A Novel Cross-SVM Framework for Wheat Rust Disease Recognition through Multimodal Fusion of Images, Mask R-CNN, and DenseNet

Deepak Kumar[1] and Vinay Kukreja[1,*]

ARTICLE INFO

Article history:
Received: 6 September 2023
Revised: 30 October 2023
Accepted: 7 February 2024
Online: 20 June 2024

Keywords:
Rust diseases
Multimodal fusion
Wheat crop
Mask-RCNN
Classification

ABSTRACT

Wheat rust diseases are one of the common threats to food security worldwide. If the rust diseases are not recognized during their development stages, then wheat yield quality loss decreases. Thus, it is necessary to develop a recognition model that will easily recognize rust diseases in its development stage. In this paper, a combined approach of multimodal fusion, DenseNet, Mask RCNN, and Cross-entropy Support Vector Machine (SVM) models have been employed that can easily tackle the complex problem of wheat rust disease recognition. Because of this integration, a comprehensive comprehension of the diseases is now possible, which takes into consideration both visual and verbal data. Taking advantage of DenseNet extensive connection to improve feature representation, the DenseNet model is utilized to perform accurate feature extraction from imagined data. While doing so, Mask RCNN contributes to the accurate localization and segmentation of rust-infected patches, thereby delivering the fine-grained information that is essential for classification. During region extraction, the Mask RCNN model has a high average precision (0.986) with fused images. The fused extracted images have been applied to layers of the DenseNet model for feature extraction purposes. The extracted features have been helpful for rust disease classification purposes. Four different types of recognition models such as Naïve Bayes, KNN, SVM, and SoftMax classifiers has been used to compare the performance of the CE-SVM model. Among from all classifiers, the CE-SVM (98.39%) model outperforms than four different classifiers in terms of their F1-score. A stable and accurate solution for the early diagnosis and management of these fatal diseases have been achieved through integration of multimodal fusion, DenseNet, Mask RCNN, and Cross-entropy SVM models.

1. INTRODUCTION

Wheat (Triticum spp.) is one of the most important cereal crops in the world, producing a staple food offer for a large majority of the world's population [1]. Because of its extensive cultivation and consumption, it plays an important role in global food security and economic stability. However, consistent and sustainable wheat production faces significant hurdles [2] from a variety of diseases that affect both output and quality. These diseases can cause significant economic losses, decreased food availability, and consequent price increases. Wheat's resistance to a wide range of diseases is due to its genetic variety, widespread farming, and the presence of several pathogens. Pathogens such as fungi, bacteria, and viruses, as well as abiotic factors [3], all contribute to the emergence and spread of these diseases. Effective wheat disease management needs a thorough understanding of the underlying processes that contribute to disease development, as well as accurate and timely detection procedures. According to national agriculture institute, a total of 6.7% [4] wheat grain quality has been reduced due to wheat fungal diseases. There are four types of fungal diseases namely as rust, leaf spot, spike infection and virus based diseases. Among from all types of fungal diseases, rust diseases are major disease that reduces the wheat grain quality in its flowering stage [5], [6], [7]. Wheat rust diseases are a collection of fungi that harm wheat plants and have the potential to lead to major losses in crop productivity and quality on a worldwide scale. The wheat rust diseases are transmitted through Puccinia triticum fungi pathogen [8]. Once, the rust diseases are transmitted the whole wheat plant is damaged along with its grain head. There are three types of rust diseases namely as stem rust, stripe rust and leaf rust. Wheat plants can suffer damage to their stems and leaves from a disease called stem rust. It generates spores that are a rusty reddish brown colour, which can give affected plants an unsightly appearance. Stem rust can cause severe yield losses by lowering a plant's ability to preserve grain-filled heads and weakening the

[1]*Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.*
**Corresponding author:* Vinay Kukreja; E-mail: onlyvinaykukreja@gmail.com.

plant's stems, both of which are susceptible to the disease. The wheat leaves are the main focus of the disease known as leaf rust. It leaves each leaf with small, spherical pustules that are a yellowish-orange color and may combine, causing the leaves to become damaged [9]. The yellow rust disease is characterized by the appearance of thin, yellowish stripes on the leaves with orange-colored pustules. Thus, the recognition of rust diseases is important which is helpful for farmers as well as food producer industries. Recognizing the importance of proactive and precise disease detection, the integration of new technologies and novel methodologies has become critical. Traditional disease detection methods, which rely mainly on visual inspections and expert knowledge, have subjectivity and time inefficiency issues. Over the years, researchers and farmers have put in a lot of work to figure out the complexities of wheat diseases, from the agents that cause them to the intricate connections among host, pathogen, and environment [10]. The rust diseases have been recognized through computer vision approaches such as machine learning, deep learning [2], [8], [11], image segmentation, and spectral imaging offer a viable avenue for improving wheat rust disease recognition rate. Machine learning and deep learning algorithms are effective technologies capable of autonomously analyzing large information and differentiating between healthy and diseased wheat samples with remarkable precision [12]. These technologies enable the rapid detection of diseases and the construction of appropriate intervention plans by dissecting detailed patterns and subtle subtleties in symptom development. Simultaneously, the integration of remote sensing and spectrum imaging improves disease detection by recording multispectral data [13], indicating physiological changes within wheat plants sometimes before observable symptoms appear. Several issues such as robust feature integration, extracting different wheat rut lesions, irregular lesion shapes, and complex backgrounds few studies [14], [15], [16] have been conducted to alleviate or remove the aforementioned obstacles for crop disease identification. To resolve these issues, a combined late fusion and Mask RCNN model is proposed in this paper that gives the advantages of both methodologies to deliver precise, thorough, and dependable rust disease detection in wheat plants. This method handles the challenges of disease detection, segmentation, and localization, resulting in better agricultural practices and food security.

The major contribution of this paper:

- Image fusion combines many modalities to improve rust disease recognition by capturing supportive features for more accurate detection and discrimination.
- An instance segmentation framework namely Mask R-CNN finds the rust leaves precisely and allows pixel-level identification.
- The pixel-level rust disease leaves help to extract the features using the Densenet feature extraction model. The resultant feature vectors are applied to the cross-entropy SVM model for performing a multi-classification of rust diseases.
- The proposed approach outperforms than standard CNN-based recognition models with the same number of parameters.

The structure of this paper is as follows: The dataset details along with the methods have been defined in section 2. The results and discussions have been presented in section 3. The conclusion of this paper has been presented in section 4.

2. PROPOSED METHODOLOGY

To recognize the rust disease precisely with different modalities, a combined approach of late fusion, mask rcnn, DenseNet and cross-entropy SVM model is performed. The overflow of the proposed methodology along with its phases has been shown in Figure 1. The phases of the proposed methodology are described hereunder.

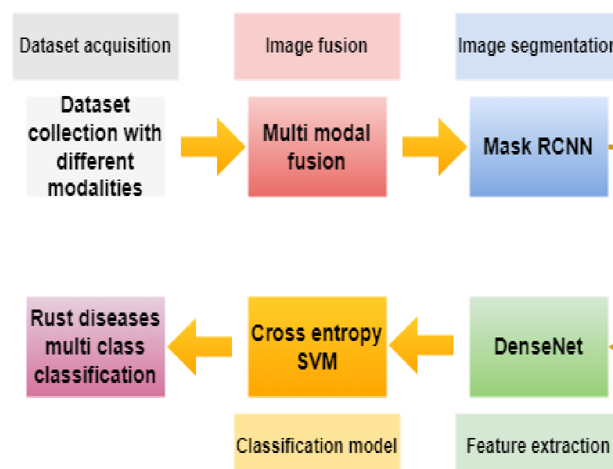

Fig. 1. Overflow of proposed methodology.

2.1. Dataset Collection with different modalities

The wheat rust diseases dataset is taken from secondary sources. A total number of 5406 images with different position modalities have been taken from web sources [22]. Among from collected images, 3950 images consist of rust diseases and 1456 images are of healthy parts. If the afflicted area appears in a center position, the recognition accuracy degrades. To improve the accuracy rate, a segmentation-based feature decomposition, and recombination technique was developed that allowed the diseased area to randomly appear at any location within the image. The images captured across various modalities are depicted in Figure 2.

Fig. 2. Samples of collected images with different modalities.

2.2. Multimodal Fusion

Multi-modal image fusion is a technique that combines data from multiple sources or modalities into a singular fused image to enhance the overall comprehension or analysis of a scene. When dealing with distinct modalities of angles, such as optical images captured from different angles, multi-modal fusion becomes especially useful. There are several steps to perform a multi modal fusion for performing fusion of images [8], [11]. The fused images are applied as an input to Mask RCNN model for leave regions extraction. Each modality is considered as a distinct model input. If you are combining RGB and thermal images, for instance, you would have two input layers, one for each modality. After the input layers, modality-specific preprocessing layers can be added for each input. These layers can perform duties such as resizing, normalization, and extraction of modality-specific features. You could, for instance, use convolutional layers to extract features from RGB images and then apply histogram equalization to the collected images. After the feature extraction layers, fusion layers can be added to combine the extracted features from various modalities. These are common fusion techniques:

- Combining attributes by concatenating them along a new attribute dimension.
- Addition or multiplication by element: Combine features by executing addition or multiplication by element.
- Utilize attention mechanisms to determine the relative relevance of features from various modalities.

Once, the fusion of two multimodal images is performed, the leave lesions are extracted through the Mask RCNN model. The output of multimodal fusion is shown in Figure 3.

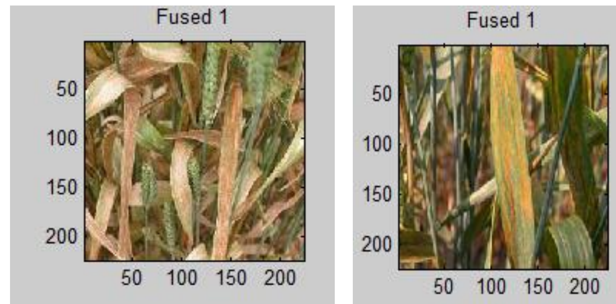

Fig. 3. Multimodal fused images.

2.3. Mask RCNN for Leave Lesions Extraction

Suppose we have a fused image A(i, j) with dimensions (X Y). where X is the number of rows, Y is the number of columns and (i, j) are the pixel values of image A (i, j). . A correlation formula is applied to the fused image A(i,j) to maximize the visual intensity of the lesion region. The purpose of this step is to provide MASK RCNN with an improved version of the original RGB image [23]. Vector 1 represents a pixel in each channel of the input image A (i, j), which is transformed into the corresponding pixel u of the output image A (i, j). RCNN mask layer consists of:

• **Backbone:** Mask RCNN mask uses regular CNN as backbone to extract features from image directly without any feature loss. A pre-trained ResNet50 model is used as the backbone. Using this CNN backbone, first a fused image of dimensions (224*224*3) is fed to the network that obtains a feature map of dimensions (1*448*1).

• **Region proposal network (RPN):** The extracted features are used to build a bounding box over the lesion area. RPN searches for anchor points in the extracted region. Even, the anchor points define the bounding box of the entire image. However, because the size of the bounding box is determined by the extent of the lesion, many anchor points are created. Only the top N anchor points are selected based on the higher probability value of the RPN.

• **Region of interest (ROI) alignment:** ROI alignment is the process of creating a very small feature map from each bounding box extracted. The alignment of the ROI layer is used to ensure that the features of the input image are properly aligned. Determine the exact values of the input feature vector at four independent locations using bilinear interpolation. This method is used to fine-tune the ROI extraction technique.

• **Mesh head:** After aligning or merging the ROIs, the feature map enters a series of convolutions.

2.4. DenseNet model for feature extraction

It is well-known for its dense connection pattern, in which each layer is feedforward coupled to every other layer. DenseNet is made up of several dense blocks, each of which has several convolutional layers followed by a concatenation process that combines the feature maps from all previous layers within the block. This extensive connectivity enables feature reuse and encourages feature propagation throughout the network [19]. DenseNet is made up of several dense blocks, each of which has several convolutional layers followed by a concatenation process that combines the feature maps from all previous layers within the block. This extensive connectivity enables feature reuse and encourages feature propagation throughout the network. There are bottleneck layers within each dense block. Before the subsequent 3x3 convolutional layers, these 1x1 convolutional layers minimize the number of channels. This design option contributes to a reduction in the number of parameters while maintaining expressive power. Instead of fully connected layers at the network's conclusion, Densenet frequently uses global average pooling. This operation computes the average of the values of each feature map over the spatial dimensions. The final softmax layer for classification or a set of fully connected layers for regression is then connected to the resulting feature vector. The feature vector (F_v) along with global average pooling (G) is represented in equation 1.

$$G = \frac{1}{h*w} \sum_{i=1}^{h} \sum_{j=1}^{w} F_v\,(i,j) \tag{1}$$

2.5. Cross Entropy SVM(CE-SVM)

In the area of machine learning or deep learning, a "cross-entropy SVM" is not a standard or commonly used word. Cross-entropy is a loss function that is frequently employed in SoftMax regression (multinomial logistic regression) [12], [21] and neural network models for multi-class classification tasks; however, it is not commonly related to Support Vector Machines (SVMs). The layers of the CE-SVM model have been described hereunder:

- Feature vector as an input layer: The input layer is given an appropriate input shape, which should match the length of your feature vectors.
- Hidden layer: These layers are responsible of learning complex representations from feature vector inputs. With a dropout rate of 0.5, approximately 50% of the neurons in the layer that came before will be "dropped out" during each training cycle. The categorical cross-entropy loss function, which is the usual loss function for multi-class classification problems, is used to assemble the model. The 'Adam' optimizer is often used, although alternative optimizers can be utilized depending on your problem.

The loss of the CE-SVM model is presented in equation 2.

$$H(x, y) = -\sum X_i \; log(y_i) \tag{2}$$

The *x* is the probability of real class *i* and *y* is probability of predicted class *i*.

3. RESULTS AND DISCUSSION

All the experiments have been performed on secondary sources datasets. The dataset details have been already described in section 3.1. As. Multimodal fusion has been performed to generate the fused dataset. The fused dataset has been used to extract leaf lesion regions. The fused leaf regions are carried through the Mask RCNN model. The MRCNN model performs a bounding box location of each leaf region. Once, the leaf regions have been extracted, the transformation of feature extraction and classification is performed. The multimodal fusion has been performed in Matlab R2013a. All the simulations of Mask RCNN, Densenet and CE-SVM is performed on python notebook. During experimentation, total accuracy, error rate, and execution time are also reported.

3.1. Segmentation Results

In this part, the segmentation results obtained by using the MASK-RCNN model are shown. This model makes heavy use of the resnet-50 model as its main component. The outcomes are showcased for layer-wise training, revealing the highest achieved Mean Overlap Coefficient (MOC) of 0.863 on ResNet50. Simultaneously, the average precision (AP) is calculated at various threshold values, specifically 0.5, 0.75, and 0.9, yielding rates of 0.986, 0.8963, and 0.6390, respectively. The segmentation results with or without image fusion with several threshold values have been presented in Table 1.

Table 1: Segmentation results with or without image fusion

Imaging type	Ap(0.5)	Ap(0.75)	Ap(0.9)	Mean over coefficient (MOC)
Without fusion	0.878	0.623	0.493	0.768
With fusion	0.986	0.8963	0.6390	0.8431

3.2. Recognition Results

In this section, the findings of the classification are discussed using a combination of tabular and graphic aids to enhance the discussion. We followed a ratio of 70:30 for training and testing, where 70% of the data was used for training and 30% was used for testing. A technique known as 10-Fold cross-validation is being utilized in order to further validate the results. As a total number of 12,532 patches in form of sub images have been extracted. The extracted patches have been used for classification purpose. The features have been extracted through different layers of DenseNet model. The extracted features are applied as an input to CE-SVM model. The performance of CE-SVM model has been compared with four different methods. The performance of recognition model has been measured through F1-score, precision, recall and false negative (FNR) parameters. The results of recognition models for wheat rust diseases multiclassification has been shown in Table 2.

Table 2: Performance evaluation

Model	F1-score (%)	Precision (%)	Recall (%)	FNR (%)
Naïve Bayes	84.73	85.56	87.63	16.36
SVM	85.36	85.9	82.38	15.69
KNN	86.89	85.63	89.32	14.8
Softmax	83.6	86.92	87.19	13.9
CE-SVM	98.39	95.98	96.72	11.3

The performance of CE-SVM model has been compared with four different classifiers. Among from all classifiers, the CE-SVM model outperforms than four different classifiers. As, CE-SVM performs multi classification of rust diseases in fused images. The parameters of cross entropy support vector machine for wheat rut diseases classification has been defined in table 3.

Table 3: Parameters details of CE-SVM model

Parameter name	Kernel function	Number of classes	C	Gamma
Parameter value	RBF	3	1.0	Scale

3.3. Result analysis

Figure 1 shows the number of processes for wheat rust diseases recognition. Most of the time, the lesion patches have brighter borders and darker centres. As a consequence of this, the framework for segmentation occasionally takes into account the darker center while ignoring the lighter boundary portions. The same issue continued to manifest itself after the MASK RCNN, and a select few regions which are part of the lesion have been disregarded. The aforementioned limits might be disregarded as a result of the exhaustive training that was performed on a substantial number of image examples. Following the completion of the MASK RCNN, the segmented pictures are applied to the DenseNet model to obtain high-level features. Lastly, the CE-SVM model performs a multi class classification of rust diseases and healthy plant leaf regions. The classifier performs different functions to evaluate the performance of recognition models.

4. CONCLUSION

The difficult subject of wheat rust diseases multiclass classification has been investigated in this paper through a combined strategy that blends multimodal fusion, DenseNet, Mask RCNN, and Cross-entropy Support Vector Machine models. The combined strategy is able to increase the classification system's overall accuracy and robustness by implementing multimodal fusion and harnessing the power of numerous data sources, including images and textual information. This allowed us to exploit the potential of multiple data sources. In both the image processing and the feature extraction processes, the integration of the DenseNet and Mask RCNN models was an essential step. DenseNet model has several different layers that effectively collected key features from the fused images, whereas Mask RCNN permitted precise localization and segmentation of the rust-infected areas. Both of these networks worked together to produce accurate results. In addition, the utilization of the CE-SVM model as the classification backbone revealed the efficiency of this technique in dealing with multiclass classification tasks. DenseNet and Mask RCNN both contributed to improved feature extraction and localization, while the multimodal fusion technique made a major contribution to the total performance improvement. The diseases were successfully classified using the Cross-entropy SVM model, which also achieved a high level of precision and recall rates. Four different types of recognition models such as Naïve Bayes, KNN, SVM and softmax classifiers has been used to compare the performance of CE-SVM model. Among from all classifiers, the CE-SVM (98.39%) model outperforms than four different classifiers in terms of their F1-score.

REFERENCES

[1] Eric Huttner, “Improving food security and farmers’ livelihood through productive and sustainable crops,” aciar.gov.au, 2021. [Online]. Available: https://www.aciar.gov.au/program/crops. [Accessed: 09-Sep-2022].

[2] Kaur, Jaspal, Ritu Bala, Harmandeep Kaur, P. P. S. Pannu, Ashok Kumar, and S. C. Bhardwaj. "Current status of wheat diseases in Punjab." Agric. Res. J 55, no. 1 (2018): 113-116.

[3] Singhi, Vishwas, Deepak Kumar, and Vinay Kukreja. "Integrated YOLOv4 Deep Learning Pretrained Model for Accurate Estimation of Wheat Rust Disease Severity." In 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 489-494. IEEE, 2023.

[4] Khan, Habib, Ijaz Ul Haq, Muhammad Munsif, Mustaqeem, Shafi Ullah Khan, and Mi Young Lee. "Automated wheat diseases classification framework using advanced machine learning technique." Agriculture 12, no. 8 (2022): 1226.

[5] Jahan, Nusrat, Paulo Flores, Zhaohui Liu, Andrew Friskop, Jithin Jose Mathew, and Zhao Zhang. "Detecting and distinguishing wheat diseases using image processing and machine learning algorithms." In 2020 ASABE Annual International Virtual meeting, p. 1. American Society of Agricultural and Biological Engineers, 2020.

[6] Kumar, Deepak, and Vinay Kukreja. "Early recognition of wheat powdery mildew disease based on mask RCNN." In 2022 International Conference on Data Analytics for

Business and Industry (ICDABI), pp. 542-546. IEEE, 2022.

[7] Kukreja, V., & Kumar, D. (2021, September). Automatic classification of wheat rust diseases using deep convolutional neural networks. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-6). IEEE.

[8] Xie, X., Zhang, X., He, B., Liang, D., Zhang, D., & Huang, L. (2016, October). A system for diagnosis of wheat leaf diseases based on Android smartphone. In Optical Measurement Technology and Instrumentation (Vol. 10155, pp. 572-580). SPIE.

[9] Kumar, Deepak, and Vinay Kukreja. "MRISVM: A object detection and feature vector machine based network for brown mite variation in wheat plant." In 2022 International Conference on Data Analytics for Business and Industry (ICDABI), pp. 707-711. IEEE, 2022.

[10] Kumar, D., & Kukreja, V. (2023). Combined CNN with STARGAN for Wheat Yellow Rust Disease Classification. International Journal of Computing and Digital Systems.

[11] Ang, S., Leeton, U., Chayakulkeeree, K., & Kulworawanichpong, T. (2018). Sine cosine algorithm for optimal placement and sizing of distributed generation in radial distribution network. GMSARN International Journal, 12(4), 202-212.

[12] T. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, “Predictive analysis of heart diseases with machine learning approaches,” Malaysian J. Comput. Sci., pp. 132–148, 2022.

[13] U. K. Lilhore, M. Poongodi, A. Kaur, S. Simaiya, A. D. Algarni, and and M. H. Hela Elmannai, V. Vijayakumar, Godwin Brown Tunze, and Mounir HamdiUmesh Kumar, M. Poongodi, Amandeep Kaur, Sarita Simaiya, Abeer D. Algarni, Hela Elmannai, V. Vijayakumar, Godwin Brown Tunze, “Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques.,” Comput. Math. Methods Med., 2022.

[14] “Rust Diseases of Wheat | Ohioline.” [Online]. Available: https://ohioline.osu.edu/factsheet/plpath-cer-12. [Accessed: 29-Mar-2022].

[15] Jaroenwanit, P., & Kantatasiri, P. (2014). Consumer Perception and Attitude Study for Market Development of Hommali Organic Rice Products from Thung Kula, Thailand. GMSARN International Journal, 89.