*. https://www.sidalc.net/search/Record/oai:wedocs.unep.org:20.500.11822-11125/Description
20. Dey B, Masum Ul Haque M, Khatun R, Ahmed R. Comparative performance of four CNN-based deep learning variants in detecting Hispa pest, two fungal diseases, and NPK deficiency symptoms of rice (*Oryza sativa*). Comput Electron Agric. 2022;202:107340. https://doi.org/10.1016/j.compag.2022.107340.
21. Ding Q, Li Y, Cao R, Chen J, Yao X, Zhang W. Spatial optimization of rice systems with ratoon rice increases production and reduces methane emissions. Eur J Agron. 2025;170:127720.
22. Elangovan A, Duc NT, Raju D, Kumar S, Singh B, Vishwakarma C, Krishnan G, Ellur S, Dalal RK, M., Swain P. Imaging sensor-based high-throughput measurement of biomass using machine learning models in rice. Agriculture. 2023;13(4):852.
23. Fang H, Liang S, Chen Y, Ma H, Li W, He T, Tian F, Zhang F. A comprehensive review of rice mapping from satellite data: Algorithms, product characteristics and consistency assessment. Sci Remote Sens. 2024;10:100172. https://doi.org/10.1016/j.srs.2024.100172.
24. Faostat FA. O. (n.d.). *Statistical Databases. 2020*.
25. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. 1997;55(1):119–39.

26. Garnaik S, Samant PK, Mandal M, Mohanty TR, Dwibedi SK, Patra RK, Mohapatra KK, Wanjari RH, Sethi D, Sena DR. Untangling the effect of soil quality on rice productivity under a 16-years long-term fertilizer experiment using conditional random forest. Comput Electron Agric. 2022;197:106965.
27. He K, Tian J, Xia Y, Zhou Y, Hu X, Xie L, Yang H, Huang Y, Huang D. Detection of the amylose and amylopectin contents of rice by hyperspectral imaging combined with a CNN-AdaBoost model. J Food Compos Anal. 2025;142:107468. https://doi.org/10.1016/j.jfca.2025.107468.
28. Hegde A, Umesh P, Tahiliani MP. Automated rice mapping using multitemporal Sentinel-1 SAR imagery using dynamic threshold and slope-based index methods. Remote Sens Applications: Soc Environ. 2025;37:101410.
29. Hernández-Lobato D. (2009). Prediction based on averages over automatically induced learners: Ensemble methods and bayesian techniques. *Ph. D. Dissertation, UNIVERSIDAD AUTONOMA de MADRID*. http://arantxa.ii.uam.es/~dhernan/docs/Thesis_color_links.pdf
30. Ho TK. The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell. 1998;20(8):832–44.
31. Hussain S, Ramzan M, Akhter M, Aslam M. Weed management in direct seeded rice. J Anim Plant Sci. 2008;18(2–3):86–8.
32. Islam MM, Himel GMS, Moazzam G, Uddin MS. Precision in Rice Variety Classification using Stacking-based Ensemble Learning. J Cereal Sci. 2025a;122:104128.
33. Islam MM, Himel GMS, Moazzam MG, Uddin MS. Artificial Intelligence-based Rice Variety Classification: A State-of-the-art Review and Future Directions. Smart Agricultural Technol. 2025b;10:100788. https://doi.org/10.1016/j.atech.2025.100788.
34. Ismail UI, Chua HN, Nordin R, Ahmed MK. A comprehensive review of deep learning approaches for rice disease detection: Datasets, methodologies, and future directions. Smart Agricultural Technol. 2025;11:100976. https://doi.org/10.1016/j.atech.2025.100976.
35. Jin Z, Liu H, Cao H, Li S, Yu F, Xu T. Hyperspectral Remote Sensing Estimation of Rice Canopy LAI and LCC by UAV Coupled RTM and Machine Learning. Agriculture. 2024;15(1):11.
36. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90. https://doi.org/10.1145/3065386.
37. Kubo M, Purevdorj M. The future of rice production and consumption. J Food Distribution Res. 2004;35(1):128–42.
38. Kumar R, Khatri A, Acharya V. (2022). Deep learning uncovers distinct behavior of rice network to pathogens response. *Iscience*, *25*(7). https://www.cell.com/iscience/fulltext/S2589-0042(22)00818-5
39. Lee H, Park Y, Kim G, Lee JH. Pre-symptomatic diagnosis of rice blast and brown spot diseases using chlorophyll fluorescence imaging. Plant Phenomics. 2025;7(1):100012.
40. Lingwal S, Bhatia KK, Singh M. A novel machine learning approach for rice yield estimation. J Exp Theor Artif Intell. 2024;36(3):337–56. https://doi.org/10.1080/0952813X.2022.2062458.
41. Liu L-W, Hsieh S-H, Lin S-J, Wang Y-M, Lin W-S. Rice blast (Magnaporthe oryzae) occurrence prediction and the key factor sensitivity analysis by machine learning. Agronomy. 2021;11(4):771.
42. Liu Q, Zuo S, Peng S, Zhang H, Peng Y, Li W, Xiong Y, Lin R, Feng Z, Li H. Development of machine learning methods for accurate prediction of plant disease resistance. Engineering. 2024;40:100–10.
43. Liu R, Tan F, Wang Y, Ma B, Yuan M, Wang L, Zhao X. Machine learning identification of Saline-Alkali-Tolerant Japonica rice varieties based on Raman spectroscopy and Python visual analysis. Agriculture. 2022;12(7):1048.
44. Liu Y, Yao X. Ensemble learning via negative correlation. Neural Netw. 1999;12(10):1399–404.
45. Lu Y, Wu X, Liu P, Li H, Liu W. Rice disease identification method based on improved CNN-BiGRU. Artif Intell Agric. 2023;9:100–9. https://doi.org/10.1016/j.aiia.2023.08.005.
46. Mao S, Jiao L, Xiong L, Gou S, Chen B, Yeung S-K. Weighted classifier ensemble based on quadratic form. Pattern Recogn. 2015;48(5):1688–706.
47. Martinez-Munoz G, Larios N, Mortensen E, Zhang W, Yamamuro A, Paasch R, Payet N, Lytle D, Shapiro L, Todorovic S. (2009). Dictionary-free categorization of very similar objects via stacked evidence trees. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 549–556. https://ieeexplore.ieee.org/abstract/document/5206574/
48. Masoudnia S, Ebrahimpour R. Mixture of experts: A literature survey. Artif Intell Rev. 2014;42(2):275–93. https://doi.org/10.1007/s10462-012-9338-y.
49. Mendes-Moreira J, Soares C, Jorge AM, Sousa JFD. Ensemble approaches for regression: A survey. ACM-CSUR. 2012;45(1):1–40. https://doi.org/10.1145/2379776.2379786.
50. Mokhtar A, Elbeltagi A, Maroufpoor S, Azad N, He H, Alsafadi K, Gyasi-Agyei Y, He W. Estimation of the rice water footprint based on machine learning algorithms. Comput Electron Agric. 2021;191:106501.
51. Muradi H, Domiri DD, Parsa IM, Yoga IK, Bustamam A, Rarasati A, Harini S, Manalu RJ, Subehi M. Rice Phenology Classification Model Based on Sentinel-1 Using Machine Learning Method on Google Earth Engine. Can J Remote Sens. 2024;50(1):2368036. https://doi.org/10.1080/07038992.2024.2368036.
52. Muthayya S, Sugimoto JD, Montgomery S, Maberly GF. An overview of global rice production, supply, trade, and consumption. Ann N Y Acad Sci. 2014;1324(1):7–14. https://doi.org/10.1111/nyas.12540.
53. Nalepa J, Kawulok M. Selecting training sets for support vector machines: A review. Artif Intell Rev. 2019;52(2):857–900. https://doi.org/10.1007/s10462-017-9611-1.
54. Omari A, Figueiras-Vidal AR. Post-aggregation of classifier ensembles. Inform Fusion. 2015;26:96–102.
55. Padhiary M, Saha D, Kumar R, Sethi LN, Kumar A. Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicle for farm automation. Smart Agricultural Technol. 2024;8:100483. https://doi.org/10.1016/j.atech.2024.100483.

56. Ramachandran A, KS SK. (2024). Border Sensitive Knowledge Distillation for Rice Panicle Detection in UAV Images. Computers Mater Continua, *81*(1). https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope= site&authtype=crawler&jrnl=1546 2218&AN=180260312&h=IRMNq1 %2F5zden7mdb3BPLZ%2FtaaWwpiBWfDY1W 9cKtJu0KXYSCx82o6jFkpSMtDxKbdXgqb%2FvN NH%2BGIsFMtR8OcQ%3D%3D&crl=c
57. Rodriguez JJ, Kunch eva LI, Alonso CJ. Rotation forest: A new cl assifier ensemble method. IEEE Trans Patter n Anal Mach Intell. 2006;28(10):1619&ndas h;30.
58. Sarkar TK, Roy DK, Kang YS, Jun SR, Park JW, Ryu CS. Ensemble of Machine Learning Algorithms for Rice Grain Yield Prediction Using UAV-Based Remote Sensing. J Biosystems Eng. 2024a;49(1):1–19. https://doi.org/10.1007/s42853-023-00209-6.
59. Sarkar TK, Roy DK, Kang YS, Jun SR, Park JW, Ryu CS. Ensemble of Machine Learning Algorithms for Rice Grain Yield Prediction Using UAV-Based Remote Sensing. J Biosystems Eng. 2024b;49(1):1–19. https://doi.org/10.1007/s42853-023-00209-6.
60. Sathiyaseelan K, Das B, Ladhalakshmi D, Venu E, Bashyal BM, Bag TK, Sinha P. (2025). Climatic drivers for assessment of false smut risk in rice ecosystems: A guide for planning effective management trials. *Heliyon*, *11*(4). https://www.cell.com/heliyon/fulltext/S2405-8440(25)00908-9
61. Sharma M, Kumar CJ, Bhattacharyya DK. Machine/deep learning techniques for disease and nutrient deficiency disorder diagnosis in rice crops: A systematic review. Biosyst Eng. 2024;244:77–92. https://doi.org/10.1016/j.biosystemseng.2024.05.014.
62. Sheng J, Zhao E, Zhu Y, Dai Y, Zhang B, Qin Q, Zhang H. A Novel Method for Mining Regulatory sRNAs Related to Rice Resistance Against Blast Fungus from Multi-Omics Data. Curr Bioinform. 2025;20(7):581–94. https://doi.org/10.2174/0115748936305102240705052723.
63. Shiri J, Kazemi MH, Karimi S, Cufí S, de Cartagena FR, Pinsach J, Arbat G. Chronological assessment of heuristic data driven approaches for soil water content simulation in subsurface drip irrigated rice. Sci Total Environ. 2024;955:177193.
64. Simhadri CG, Kondaveeti HK, Vatsavayi VK, Mitra A, Ananthachari P. Deep learning for rice leaf disease detection: A systematic literature review on emerging trends, methodologies and techniques. Inform Process Agric. 2025;12(2):151–68. https://doi.org/10.1016/j.inpa.2024.04.006.
65. Simonyan K, Zisserman A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. https://doi.org/10.48550/arXiv.1409.1556
66. Srivastava RK, Greff K, Schmidhuber J. (2015). *Highway Networks* (arXiv:1505.00387). arXiv. https://doi.org/10.48550/arXiv.1505.00387
67. Tian L, Xue B, Wang Z, Li D, Yao X, Cao Q, Zhu Y, Cao W, Cheng T. Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection. Remote Sens Environ. 2021;257:112350.
68. Tian S, Sheng Q, Cui H, Zhang G, Li J, Wang B, Xie Z. Rice recognition from Sentinel-1 SLC SAR data based on progressive feature screening and fusion. Int J Appl Earth Obs Geoinf. 2024;134:104196.
69. Ting KM, Witten IH. (1997). *Stacking bagged and dagged models*. https://researchcommons.waikato.ac.nz/items/37f19edd-e874-4b2c-a26b-0b4a202ca555
70. Wang J, Feng M, Guo S, Li H, Han Y, Chen J. A knowledge-informed machine learning (KIML) approach to disentangle single-and double-season rice yields and attainable yield gaps in the mixed cropping systems. Comput Electron Agric. 2025;237:110616.
71. Wang S, Li T, Wang Y, Chen L, Jiang F, Zhang X, Wei M, Chen S, Xu L, Yang N. MOS sensor array based on multi-modal data weighted composite membership optimization for rice blast detection in symptomless stage. Comput Electron Agric. 2025;232:110153.
72. Wang W, Man Z, Li X, Chen R, You Z, Pan T, Dai X, Xiao H, Liu F. Response mechanism and rapid detection of phenotypic information in rice root under heavy metal stress. J Hazard Mater. 2023;449:131010.
73. Webb GI. MultiBoosting: A Technique for Combining Boosting and Wagging. Mach Learn. 2000;40(2):159–96. https://doi.org/10.1023/A:1007659514849.
74. Wickramasinghe L, Weliwatta R, Ekanayake P, Jayasinghe J. (2021). Modeling the Relationship between Rice Yield and Climate Variables Using Statistical and Machine Learning Techniques. *Journal of Mathematics*, *2021*, 1–9. https://doi.org/10.1155/2021/6646126
75. Wolpert DH. Stacked generalization. Neural Netw. 1992;5(2):241–59.
76. Yamaguchi T, Katsura K. A novel neural network model to achieve generality for diverse morphologies and crop science interpretability in rice biomass estimation. Comput Electron Agric. 2024;218:108653.
77. Yang Y, Lv H, Chen N. A Survey on ensemble learning under the era of deep learning. Artif Intell Rev. 2023;56(6):5545–89. https://doi.org/10.1007/s10462-022-10283-5.
78. Yenkikar A, Mishra VP, Bali M, Ara T. (2025). An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction. MethodsX, 103442.
79. Yu F, Bai J, Fang J, Guo S, Zhu S, Xu T. Integration of a parameter combination discriminator improves the accuracy of chlorophyll inversion from spectral imaging of rice. Agric Commun. 2024;2(3):100055.
80. Zagoruyko S, Komodakis N. (2017). *Wide Residual Networks* (arXiv:1605.07146). arXiv. https://doi.org/10.48550/arXiv.1605.07146
81. Zeiler MD, Fergus R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars, editors, *Computer Vision – ECCV 2014* (Vol. 8689, pp. 818–833). Springer International Publishing. https://doi.org/10.1007/978-3-319-10590-1_53
82. Zhang P, Zhang S, Wang J, Sun X. Identifying rice lodging based on semantic segmentation architecture optimization with UAV remote sensing imaging. Comput Electron Agric. 2024;227:109570. https://doi.org/10.1016/j.compag.2024.109570.
83. Zhang Y, Jin X, Shi L, Wang Y, Qiao H, Zha Y. A hybrid method for water stress evaluation of rice with the radiative transfer model and multidimensional imaging. Plant Phenomics. 2025;7(1):100016.
84. Zhao L, Qing S, Bai J, Hao H, Li H, Shi Y, Xing X, Yang R. A hybrid optimized model for predicting evapotranspiration in early and late rice based on a categorical regression tree combination of key influencing factors. Comput Electron Agric. 2023;211:108031.

85. Zhou Z-H, Wu J, Tang W. Ensembling neural networks: Many could be better than all. Artif Intell. 2002;137(1–2):239–63.
86. Zhu Q, Gao Y, Yang B, Zhao K, Wang Z, Huang F, Cheng F, Zhao Q, Huang J. Advanced data-driven interpretable analysis for predicting resistant starch content in rice using NIR spectroscopy. Food Chem. 2025;486:144311.

Figures

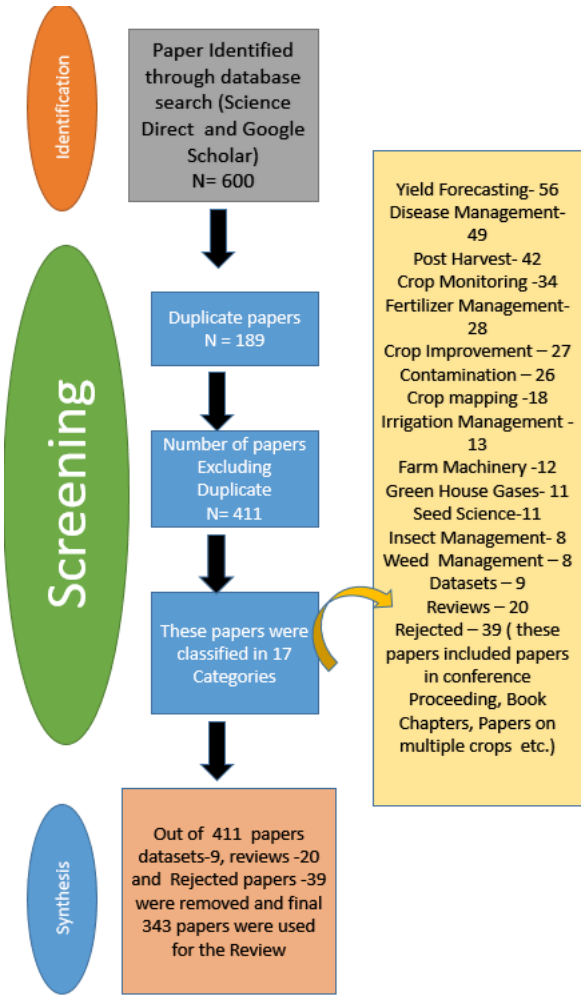


Figure 1

Flow diagram of the Database Search

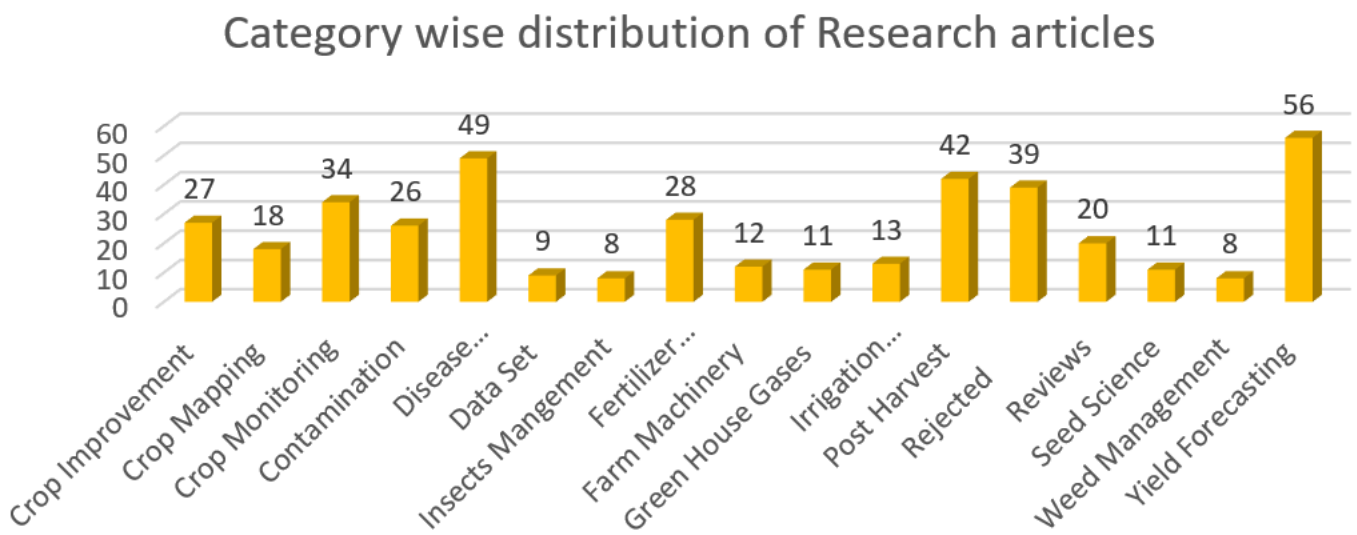


Figure 2

Distribution of research articles in different categories pertaining to rice cultivation

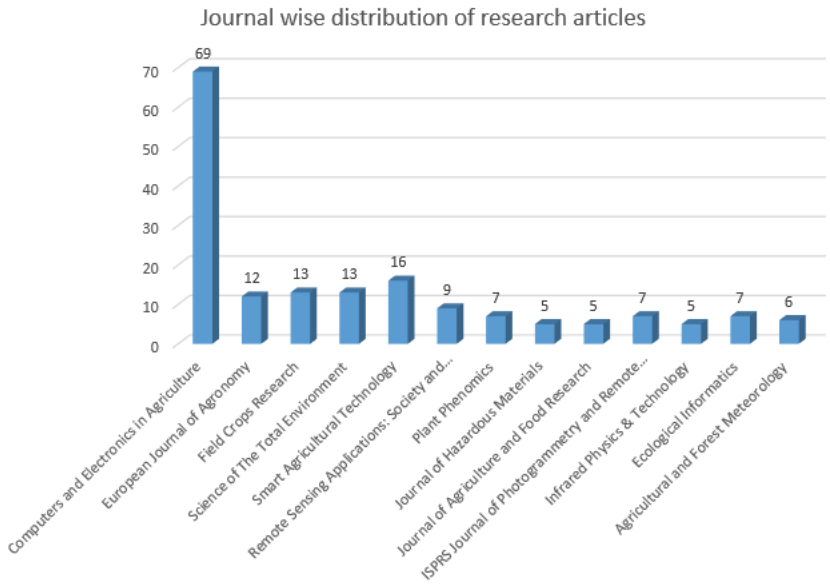


Figure 3

Journal-wise distribution of the research articles. Only journals publishing more than 4 research articles are displayed.

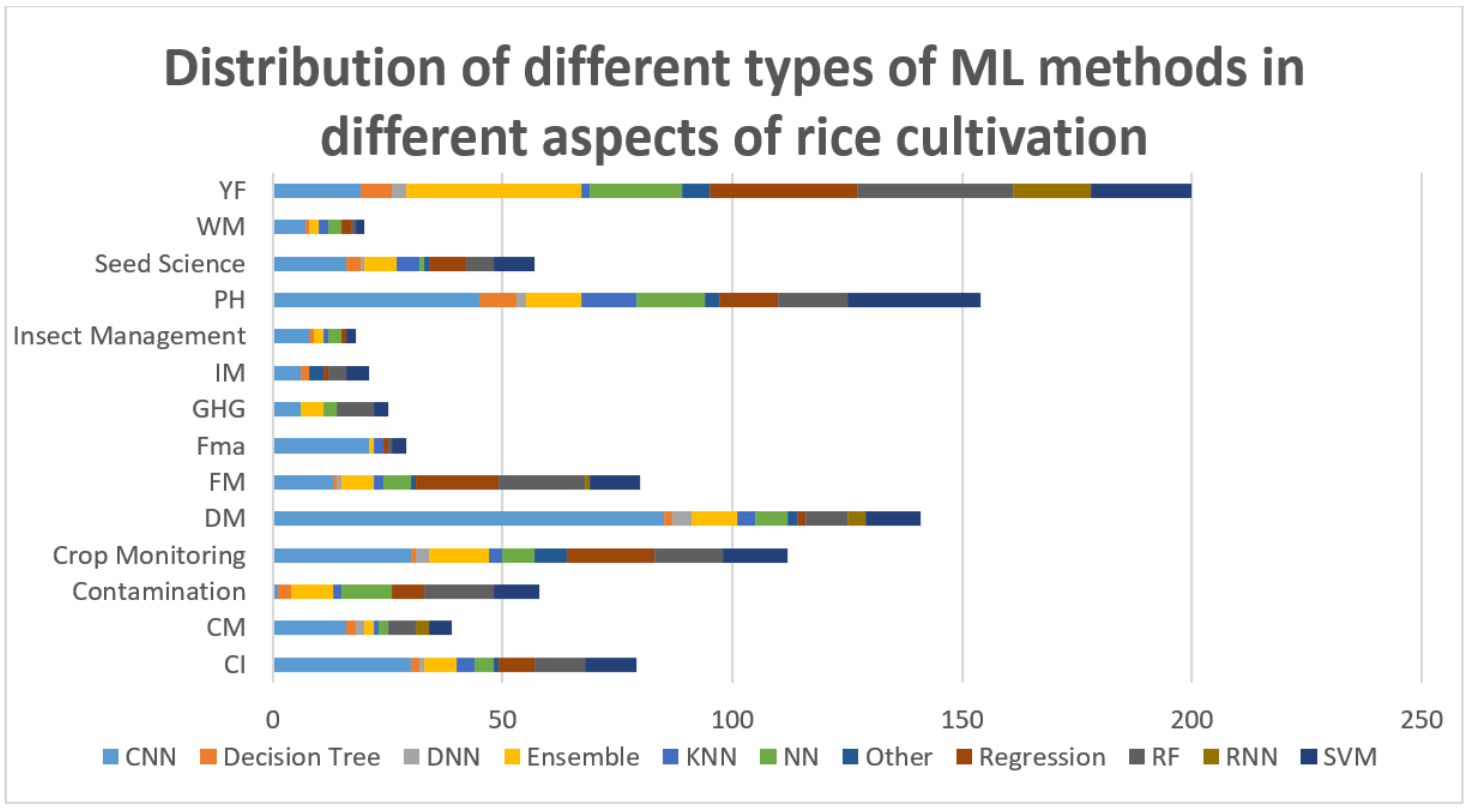


Figure 4

Distribution of different types of ML methods in different aspects of rice cultivation

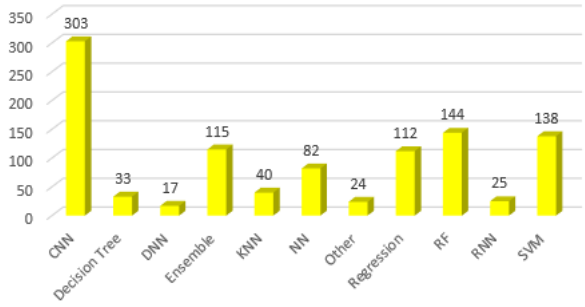


Figure 5

The use of various ML methods in the research papers.

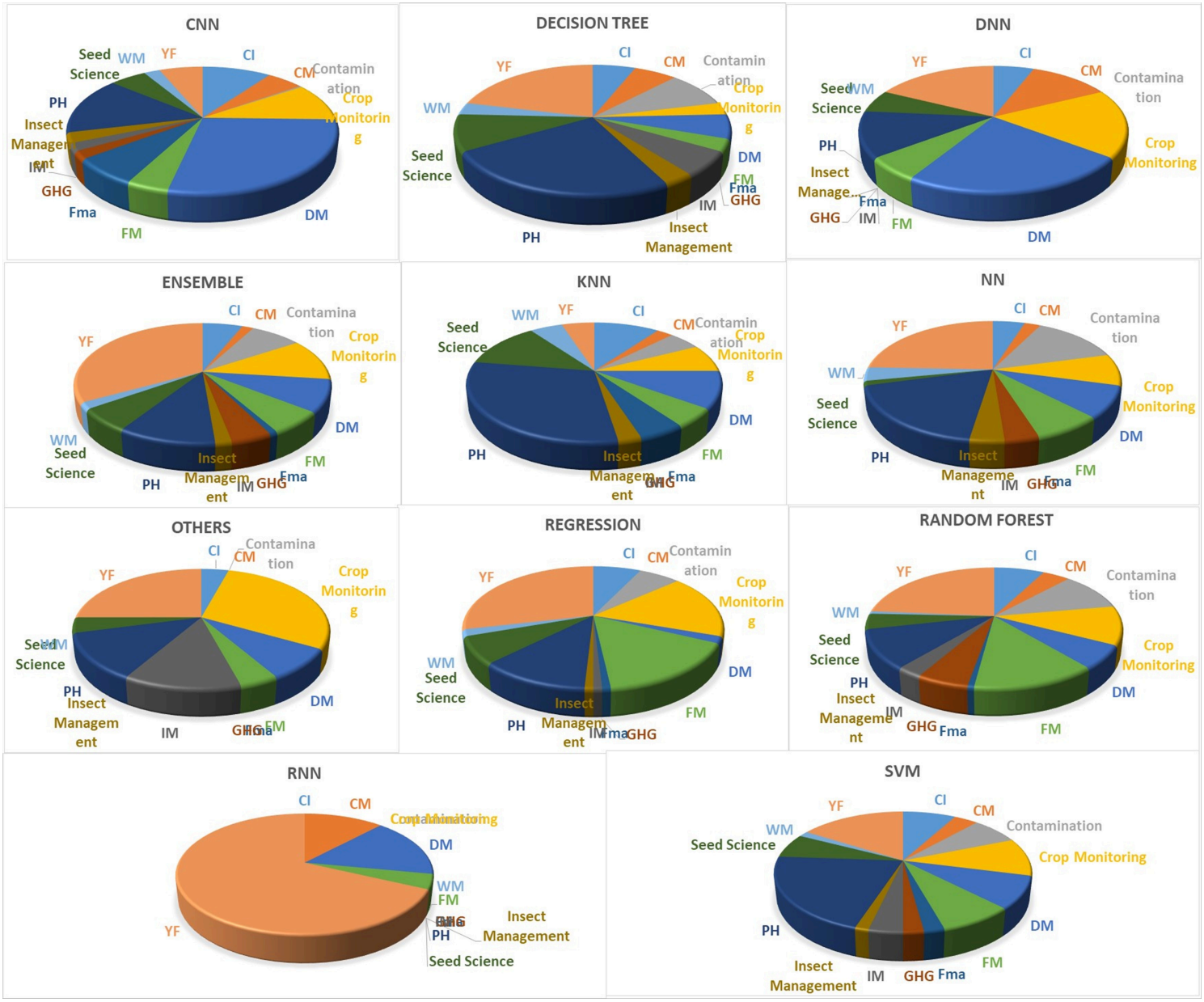


Figure 6

Extent of usage of different ML methods in different aspects of rice cultivation.

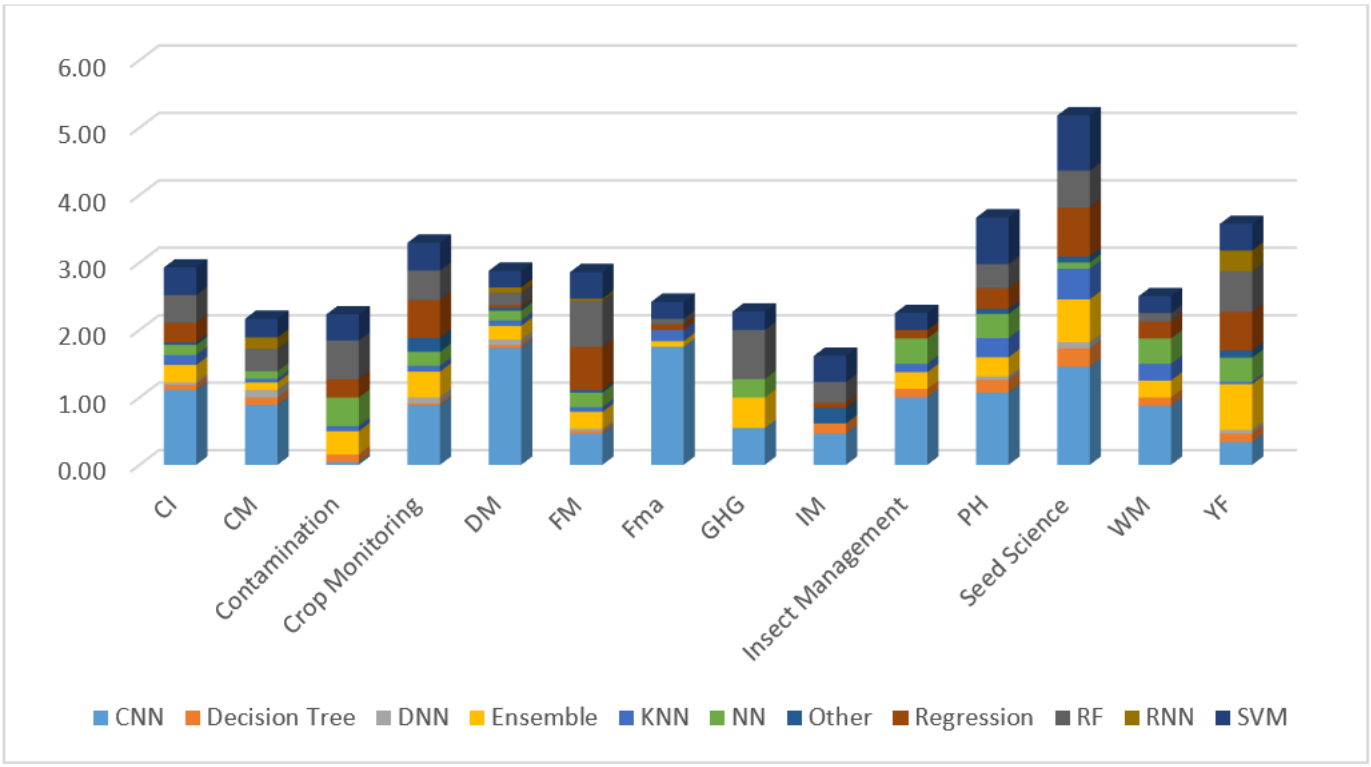


Figure 7

No. of times ML methods used in different categories per paper