Rice Crop Disease Prediction Using Machine Learning Technique

Bharati Patel, National Institute of Technology, Raipur, India
Aakanksha Sharaff, National Institute of Technology, Raipur, India

ABSTRACT

Crop yields are affected at large scale due to the spread of unchecked diseases. The spread of these diseases is similar to the spreading of cancer in the human body. But, unlike cancer, these diseases can be identified at early stages through plant phenotyping traits analysis. In order to effectively identify these diseases, effective segmentation, feature extraction, feature selection, and classification processes must be followed. Selection of the best combination for the given methods is very complex due to the presence of a large number of the aforementioned methods. Thereby, disease prediction models are generally found to be ineffective. This paper proposes a highly effective machine learning-based formulation approach to select a proper classification process that improves the overall accuracy of crop disease detection with different dimensionalities of plant datasets and includes maximum features. Hence, the proposed adaptive learning algorithm gives 99.2% accuracy compared to other techniques like back-propagation neural network (BPNN), convolutional neural network (CNN), and SVM.

KEYWORDS

Adaptive Feature Selection, Disease Prediction, Extensive Learning, Intensive Learning, Machine Learning Classification

INTRODUCTION

The agriculture field contains so many challenging applications for yield prediction. In Chhattisgarh, farmers mostly depend upon rice production for their livelihoods and it is advantageous over the Nation also. Therefore, production growth should be proper even with some uncertainties like weather disturbance, pests’ attack, virus attacks, bacterial and fungal attacks etc. The selection of a method for profitable production growth requires plant phenotyping traits analysis. Recently, applied various popular phenotypic traits also have significant limitations for cost, performance, space and time coverage (Li et al., 2020). One of the other limitations for learning patterns from the plant spectra is a crucial task for predictive plant phenotyping applications discussed by Rehman et al. (2020). Some other challenging applications using machine learning approaches has listed as: digitized regularization, smart farming using IOT, irrigation method, precision agriculture using Artificial Intelligence, and market prediction analysis discussed by Sharaff and Choudhary (2018) etc. These traits having major utility for the reduction of the complex genome executions over real time data analysis in the field of agriculture.

The main contribution of the proposed work includes:

DOI: 10.4018/IJAIEIS.20211001.oa5

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.
1. Early disease detection for high yield production.
2. Encouragement towards the interdisciplinary approach for smart farming by using computer vision techniques.
3. Field farming needs digitization/computer-aided techniques, so that performance rate gets increases. It avoids the complexity of manual calculation, time management, hardware management, and so on.
4. The proposed work improves the phenotyping traits analysis over complexity of genomic selection. Traits analysis (Tong and Nikoloski, 2021) of plant has multiple tasks, but it is exclusively an abstract structure over the genotyping to produce fruitful results by applying it.
5. It gives the vision towards the heterogeneous environment. Prediction over the environmental uncertainties like rainfall prediction, temperature management (field conditions are heterogeneous by nature that will not give identical results), weather forecasting and avoiding global warming problem over the world.
6. Management of Eco-system is having all the above dependencies over smart farming or forest-scale phenotyping (Bombrun et al., 2020) etc. Smart farming requires the computer vision techniques to limit the error rate calculation and maximize the early prediction rate.
7. Utility of machine learning algorithms make a trend for working efficiently in the field of agriculture.

Figure 1 represents the significance of plant phenotyping over genotyping data analysis. Genotyping creates a complexity to analyze the plant features for regular analysis. In genome data analysis, GWAS relates the gene functionality of plant data. In contrast, plant phenotyping data analysis describes with some of the plant feature analysis such as height, leaf disease, biomass, LAI (leaf area index), CCC (canopy chlorophyll content), CT (computed tomography) (Liu et al., 2017; Zhang et al., 2018), color index based segmentation and so on. Plant phenotyping traits analysis contains identification or prediction of crop diseases with multi-domain image processing task.

In this paper, every descended section is distributed for their respective purposes. According to the second module, literature survey has been discussed, which gives knowledge about the
latest concerning area and problem statement for the future work. The third module describes the methodology and process of the proposed work with the enhanced result analysis. The fourth module has an experimental setup to process step-wise execution of all the learning phases. The next section relates tabular representation of the result analysis and the last section concluded this proposed work, which indicates utility towards yield prediction by applying machine learning techniques over the real time dataset.

LITERATURE REVIEW

In this module, machine learning techniques and various technologies have been discussed briefly. Plant disease detection relates to multiple factors for production growth. The need of various parameters is incorporated the production growth of rice crop listed with some specifications such as:

Platform Dependent Machine Learning (ML) Approach for Disease Detection

Detection of plant-based diseases has become very complex due to the wide variety of data-types (Image dataset) available for processing. Plant Phenotyping includes maximum features but It is not limited to top view images, frontal images, side view images, etc. All these images have different algorithms for processing. These algorithms are mainly based on deep learning to cover maximum features for high accuracy. Cristin et al. (2020) mentioned one such algorithm, wherein Deep Neural Networks (DNNs) algorithm is applied to detect the diseases in plants, such as Rider-Cuckoo Search Technique (RCSA). The method uses piecewise Fuzzy-C means and extracted a histogram of gradient-based features to represent the images. The features are given to a Deep Neural Network-based on RCSA to perform training and evaluation. The algorithm achieves 87% accuracy on the training plus testing set, which was highly considering that the researchers have used real-time images for both purposes (Training plus Testing). Other real-time applications such as apple leaf disease detection and prediction algorithm using Convolutional Neural Networks (CNNs) proposed by Jiang et al. (2019). Wherein researchers claim to have achieved an accuracy of 78.8% (20 frames per second video), and it helps in using the research for high-quality video processing also. Zhang et al. (2018) proposed Deep Convolutional Neural Networks (DCNN) to evaluate maize leaf diseases. The system uses rectified linear unit operator to assess the diseases. And they have claimed to have achieved 98% of training set accuracy. The test set precision is not apparent, but researchers have been considering it to be nearly more than 80% due to CNN-inspired networks’ usage.

Feature Dependent ML Approach for Disease Detection

Jagtap and Hambarde (2014) mentioned a more straightforward method for detecting diseases using morphological feature extraction. Wherein histogram-based filtering is applied here for preprocessing, followed by Fuzzy-C Means for segmentation. Then color, texture/ venation, and shape features are used to train a classification-based neural network. The results show a testing set accuracy of more than 70%, improved by CNN in the same system. These days, the image capturing process is usually by mobile phones. Raut and Nage (2019) proposed a model of smartphone camera images to evaluate the disease type of plant images. The system also uses CNN to perform classification. The system achieves more than 75% accuracy, and it is improved using advanced Deep Learning (DL) and Machine Learning (ML) classification techniques. Proposed work would suggest readers refer to Gaikwad and Musande (2017) to evaluate the most effective classifiers. Machine learning techniques such as Grey-Level Co-Occurrence Matrix (GLCOM) and Multi-Class Support Vector Machine (SVM) obtained the right classification level. The researchers have added that classification accuracies up to 80% using this combination. Similar to work by Raut et al. (2019), Manso et al. (2019) also proposed an advanced classification system using smartphone cameras. Their work has limitations for coffee-leaf disease detection; instead, it can extend to any plant’s leaf disease detection. They have used different color spaces like HSV and YCbCr to evaluate diseases in plants. They have taken more explicit images
for comparison and used basic methods like K-Means for segmentation. But their novelty is the use of extreme machine learning-based classifiers, which provides a high level of accuracy compared to other classifiers. The system can produce results that are more than 80% accurate.

**Plant Phenotyping Traits Analysis Using ML Approaches**

Rice crop having different types of traits analysis like leaf disease detection trait contains biotic and abiotic stresses (Zhao et al., 2019). Diseases come with pests, which comes under biotic stress, and identification of pests can help identify diseases. Such an identification system was proposed by Selvaraj et al. (2019), wherein Artificial Intelligence (AI) is applied to evaluate banana leaf diseases. They have used Multi-Level Convolutional Neural Networks and combined them using a transfer learning approach. The results are evaluated and suggested by research that more than 98% accuracy can be achieved for training set, while testing set accuracy will be more than 85%. These numbers are highly acceptable for the use of any real-time classification environment. Hyperspectral imaging can also be used to classify plant-leaf diseases. Like the work mentioned by Nagasubramanian et al. (2019), hyperspectral imaging is used to evaluate diseases in soya bean plantations. The results demonstrate that Convolutional Neural Networks combined with Hyperspectral Fusion and feature extraction methods like brovey and (Principle Component Analysis) PCA can achieve a more than 90% classification accuracy. They are making the system a top choice for real-time deployment. Abiotic stress includes weather inconsistencies like temperature, mentioned in the paper of Sharaff and Roy (2018) that how the temperature calculation is easy to evaluated by machine learning technique. One of the useful techniques: regression model vs. (Back-Propagation Neural Network) BPNN helped for abiotic stress calculation. Instead of classification between multiple diseases, the system classifies between healthy and non-healthy images. It made the classification simpler and more efficient and it is one of the great objectives for real data analysis also.

**ML Techniques and Method Dependent Data Analysis for Disease Detection**

A detailed survey of classification and processing methods has given by Kaur et al. (2019), which indicates that a Convolutional Neural Network is the de-facto choice for any plant disease classification system. This claim is further proven to be accurate by the work of Patel and Sharaff (2019). Faster Region-based Convolutional Neural Network (FRCNN) achieved a testing-set accuracy of more than 80% for a better classification system proposed by Akila and Deepan (2018). Like previous work, Sladojevic et al. (2016) also used CNN to obtain a high level of accuracy for test images. Results of Tetila et al. (2019) also suggested using Deep CNN for classification. From these above observations, the authors concluded that Deep learning-based CNN is better for the classification of different plant diseases. Tetila et al. (2019) proven for coffee plants that Deep learning is useful in preprocessing and dimensionality reduction (Sorte et al., 2019). Barbedo (2013) also told Deep network has various models for the classification of complex traits. Same as the proposed algorithm is inspired by CNN and achieves a high testing set accuracy of more than 85%. Patel and Sharaff (2020) discussed some other recent techniques, as Feature Fusion Technique gives a high-speed feature selection procedure for better classification. In contrast, the Contour-based Masking Technique provided by Patel and Sharaff (2020) directs the pixel value of the crop by using image data for the feature extraction. Here, the proposed algorithm is different from the techniques mentioned above are described in the next section, followed by its result analysis and comparison.

**METHODS AND MATERIALS**

This section is describing stepwise execution of combined image processing phases with machine learning algorithms and stepwise procedure is presented in Figure 2. Machine learning classifiers is a significant phase for the rice crop disease prediction. The classifier is giving an efficient result with the following listed factors:
Dataset Collection

Dataset combines the categories by identifying images that are most suited for the task of classification. These images must vary in terms of disease types, light intensity, crop types, disease texture/venation, color levels, etc. A more substantial dataset has been selected; thereby a stronger classification model will be created. In this proposed work online dataset has been taken for the observation, and the source is https://www.kaggle.com/vbookshelf/rice-leaf-diseases. Chhattisgarh is one of the significant sources of rice production over the Nation. As a future work, this proposed technique will be applied for the dataset taken from Indira Gandhi Agriculture University, Raipur (Chhattisgarh) for different disease prediction of rice crop to make this technique global and effective (Shrivastava and Pradhan, 2020).

Pre-Processing of Dataset

In the preprocessing stage, operations like denoising, smoothening, etc., are performed. A sufficiently preprocessed image is ideal for processing further because it clarifies the regions of interest extracted by the segmentation algorithm (Boulent et al., 2019).

Segmentation of Defected Regions

The segmentation algorithms play a significant role in the overall classification process. Segmentation algorithms like saliency maps, silhouette-based segmentation, edge-color-texture-based segmentation, color index-based segmentation (Castillo-Martínez et al., 2020), etc. must be carefully selected as per the database. A properly tuned segmentation algorithm will yield the most useful classification results. Researchers and system designers prefer the segmentation algorithm (or a combination of them) based on their efficiency on the dataset under test.

Feature Extraction of Segmented Area

Real-time applications of direct automation or identification of plant type can be easily identified even with similar plant leaves in all respective environmental conditions (Sachar and Kumar, 2020). Therefore, after the segmentation, desired images are given to a feature extraction unit. Here, features like color maps, edge maps, texture maps, region-based features, etc., are evaluated. These features can convert the segmented regions into numerical values that can be used for further comparison.
Process of Feature Selection

The feature selection process is applied to determine the best & most varying features for the given dataset. Better features mean a higher chance of proper classification. Algorithms like variance-based methods, key-point selection methods, etc., are used to find the optimum feature set (Li et al., 2020).

Machine Learning Classifier for Classification

After the feature evaluation, a series of classifiers are applied to obtain the final disease category of the input image. These classifiers generally use a feedback mechanism like backpropagation neural networks (Fulari et al., 2020) to learn from their results and obtain the best classification accuracy. Almost 30% of all research in plant phenotyping applications is done to find out the best classifier method. This proposed paper has worked on the feature selection, extraction, and classification process to obtain a useful and accurate crop disease prediction method.

Post-Processing

The proposed work compared different crop disease prediction methods for the next section, followed by the proposed machine learning-based algorithm. Finally, the paper concludes with some interesting observations about the compared algorithms and the proposed algorithm, followed by the future work that researchers can perform to further analyze & required enhancement of these proposed algorithms.

EXPERIMENTAL SETUP

This paper proposed a machine-learning algorithm to detect diseases in the crop images and implemented it in two phases, listed as following subsections:

Phase One: The Pre-execution/ Intensive Learning Phase

Perform the Works in the Following Steps:

This phase pre-execution task has been implemented with the initialization of all the desired parameters and learning solutions.

**Input Parameters**: Solution Numbers

- Selection of Classifiers (CFs)
- Selected Features
- Lc Value

**Output Parameters**: Accuracy (%)

Step 1: Initialization

- Initialize the desired learning parameters, included as:
  - Number of learning rounds/ counts = Nr
  - Number of learning solutions = Ns
  - Initial learning rate = Lr
  - Max features per solution = Fmax
  - Max classifiers per solution = Cmax
  - Max number of images per learning set = Imax

Step 2: Iterative Execution

- In this round each round give new changed solution, performing the following task to find a new approach for better solution:
1. Selecting a random set of images from the given dataset,
2. Make sure this dataset contains at most Imax images,
3. Evaluate features such as color, shape, texture, and other features from these given images dataset,
4. Selecting Fmax features from this image for feature selection,
5. Next step makes the selection of these Fmax features based on the distance of features from one another (Because closely connected features generally have the same information, while features that are far from each other have more different information),
6. Selecting Cmax number of random classifiers, list of classifiers is given as:
   i. K-nearest Neighbours (KNN)
   ii. Support Vector Machine (SVM)
   iii. Neural Network with Different Layer Configurations (NN)
   iv. Quadratic Linear Classifier (QL)
   v. Mahala Nobis Classifier (MH)
   vi. Random-forest Classifier (RF)
   vii. Naïve Bayes Classifier (NB)
7. Apply classifier learning for all the Cmax classifiers using these Fmax features on each of the selected Imax images,
   Accuracy evaluation of the classifier system, and mark it as the learning convergence for this solution, using the following formula,

\[
Lc = \frac{\sum_{i=1}^{Imax} Ai / Ndi}{Imax}
\]  

(1)

Where \( A_i \) = Accuracy for \( i \)th image
\( Ndi \) = Normalized delay needed for processing the \( i \)th image

Evaluate the normalized delay using the following formula,

\[
Ndi = \frac{di}{\sum di}
\]  

(2)

Where \( di \) = delay needed to process the \( i \)th image

10. Observe this approach of iterative execution for a better solution, and keep it for ready reference.

Learning convergence evaluation for each of the coming solutions, and then evaluating the learning threshold value,

\[
Lth = \frac{\sum_{i=1}^{Ns} Lci}{Ns} * Lr
\]  

(3)

Where \( L_i \) is the learning rate
For each coming solution that should satisfy the equation (4), pass it onto the next round; else, discard the solution and replace it with step (2),

\[ L_{ci} > L_{th} \] (4)

Step 5: Repeat steps (2) to (4) for Nr rounds, and prepare the following input/output parameters at the end of the Nr round.
Step 6: From the input/output parameters, select the solution with the highest Lc value and the highest value of accuracy and use it for the execution phase. CFs stands for selection of classifiers.

Due to the intensive learning phase, the proposed work gets many solutions, which are kept for further evaluation in the actual execution phase. The entire process can be demonstrated with the help of the following flowchart:

**Phase Two: Extensive Learning Phase:**

In this phase, evaluation of classifier has been discussed over the image dataset. The following steps are performed in the execution phase:
Step 1: Select the best accuracy entry from the learning input/output parameters.
Step 2: For each of the input images in the dataset, apply the feature selection method.
Step 3: Apply the classifiers as mentioned and evaluate the image class (healthy and disease).
Step 4: Inject random training set entries for evaluation, and repeat steps (1) to (3) for these random entries.
Step 5: Evaluate the accuracy for these random entries and evaluate Lc’s value for each of these entry sets.
Step 6: If Lc’s value for a given set is lower than the selected input/output parameters, then update the parameters with this value.
Step 7: Select the next best entry from input/output parameters, and repeat the process for each of the images.
Step 8: In case more than half of the entries of input/output parameters are replaced, then retrain the algorithm with the help of the pre-execution step, and re-create parameters with better entries/thresholds of Lc.

**Phase Three: Adaptive Feature Selection Algorithm:**

Due to the applied continuous learning process in this algorithm, the overall performance of the system improves the accuracy, and proposed work gets a better classification rate other than the individual algorithms. Therefore, the proposed adaptive feature selection algorithm is designed using the following steps:
Step 1: Features Extraction Algorithm
Initially, the input dataset is given in image form to a key point extraction algorithm, which calculates Maximal Stable Extremal Regions (MSER). It applies Speed Up Robust Features (SURF) method to each of these regions for ROI. It also uses color maps, edge maps, and other feature extraction algorithms on the image dataset. The SURF method evaluates the selected features for each of the regions; these key points’ distribution and values help in evaluating image properties. Once the chosen features are evaluated, they are tagged/marked using the training phase.
Step 2: Training the System for High Performance
The training phase stores information about the image features and the tag, which consists of the image class. This information helps identify the features, which are later used for comparison by the adaptive key-point selection algorithm and identify the disease type.
Step 3: Adaptive Key Point Selection Method for Disease Detection
The tagged database and full key point selected features of the given image were tested with the help of the adaptive key-point classifier. The adaptive key-point classifier works via the following process,

1. Feature-length helps to distinguished the query image feature for a given region,
2. Feature matching procedure: For each selected region, the number of features equal to the features in the given database is compared using standard keypoint matching. Score value (S1) is evaluated according to the number of features matched.
3. For another unequal feature-length, the following process is followed,
   i. If the number of features of the query image is not more than the number of features of the given database entries, then the database entries are trimmed, and a comparison is made to find out the score value (S2).
   ii. If the number of features of the query image is more than the number of features of the given database entries, then the input image feature entries are trimmed, and a comparison is made to find out the score value (S3).
4. The total score value of the tested image is evaluated using the following formula, $S_i = S_1 + S_2 + S_3$ (sum of all the score values). There is a value of $i = 1$ to 3.
5. This is performed for each database entry, and score values are evaluated for all data entries.
6. Finally, the given data entry with maximum score is selected, and classification is evaluated for a better solution.

Based on this machine learning classification, the proposed work obtains the input image as initial input belongs to either disease class or healthy crop for a better solution. The following parameters will evaluate disease prediction of rice crop: Accuracy, Precision, Recall. Delay is one of the other parameters taken for the proposed work to calculate the processing speed of the proposed work compared to other machine learning techniques.

**RESULT AND DISCUSSION**

The proposed algorithm is evaluating the results in terms of delay, accuracy, precision, and recall. The proposed approach achieves maximum accuracy compared to other state-of-art approaches based on neural networks. The result evaluation and their outcomes has been presented in Table 1. This section describes the quantitatively evaluated machine learning parameters with respect to learning phases. The proposed approach results have been compared with different neural networks model (Kaur et al., 2019) and is presented in Table 1.

Table 1 shows that the results of the proposed algorithm are better than the most effective CNN classification system. The previous work has discussed that classification for the real-time dataset accuracy is a challenging task. The accuracy of the images and the number of images that are taken for the analysis are very significant for this proposed algorithm. Moreover, due to the 2-stage classification nature of the proposed system, there is a reduction in processing delay, which can be observed from the following Table 2. Delay comparison Table 2 describes delay measurement (in ms) vs. given range of images for the proposed work with recent techniques Back-Propagation Neural Network (BPNN), Convolutional Neural Network (CNN), and Support Vector Machine (SVM) list of machine learning classifiers. For delay comparison number of images is also very significant to capture the difference. The classifier needs to identify this delay comparison for the processing speed calculation of the proposed work. Due to the intensive training phase, delay of the actual classification phase is reducing by more than 20% compared to other methods. The intensive training phase describes the iterative solutions for all the possible ranges of a given dataset.
### Table 1. Accuracy Comparison of Proposed Work

<table>
<thead>
<tr>
<th>Number of Images</th>
<th>Accuracy (%) BPNN (Kaur et al. 2019)</th>
<th>Accuracy (%) CNN (Kaur et al. 2019)</th>
<th>Accuracy (%) SVM (Kaur et al. 2019)</th>
<th>Accuracy (%) Proposed Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>80</td>
<td>90</td>
<td>68</td>
<td>90</td>
</tr>
<tr>
<td>20</td>
<td>85</td>
<td>90</td>
<td>70</td>
<td>95</td>
</tr>
<tr>
<td>30</td>
<td>86</td>
<td>94</td>
<td>72</td>
<td>95</td>
</tr>
<tr>
<td>50</td>
<td>86.5</td>
<td>95</td>
<td>72.6</td>
<td>96</td>
</tr>
<tr>
<td>75</td>
<td>86.9</td>
<td>95.5</td>
<td>72.96</td>
<td>96.5</td>
</tr>
<tr>
<td>100</td>
<td>87.2</td>
<td>95.6</td>
<td>73.12</td>
<td>97.5</td>
</tr>
<tr>
<td>125</td>
<td>87.4</td>
<td>96.3</td>
<td>73.48</td>
<td>97.8</td>
</tr>
<tr>
<td>150</td>
<td>87.5</td>
<td>96.8</td>
<td>73.72</td>
<td>98.2</td>
</tr>
<tr>
<td>200</td>
<td>87.7</td>
<td>96.9</td>
<td>73.84</td>
<td>98.6</td>
</tr>
<tr>
<td>250</td>
<td>87.9</td>
<td>97.3</td>
<td>74.08</td>
<td>98.7</td>
</tr>
<tr>
<td>300</td>
<td>88.1</td>
<td>98.1</td>
<td>74.48</td>
<td>98.9</td>
</tr>
<tr>
<td>400</td>
<td>88.2</td>
<td>98.1</td>
<td>74.52</td>
<td>99.1</td>
</tr>
<tr>
<td>500-1000</td>
<td>88.5</td>
<td>98.1</td>
<td>74.64</td>
<td>99.2</td>
</tr>
</tbody>
</table>

### Table 2. Delay Comparison of Proposed Work

<table>
<thead>
<tr>
<th>Number of Images</th>
<th>Delay (ms) BPNN (Kaur et al. 2019)</th>
<th>Delay (ms) CNN (Kaur et al. 2019)</th>
<th>Delay (ms) SVM (Kaur et al. 2019)</th>
<th>Delay (ms) Proposed Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7.20</td>
<td>10.80</td>
<td>6.00</td>
<td>5.33</td>
</tr>
<tr>
<td>20</td>
<td>7.30</td>
<td>10.95</td>
<td>6.08</td>
<td>5.41</td>
</tr>
<tr>
<td>30</td>
<td>7.45</td>
<td>11.18</td>
<td>6.21</td>
<td>5.52</td>
</tr>
<tr>
<td>50</td>
<td>7.57</td>
<td>11.35</td>
<td>6.31</td>
<td>5.60</td>
</tr>
<tr>
<td>75</td>
<td>7.69</td>
<td>11.54</td>
<td>6.41</td>
<td>5.70</td>
</tr>
<tr>
<td>100</td>
<td>7.82</td>
<td>11.73</td>
<td>6.51</td>
<td>5.79</td>
</tr>
<tr>
<td>125</td>
<td>7.94</td>
<td>11.91</td>
<td>6.62</td>
<td>5.88</td>
</tr>
<tr>
<td>150</td>
<td>8.07</td>
<td>12.10</td>
<td>6.72</td>
<td>5.98</td>
</tr>
<tr>
<td>200</td>
<td>8.19</td>
<td>12.29</td>
<td>6.83</td>
<td>6.07</td>
</tr>
<tr>
<td>250</td>
<td>8.32</td>
<td>12.48</td>
<td>6.93</td>
<td>6.16</td>
</tr>
<tr>
<td>300</td>
<td>8.44</td>
<td>12.66</td>
<td>7.03</td>
<td>6.25</td>
</tr>
<tr>
<td>400</td>
<td>8.57</td>
<td>12.85</td>
<td>7.14</td>
<td>6.35</td>
</tr>
<tr>
<td>500-1000</td>
<td>8.69</td>
<td>13.04</td>
<td>7.24</td>
<td>6.44</td>
</tr>
</tbody>
</table>
### Table 3. Precision Comparison of Proposed Work

<table>
<thead>
<tr>
<th>Number of Images</th>
<th>Precision (%) BPNN (Kaur et al. 2019)</th>
<th>Precision (%) CNN (Kaur et al. 2019)</th>
<th>Precision (%) SVM (Kaur et al. 2019)</th>
<th>Precision (%) Proposed Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>81</td>
<td>91</td>
<td>68.8</td>
<td>95.56</td>
</tr>
<tr>
<td>20</td>
<td>86</td>
<td>91.2</td>
<td>70.88</td>
<td>98.3</td>
</tr>
<tr>
<td>30</td>
<td>87</td>
<td>93.2</td>
<td>72.08</td>
<td>98.6</td>
</tr>
<tr>
<td>50</td>
<td>87.5</td>
<td>95.3</td>
<td>73.12</td>
<td>98.7</td>
</tr>
<tr>
<td>75</td>
<td>87.2</td>
<td>95.6</td>
<td>73.12</td>
<td>98.9</td>
</tr>
<tr>
<td>100</td>
<td>87.6</td>
<td>95.7</td>
<td>73.32</td>
<td>99.1</td>
</tr>
<tr>
<td>125</td>
<td>87.9</td>
<td>95.9</td>
<td>73.52</td>
<td>99.2</td>
</tr>
<tr>
<td>150</td>
<td>88.2</td>
<td>96.3</td>
<td>73.8</td>
<td>99.3</td>
</tr>
<tr>
<td>200</td>
<td>88.3</td>
<td>96.7</td>
<td>74</td>
<td>99.4</td>
</tr>
<tr>
<td>250</td>
<td>88.5</td>
<td>96.9</td>
<td>74.16</td>
<td>99.5</td>
</tr>
<tr>
<td>300</td>
<td>88.7</td>
<td>97.1</td>
<td>74.32</td>
<td>99.5</td>
</tr>
<tr>
<td>400</td>
<td>88.6</td>
<td>97.6</td>
<td>74.48</td>
<td>99.6</td>
</tr>
<tr>
<td>500-1000</td>
<td>88.7</td>
<td>97.8</td>
<td>74.6</td>
<td>99.6</td>
</tr>
</tbody>
</table>

### Table 4. Recall Value Comparison of Proposed Work

<table>
<thead>
<tr>
<th>Number of Images</th>
<th>Recall (%) BPNN (Kaur et al. 2019)</th>
<th>Recall (%) CNN (Kaur et al. 2019)</th>
<th>Recall (%) SVM (Kaur et al. 2019)</th>
<th>Recall (%) Proposed Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>65.00</td>
<td>75.00</td>
<td>56.00</td>
<td>77.78</td>
</tr>
<tr>
<td>20</td>
<td>68.00</td>
<td>76.00</td>
<td>57.60</td>
<td>80.00</td>
</tr>
<tr>
<td>30</td>
<td>72.00</td>
<td>78.00</td>
<td>60.00</td>
<td>83.33</td>
</tr>
<tr>
<td>50</td>
<td>79.00</td>
<td>79.00</td>
<td>63.20</td>
<td>87.78</td>
</tr>
<tr>
<td>75</td>
<td>81.00</td>
<td>82.00</td>
<td>65.20</td>
<td>90.56</td>
</tr>
<tr>
<td>100</td>
<td>85.00</td>
<td>86.00</td>
<td>68.40</td>
<td>95.00</td>
</tr>
<tr>
<td>125</td>
<td>86.00</td>
<td>87.00</td>
<td>69.20</td>
<td>96.11</td>
</tr>
<tr>
<td>150</td>
<td>87.00</td>
<td>88.00</td>
<td>70.00</td>
<td>97.22</td>
</tr>
<tr>
<td>200</td>
<td>88.00</td>
<td>89.50</td>
<td>71.00</td>
<td>98.61</td>
</tr>
<tr>
<td>250</td>
<td>88.50</td>
<td>90.20</td>
<td>71.48</td>
<td>99.28</td>
</tr>
<tr>
<td>300</td>
<td>88.90</td>
<td>90.30</td>
<td>71.68</td>
<td>99.56</td>
</tr>
<tr>
<td>400</td>
<td>89.00</td>
<td>90.50</td>
<td>71.80</td>
<td>99.72</td>
</tr>
<tr>
<td>500-1000</td>
<td>89.10</td>
<td>90.70</td>
<td>71.92</td>
<td>99.89</td>
</tr>
</tbody>
</table>
Proposed system can be used for high-end video processing disease detection systems as a future work with high efficiency. Similar results were obtained for precision and recall, which can be observed from the following Table 3 and Table 4.

The previous Table shows the precision of the proposed work with other techniques. Similar to that, recall values were also compared, which are shown in the following Table 4. From these results, the proposed machine learning algorithm is superior in terms of accuracy and delay to other state of the art methods.

PRACTICAL IMPLICATIONS OF PROPOSED WORK

Disease detection of rice crops creates such a significant role in plant phenotyping analysis. According to previous work, it has been found that machine learning techniques are beneficial for disease detection, and the proposed adaptive learning technique is giving an efficient result with 99.2% for a massive number of image analysis on rice crop disease prediction. There are some of the significant implications of the proposed work listed as:

1. It is useful for early disease prediction of rice crops for the application of better yield prediction.
2. Leaf analysis provides the information of stress observation and growth analysis.
3. Classification techniques are very significant for filled and broken grain prediction for the heterogeneous dataset as a future work.
4. It is highly recommended for grain, panicle and spikelet’s disease prediction for the high rate of production as a future requirement.

CONCLUSION AND FUTURE SCOPE

Proposed work demonstrate that the adaptive machine learning algorithm outperforms standard state-of-the-art algorithms like SVM, BPNN, and CNN in terms of delay, accuracy, precision, and recall values. Result and discussion of proposed algorithm included all the phases with better accuracy compared to other techniques. The accuracy is improved by 25%, while the delay is reduced by more than 50% compared to the methods mentioned above. Improvement in precision and recall are also more than 25%, which indicates a superior performance of the proposed algorithm. Thus, the algorithm can be used for real-time evaluation of plant leaf diseases for heterogeneous plant data and can also be used as a future work for video processing-based systems for leaf-disease as well as grain disease detection. The number of tillers and grain counting is also a future research work of this proposed technique.

FUNDING AGENCY

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

ACKNOWLEDGMENT

The authors would like to thank the National Institute of Technology, Raipur, to provide a research facility and give encouragement for real-world data analysis.
REFERENCES


Bharati Patel has completed her graduation in Information Technology in 2012 from Shri Shankaracharya College of Engineering & Technology, Bhilai (C.G.). She has completed her postgraduation Master of Technology in 2015 in Computer Science & Engineering (Specialization- Computer Technology and Application) from Shri Shankaracharya College of Engineering & Technology, Bhilai (C.G.) and currently pursuing Ph.D. degree in Computer Science & Engineering admission in 2018 from National Institute of Technology Raipur, India. Her area of interest is Image Processing, Data Mining, and Information Retrieval.

Aakanksha Sharaff (PhD) is working as a Faculty (Assistant Professor) in Department of Computer Science & Engineering at National Institute of Technology Raipur Chhattisgarh India since July 2012. She has been actively involved in research activities leading to Data Science research and related areas. She holds Doctor of Philosophy in Computer Science & Engineering from National Institute of Technology Raipur (An Institute of National Importance) in 2017; Master of Technology from National Institute of Technology Rourkela (An Institute of National Importance) with Honours in 2012; and Bachelor of Engineering from Government Engineering College Bilaspur Chhattisgarh with Honours in 2010. She has received gold medal during her graduation and post-graduation. Till date she pursuists for excellence and various academic success including the Top Student in Post-Graduation Master of Technology (2012), Bachelor of Engineering (2010) and throughout her schooling. She has received the gold medal for being the Top Student in Higher Secondary School Certificate Examination (2006) and High School Certificate Examination (2004). She has completed all her degrees and schooling with HONOURS (Distinction) and studied from reputed national institutions. She has achieved various merit certifications including All India Talent Search Examination during her schooling. She is the Vice Chair of Raipur Chapter of Computer Society of India and Secretary of IEEE Newsletter of Bombay Section. She is actively involved in various academic and research activities. She has received Young Women in Engineering Award for the contribution in the field of Computer Science and Engineering in 3rd Annual Women’s Meet AWM 2018 by Centre for Advanced Research and Design of Venus International Foundation. She has received Best Paper Award for several research papers. She contributes in various conferences as Session Chairs, Invited/Keynote Speakers and has published good number of research papers in reputed International Journals and Conferences. She is contributing as an active technical reviewer of leading International journals of IEEE, Springer, IGI and Elsevier etc. Dr. Sharaff has supervised many undergraduate and postgraduate projects. Currently she is guiding four research scholars for Ph.D. She has visited Singapore and Bangkok, Thailand for professional as well as personal reasons. Her research areas focus mainly on Data Science, Text Analytics, Sentiment Analysis, Information Retrieval, Soft Computing, Artificial Intelligence, Machine and Deep Learning. She is editing two books on “Data Science and its Applications” with Taylor and Francis (CRC) publisher and “New Opportunities for Sentiment Analysis and Information Processing” with IGI Publisher.