The classification of wheat yellow rust disease based on a combination of textural and deep features

Tolga Hayıt[1] · Hasan Erbay[2,3] · Fatih Varçın[4] · Fatma Hayıt[5] ⬥ · Nilüfer Akci[6]

Abstract

Yellow rust is a devastating disease that causes significant losses in wheat production worldwide and significantly affects wheat quality. It can be controlled by cultivating resistant cultivars, applying fungicides, and appropriate agricultural practices. The degree of precautions depends on the extent of the disease. Therefore, it is critical to detect the disease as early as possible. The disease causes deformations in the wheat leaf texture that reveals the severity of the disease. The gray-level co-occurrence matrix(GLCM) is a conventional texture feature descriptor extracted from gray-level images. However, numerous studies in the literature attempt to incorporate texture color with GLCM features to reveal hidden patterns that exist in color channels. On the other hand, recent advances in image analysis have led to the extraction of data-representative features so-called deep features. In particular, convolutional neural networks (CNNs) have the remarkable capability of recognizing patterns and show promising results for image classification when fed with image texture. Herein, the feasibility of using a combination of textural features and deep features to determine the severity of yellow rust disease in wheat was investigated. Textural features include both gray-level and color-level information. Also, pre-trained DenseNet was employed for deep features. The dataset, so-called Yellow-Rust-19, composed of wheat leaf images, was employed. Different classification models were developed using different color spaces such as RGB, HSV, and L*a*b, and two classification methods such as SVM and KNN. The combined model named CNN-CGLCM_HSV, where HSV and SVM were employed, with an accuracy of 92.4% outperformed the other models.

1 Introduction

Yellow rust disease caused by *Puccinia striiformis* f. sp. *tritici* pathogen leads to significant losses in wheat production worldwide because of its high devastating property [37, 54]. It is a global disease and has been determined in more than 60 countries [54]. Turkey also

✉ Fatma Hayıt
  fatma.hayit@bozok.edu.tr

Extended author information available on the last page of the article.

suffers from the wheat yellow rust epidemic [13]. It caused crop damage estimated at ten million US dollars in Central Anatolia in 2010 [6]. Due to disease, yield losses might reach up to 100% when hypersensitive cultivars are used and when suitable weather conditions occur [11]. Moreover, yellow rust limits not only the yield but also the quality of wheat. As a result, it is critical to monitor the development of yellow rust in natural conditions using reliable methods for minimizing economic losses. The disease causes physical changes on wheat leaves such as leaf color, leaf texture, and leaf morphology [26], the severity of the disease is evaluated through these physical deformations. In other words, texture features provide enough evidence about wheat yellow rust [40]. However, the process of determining the severity level of yellow rust is carried out according to certain standards through visual inspection in field conditions by specialists. This process is subjective, labor-intensive, and time-consuming. Failure to perform accurate severity level evaluation may pose problems in later growth stages, especially in precision agriculture in which detecting infection type is critical in taking necessary precautions against the disease.

Texture features analysis over two-dimensional images is one of the most popular approaches in digital image processing problems [2]. It focuses on partitioning images into regions of interest and, later, classifying those regions. It utilizes existing information on the spatial arrangement of colors or intensities in an image. It extracts information about texture features, structural features of surfaces, and their relationship with each other [17]. Texture feature extraction methods have been used for about half a century [18] in various computer vision applications, including object detection and image classification [55]. Over time, the number of features has increased to better adapt to the various tasks. The methods consist of various techniques such as Gray Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), Gabor Filter, Segmented Fractal Texture Analysis (SFTA) each based on different calculations. GLCM is a popular measurement among the other texture feature extraction methods designed to perform statistical calculations over different image textures and forms [17]. It is often used especially in the fields of texture analysis and machine learning. GLCM performs calculations using statistical distributions of combinations of different intensity values for an image [48]. However, capturing information only from grayscale images limits GLCM features' effectiveness with color images. Thus, recently, color spaces have attracted the attention of researchers as they reveal hidden patterns in color channels [15, 20].

On the other hand, recent developments in both information technologies and image analysis offer new methods to extract data-dependent features to develop more sophisticated systems for image-based problems [36, 47, 50]. Convolutional Neural Network (CNN) is in the leading position among these methods due to its ability to have both automatic feature extraction and classification. It also has the advantage of spatial information over other types of deep networks. CNN is a deep learning (DL) method that is especially popular in classifying images, text, sounds, and videos. One of CNN's strongest sides is that it automatically extracts data-dependent appropriate features from images in the dataset. Researchers are still on the hunt to improve the effectiveness and generalizability of deep learning models [35, 51].

Herein, various models have been proposed to determine the severity of yellow rust disease in wheat. The models were trained on a dataset that included not only CNN-based auto-generated data representation features, but also traditional texture features. The study focuses on the hypothesis of *"diversity in the feature set is crucial in developing state-of-art learning models trained on image datasets even though dataset-representative image-based features are available"*. Pre-trained DenseNet-201 was employed to extract CNN-based features. DenseNet-201 was fine-tuned by the so-called Yellow-Rust-19 dataset [19] to obtain

task-specific features. DenseNet-201's Global Average Pooling (GAP) layer was utilized for extracting 1920 features. Also, using both the Gray Level Co-occurrence Matrix (GLCM) and the Color-Gray Level Co-occurrence Matrix (CGLCM), a total of 88 texture features were obtained, 22 features corresponding to each grayscale and 22 features corresponding to each color channel for each leaf image in Yellow-Rust-19 dataset [19]. Moreover, three different color spaces such as Red-Green-Blue (RGB), Hue-Saturation-Value (HSV), and Lightness-Red/Green-Blue/Yellow (L*a*b) were utilized for obtaining color textural features.

Experimental results confirm that models trained using combined features sound and are promising. Experimental results also confirm that the models developed using combined features perform better than the models developed by using one of the CNN-based features and textural features. The support Vector Machine (SVM) method and K-Nearest Neighbors (KNN) were employed as the classifiers. Figure 1 shows the overall workflow.

The remaining part of this paper is organized as follows. In the next section, some studies' reviews related to the methodologies used in the study are presented. Section 3 explains the methodological framework. This section gives information about the final data set and feature extraction methods. The combined textural analysis results based on different color spaces and CNN results are presented in Section 4. In the last section, the results are summarized and recommendations for further studies are given.

2 Related work

Precision farming is heavily dependent on information technology, with data and connectivity at its core. With artificial intelligence, analytics, and connected sensors, it can further increase yields, detect diseases and build sustainability and resilience across crop cultivation. Kundu et al. [25] developed a CNN-based model called *Custom-Net* that collects the imagery and parametric data from the pearl millet farmland and predicts the blast and rust diseases in pearl millet. Hayit et al. [19] introduced a multi-label deep convolutional neural networks-based model to classify the severity of yellow rust disease on wheat. The model inputs the wheat leaf image and classifies it by the percentage of rust. However, Abayomi-Alli et al. [1] aimed to improve the accuracy of deep learning models on low-quality test images using data augmentation. They generated synthetic images by histogram transformation technique to recognize important colour-based features, which are less sensitive to the deficiencies of low-quality images.
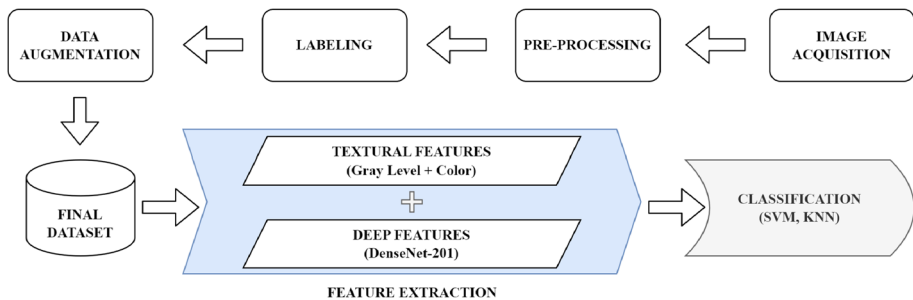

Fig. 1 The workflow of the study

On the other hand, there exists recent studies on the diagnosis of plant diseases through rich, informative, and data-representative features. Almadhor et al. [3] developed an artificial intelligence (AI)-based model to detect and classify the most common guava plant diseases by utilizing a combination of color histograms and Local Binary Patterns (LBP) textural features. However, GLCM is a widely used statistical method for texture analysis, but it can only capture information from grayscale images [18, 20], which limits its effectiveness with color images. Recently, color spaces have also attracted researchers' attention due to their ability to reveal color data hidden in pixel values. Thus, many efforts have been made to extend the effectiveness of GLMC by combining or integrating the color in it. For example, Hossain and Parekh [20] studied the possibilities of including color information from color texture images to improve texture analysis. The results proved that color GLCM combination provided better accuracies as compared to standard GLCM. In another study, Gui et al. [15] demonstrated the texture feature extraction based on Color Co-occurrence Matrix is superior over the GLCM-based texture feature extraction methods. HSI and RGB color spaces were used for texture feature extraction. More on combining color and grey-level co-occurrence matrix features, see [23, & references therein].

Nowadays, with the emergence of deep learning, especially CNNs, there has been a shift to the use of both hand-crafted and learning-based features. Zhou et al. [58] employed hand-crafted features like color histogram and LBP for image retrieval as well as the deep features extracted from the various CNNs such as VGG16, ResNet, and AlexNet. Zhang et al. [56] proposed a classification model for medical images, which the model was trained by utilizing deep and handcrafted visual features. In [53], Wei et al. proposed a hybrid feature descriptor based on DL by integrating the deep and handcrafted features. An SVM-based classification model was used for training. Experimental results show that the proposed method outperforms the state-of-the-art predictors.

Nowadays, with the emergence of deep learning, especially CNNs, there has been a shift to the use of both handcrafted and learning-based features. Zhou et al. used deep features extracted from various CNNs such as VGG16, ResNet and AlexNet as well as handcrafted features such as color histogram and LBP for image acquisition. Zhang et al. [58] proposed a classification model for medical images, in which the model is trained using deep and handcrafted visual features. In [53], Wei et al. proposed a hybrid feature descriptor based on DL by integrating deep and handcrafted features. A SVM-based classification model was used for training. Experimental results show that the proposed method outperforms state-of-the-art predictors. Nanni et al. [32] proposed a model for the classification of virus images by using both hand-crafted features and pre-trained deep neural network features. SVM was used for features training. Various pattern analysis methods were used as handcrafted features extraction methods. The experiment results showed that the combined method with hand-crafted and deep features improve the accuracy. Luz et al. [27] presented an urban sound classification method based on deep and handcrafted features combination. CNN model was used for extraction deep features. Random Forest and SVM were used as classifiers and also two separate data sets were used. For both classifiers, the proposed method outperformed handcrafted features. Zhang et al. [57] used a combination of handcrafted and deep CNN features extracted from X-Ray images for classify between healthy, widespread pneumonia, and Covid-19 patients. The performance of the proposed approach was compared with a standard CNN and the SVM trained with texture features. It was reported that combining the features improves the accuracy of the classification compared to the independent application of handcrafted and deep features. Naz et al. [33] proposed a hybrid method based on the texture features and deep features for automated detection of gastrointestinal diseases by using wireless capsule endoscopy images. Texture features were extracted by

using SFTA and LBP texture analysis methods. Also, deep features were extracted by using VGG-16 and Inception-V3 pre-trained CNN models. Extracted deep features were combined serially along with the obtained texture features to obtain an ensemble deep feature vector. It was reported that the results of proposed approach are promising.

In addition, state-of-the-art studies utilized the DenseNet [21] model for similar tasks that this study is concerned with [12, 31, 44, 49]. Each of these studies used leaf images of different plants as a dataset and successfully detected diseases. While some of these studies use the basic DenseNet model, others modified the DenseNet to their specific task.

The studies summarized above indicate that a combination of textural and deep features improves the performance of learning-based models, which motivated us to explore a learning-based classification performed on a combination of textural and deep features to determine the severity level of yellow rust disease in wheat.

3 Materials and method

3.1 Wheat yellow rust infection types

Yellow rust disease causes physical changes in wheat leaves such as leaf color, leaf texture, and leaf morphology [26], which provide sufficient evidence of wheat yellow rust disease [40] to determine the type of yellow rust infection. Relative to the resistance of wheat to the disease, there are six different infection types according to the modifed Cobb scale [38]. Infection types no disease (Immune), Resistant (R), Moderately Resistant (MR), Moderately Resistant—Moderately Susceptible (MR-MS), Moderately Susceptible (MS), and Susceptible (S) in Table 1 presented by Johnston and Browder [22] were used for disease assessment. However, the process of determining the severity level of yellow rust is carried out according to certain standards through visual inspection by specialists.

3.2 Dataset

The classification has been performed on the dataset called *Yellow-Rust-19*, which was initially created by Hayıt et al. [19] and composed of wheat leaf images. In order to obtain images, first, wheat leaves were collected from the cultivation area belonging to the Republic of Türkiye Ministry of Agriculture and Forestry Directorate of Field Crops Central Research Institute located in Ankara, Türkiye. After that each leaf was examined to determine the severity level of yellow rust by two specialists. Then, they were photographed to yield a raw dataset. After that, each image in the raw dataset went through some image

Table 1 Infection types of wheat leaf rust used in disease assessment

| Immune	No visible infection on plant.
R	Resistant: visible chlorosis or necrosis, no uredia are present.
MR	Moderately Resistant: small uredia are present and surrounded by either chlorotic or necrotic areas.
MR-MS	Intermediate: variable sized uredia are present; some with chlorosis, necrosis, or both.
MS	Moderately Susceptible: medium sized uredia are present and possible surrounded by chlorotic areas.
S	Susceptible: Large uredia are present, generally with little or no chlorosis and no necrosis.
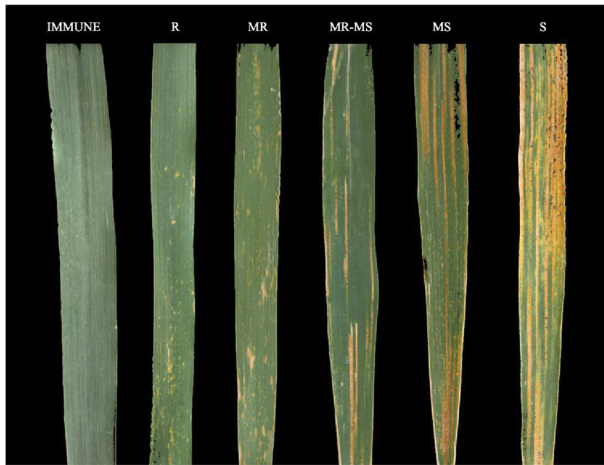
Fig. 2 Sample leaf image of each class [19]

pre-processing. Finally, Yellow-Rust-19 was created by the labeling and data augmentation process, for details see [19]. Sample leaf images from Yellow-Rust-19 are represented in Fig. 2 and Table 2 shows the statistics for the dataset.

3.3 Color spaces

Texture, the spatial arrangement of pixels, has been one of the most popular features in image processing tasks [39]. Although grayscale textures provide enough information to solve many tasks, many researchers have started to consider color information in recent years [4, 7, 14, 15, 20]. And, in these studies, it has been observed that structural feature extraction through color texture has led to a considerable increase in sensitivity. Recall that color space is an abstract mathematical model for representing colors numerically, that is, the color space establishes a mapping between numeric values and specific colors. However, there are different color spaces for various purposes in image analysis. In this study, three different color spaces such as Red-Green-Blue (RGB), Hue-Saturation-Value (HSV), and Lightness-Red/Green-Blue/Yellow (L*a*b) were utilized for obtaining color textural features.

3.4 Feature extraction

3.4.1 Gray Level Co-occurrence Matrix (GLCM) and Color-Gray Level Co-occurrence Matrix (CGLCM) feature extraction

Visual texture (in short texture) is a feature of repeating patterns of local variations in image intensity utilized for determining objects or regions of interest in an image.

Table 2 Final dataset statistics

Immune	R	MR	MR-MS	MS	S	TOTAL
2500	2500	2500	2500	2500	2500	15000

It provides information about the spatial arrangement of colors or intensities in the image and is described by the spatial distribution of intensity levels in a neighborhood. The statistical texture, on the other hand, is a quantitative texture measure of the arrangement of intensities in a region. Gray Level Co-occurrence Matrix (GLCM) contains information about the positions of pixels having similar gray level values and, mathematically defined by,

$$G_{(dx,dy)}(i, j) = \sum_{x=1}^{n} \sum_{y=1}^{n} \begin{cases} 1, & \text{if } I(x, y) = i \text{ and } I(x + dx, y + dy) = j \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where i and j represent the pixel values, x and y represent the spatial positions in the digital image I, dx and dy represent the displacement values and define the spatial relation for which this matrix is calculated, and $I(x, y)$ represents the pixel value at the spatial

Table 3 Texture features computed from GLCM $P(i, j)$

Feature	Formula
Contrast	$\sum_{i=1}^{n} \sum_{j=1}^{n} (i - j)^2 P(i, j)$
Correlation	$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(i - \mu_x)(i - \mu_y)}{N_x N_y} P(i, j)$
Entropy	$-\sum_{i=1}^{n} \sum_{j=1}^{n} P(i, j) \log P(i, j)$
Homogeneity	$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P(i, j)}{1 + (i - j)^2}$
Inverse difference moment	$\sum_{k=0}^{n_g - 1} \frac{P_{x-y}^{(k)}}{1 + k^2}$
Sum of squares: Variance	$\sum_{i=1}^{n} \sum_{j=1}^{n} (i - \mu)^2 P(i, j)$
Sum variance	$\sum_{i=2}^{2n_g} (i - SA)^2 P_{x+y}(i)$
Sum average	$\sum_{i=2}^{2n_g} [i P_{x+y}(i)]$
Sum entropy	$-\sum_{i=2}^{2n_g} P_{x+y}(i) \log_2[P_{x+y}(i)]$
Difference variance	$\sum_{i=0}^{2n_g - 1} i^2 P_{x+y}(i)$
Difference entropy	$\sum_{i=0}^{2n_g - 1} P_{x+y}(i) \log_2[P_{x+y}(i)]$
Information measure of correlation 1	$\frac{HXY - HXY1}{\max(HX, HY)}$
Information measure of correlation 2	$\sqrt{(1 - \exp[-2(HXY2 - HXY)])}$
Maximum correlation coefficient	$\sqrt{\sum_k \frac{P(i,k)P(j,k)}{P_x(i)P_y(k)}}$
Energy	$\sum_{i=1}^{n} \sum_{j=1}^{n} P(i, j)^2$
Dissimilarity	$\sum_{i=1}^{n} \sum_{j=1}^{n} P(i, j)|i - j|$
Cluster shade	$\sum_{i=1}^{n} \sum_{j=1}^{n} [i + j - \mu_x - \mu_y]^2 P(i, j)$
Cluster prominence	$\sum_{i=1}^{n} \sum_{j=1}^{n} [i + j - \mu_x - \mu_y]^4 P(i, j)$
Maximum probability	$\max(P(i, j))$
Inverse difference normalized	$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P(i, j)}{1 + \frac{|i - j|^2}{n^2}}$
Inverse difference moment normalized	$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P(i, j)}{1 + \frac{(i - j)^2}{n^2}}$
Autocorrelation	$\sum_{i=1}^{n} \sum_{j=1}^{n} (ij) P(i, j)$

position (x, y). In words, $G_{(dx,dy)}(i, j)$ counts all pairs of pixels having gray values i and j occur within (dx, dy) neighbors of $I(x, y)$. We observe that $G_{(dx,dy)}$ is an n-by-n matrix. However, the normalized GLCM is defined by

$$P(i, j) = \frac{G_{(dx,dy)}(i, j)}{\sum_{i=1}^{n} \sum_{j=1}^{n} G_{(dx,dy)}(i, j)} \tag{2}$$

where the entries are between 0 and 1.

Haralick et al. [18] proposed a set of 14 textural features that can be inferred from the co-occurrence matrix given at (2) in 1979. The set contains information about image textural characteristics such as homogeneity, linearity, and contrast. Later on, additional eight textural features were introduced [8, 17, 43]. Some of these features have intuitive and familiar definitions, for example, Energy is the second moment of $P(i, j)$. The mathematical definitions of features are listed at Table 3.

On the other hand, for colored images, relatively little is known about the relationship between color information and GLCM textural features. It is known that GLCM textural features tend to be globally adaptable but not locally optimized. Shearer [42] suggested the use of color texture analysis to overcome the shortcomings of traditional gray-level texture analysis. Within the scope of this study, all textural features listed in Table 3 were obtained from each color image, i.e. 22 features for each channel. Moreover, each image was converted to grayscale, and the similar feature set was obtained for each grayscaled image. Thus, in total, 88 features were contributed to the dataset by each image.
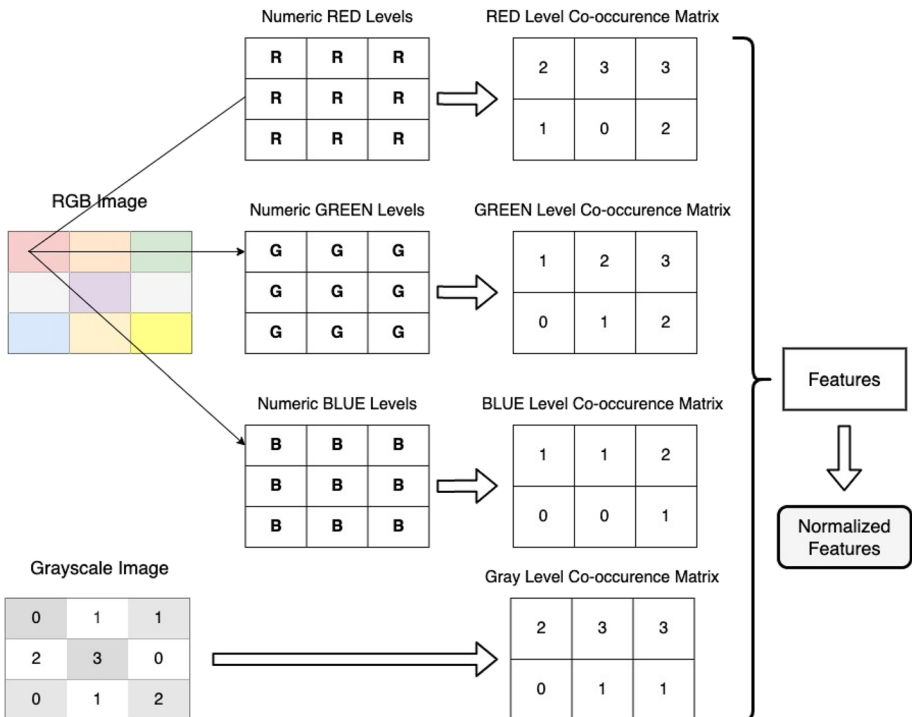

Fig. 3 A Sample representation for CGLCM_RGB

Furthermore, each leaf was represented in three different color spaces such as RGB, HSV, and L*a*b, making three different textural datasets corresponding to each image color space. Here is a good place to point out that each image was resized to 224 × 224 before the textural feature processing was applied, as the DenseNet network converts the input images to 224 × 224 before any other processing. Finally, each feature was normalized using the min-max normalization method to reduce the classification time. Figure 3 depicts textural features for an image in RGB color space.

As a result of all these processes, three different feature matrices, corresponding to each color space, with 15.000 rows, corresponding to each image, and 88 columns, corresponding to each textural feature, were obtained for the dataset *Yellow-Rust-19*.

3.4.2 DenseNet-201 feature extraction

DenseNet (Dense Convolutional Network) [21] is a convolutional neural network with dense connection, fewer parameters, but high accuracy. Huang et. al. [21] introduced DenseNet because of the observation that convolutional networks could be substantially deeper, more accurate, and efficient to train if they contained shorter connections between layers close to the input and those close to the output. DenseNet concatenates the outputs of the previous layer and the future layer, meaning each layer is directly connected to every other layer in a feed forward structure within each dense block. DenseNet is one of the new technologies
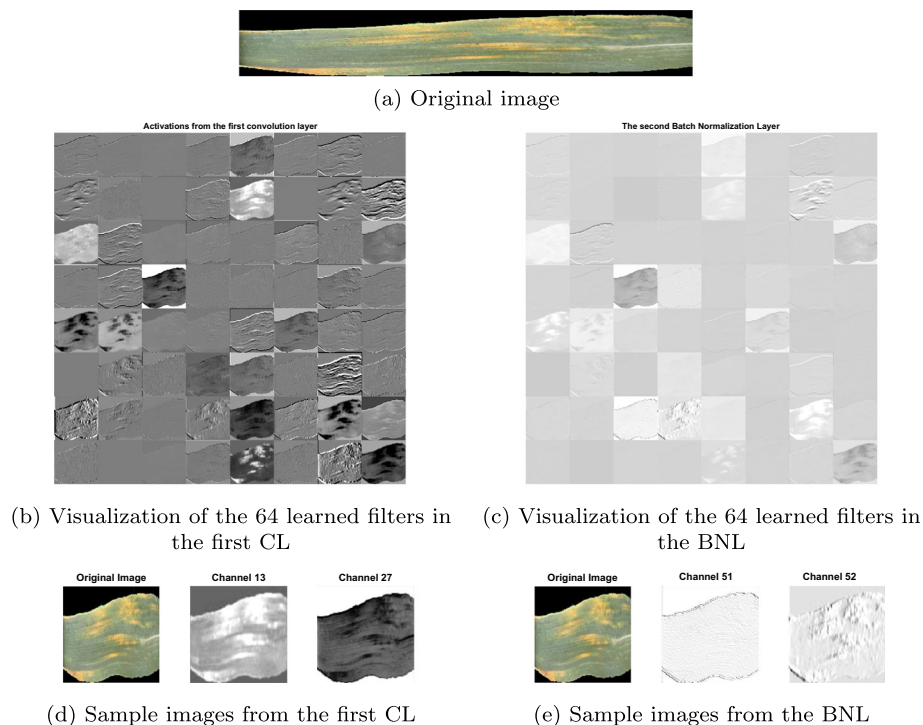

(a) Original image

(b) Visualization of the 64 learned filters in the first CL

(c) Visualization of the 64 learned filters in the BNL

(d) Sample images from the first CL

(e) Sample images from the BNL

Fig. 4 Visualization of a sample leaf image from the first Convolution Layer (CL) and the Batch Normalization Layer (BNL) in DenseNet-201 architecture

in deep learning for visual object recognition and has different versions such as DenseNet-121, DenseNet-160, DenseNet-201 and DenseNet-264. DenseNet accepts 224-by-224-by-3 input images. DenseNet diminishes the vanishing-gradient problem, consolidates feature propagation, promotes feature reuse, and considerably reduces the number of parameters. Herein, DenseNet-201 was preferred for deep features extraction.

DenseNet-201 is a convolutional neural network that is 201 layers deep. DenseNet-201 consists of four dense blocks with varying number of layers [6,12,48,32]. In each dense block, each layer puts some additional features on top of the existing feature maps. Inside the dense blocks, the feature maps' size remains the same. DenseNet-201 was trained on the ImageNet database [10]. Like other convolutional neural networks, DenseNet-201 network creates a hierarchical representation of the input images. Deeper layers contain higher-level features created using lower-level features of previous layers. Figure 4 shows a visualization of a sample leaf image from the first Convolution Layer (CL) and the Batch Normalization Layer (BNL) in DenseNet-201 architecture. Features extracted at the Global Average Pooling (GAP) layer at the end of the network were utilized in the classification phase. The GAP provided 1920 features in total. See Table 4 for architecture details.

And, Table 5 presents the hyperparameters utilized in training the model.

3.5 Classification

This study aimed to investigate the feasibility of using only textural features of colored images compared to the combination of textural features of colored images and DenseNet-201 deep features to determine the severity level of yellow rust disease in wheat. The dataset was labeled, meaning each data was attached with the severity level of yellow rust disease. There are six severity levels such that Immune, R, MR, MR-MS, MS, and S. Thus, it is a multi-class classification task. Many multi-class classification methods exist in the literature, however, Multiclass Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) were utilized in the classification phase. Indeed, during experimental studies, other

Table 4 DenseNet-201 architecture

Layers	Output size	Description
Convolution	112×112	7×7 conv, stride 2
Pooling	112×112	7×7 conv, stride 2
Dense block 1	56×56	7×7 conv, stride 2
Transition layer	56×56	7×7 conv, stride 2
	28×28	7×7 conv, stride 2
Dense block 2	112×112	7×7 conv, stride 2
Transition layer	112×112	7×7 conv, stride 2
	112×112	7×7 conv, stride 2
Dense block 3	112×112	7×7 conv, stride 2
Transition layer	112×112	7×7 conv, stride 2
	112×112	7×7 conv, stride 2
Dense block 4	112×112	7×7 conv, stride 2
Classification layer	112×112	7×7 conv, stride 2
	112×112	7×7 conv, stride 2

Table 5 DenseNet-201 hyperparameters

Parameter	Value
Optimizer	Adam
Loss function	Categorical cross entropy
Momentum	0.9
Learning rate	1.0000e-03
Early stopping patience	10
Maximum epoch	50
Mini batch size	32
Shuffle	Every epoch

classifiers such as Artificial Neural networks, Decision trees were also tested, but SVM and KNN performed better than others, thus, the current study presents results by SVM and KNN.

3.5.1 Support vector machines

SVM is a widespread technique in texture classification tasks [29], was employed in many texture classification studies [5, 24, 46].

SVM seeks to find an optimal boundary between the possible labels. It is based on maximizing the minimum distance from the separating hyperplane to the nearest label. While the fundamental SVM only supports binary classification, its extensions are designed to handle the multiclass classification situation as well. To distinguish K classes, the classifier can use multiple SVMs, each SVM predicting membership in one of the K classes. The idea is to map data points to the hyperplane to achieve mutual linear separation between every pair of classes. There are two approaches such that *One-to-One approach* and *One-to-Rest*. The current study utilized One-to-One approach, which puts a hyperplane to separate every two classes, neglecting the data points of the other classes, thus, it uses $\frac{K(K-1)}{2}$ SVMs. In this case, the classifier used 15 SVMs. The parameters for the multiclass SVM classifer is outlined at Table 6.

3.5.2 K-Nearest neighbors

KNN, a widely used and non-parametric classification method, relies on visible pattern similarities and distance metrics to generate precise predictions, in short, it is a similarity measure. It is frequently utilized in texture classification problems [16] since patterns

Table 6 Multiclass SVM hyperparameters

Parameter	Value
Optimizer	Bayesian
Kernel	Cubic polynomial
Max. Iteration	30
Multiclass Method	One vs One
Number of SVMs	15

Table 7 Training and test sets statistics

Groups	Immune	R	MR	MR-MS	MS	S	Total
Training	2250	2250	2250	2250	2250	2250	12250
Test	250	250	250	250	250	250	750
Total	2500	2500	2500	2500	2500	2500	15000

close to each other in the feature space most likely belong to the same pattern class. KNN estimates the likelihood that a data point is a member of one group or another based on which group the data points closest in distance belong to. There are different distance metrics such as Euclidean distance, Manhattan distance, Minkowski distance, etc. employed with KNN, however, Euclidean distance is the most accepted metric and, for the data points $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and $\mathbf{y} = (y_1, y_2, \cdots, y_n)$, it is defined as:

$$\text{Euclidean distance}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} x_i^2 - y_i^2}. \tag{3}$$

On the other hand, The k value in KNN classifier represents the number of nearest neighbors to be included in groups during the voting process. The selection of k is crucial and depends upon the data, however, generally, while large k values reduce the effect of the noise on the classification, they may cause boundaries between classes less distinct.

The current study uses Euclidean distance to measure the distance between two data points and takes k as 10. Moreover, it has used inverse distance weighting to assign higher weights to the closest training samples.

4 Results

The study investigates the feasibility of using only colored images' textural features compared to the combination of colored images' textural features and DenseNet-201 deep features to determine the severity level of yellow rust disease in wheat, among one of six severity levels such that Immune, R, MR, MR-MS, MS, and S. Initial dataset was *Yellow-Rust-19*, which contains 15000 wheat leaf images with varying size. Initially, each image was converted to the dimension of 224 × 224. Then, the dataset was randomly split into two parts using the stratified selection strategy that guarantees an even distribution of images from each class to allocate 90% training samples for model development and a 10% test sample for the model's generalization evaluation. Table 7 shows the training set and test set numbers.

The implementation, as well as training of the proposed transfer model, was carried out using both Python and MATLAB$^\circledR$ programming environments.

For each leaf image, 66 textural colored features were obtained corresponding to each color space as RGB, HSV, and L*a*b, making three different colored textural features datasets. Moreover, for each image 22 gray level textural features were also obtained from each grayscale leaf image. Furthermore, for each image 1920 DenseNet-201 deep features provided by the GAP layer were obtained.

Different models on different datasets with different classifiers were developed. A total of twelve models, six different multiclass models for each SVM and KNN classifier, were developed and evaluated, six of them, three for each classifier, were trained using only

Table 8 Datasets, color spaces

Dataset name	Color Space	Dataset description
CGLCM_RGB	RGB	22 Grayscale plus 66 Color features
CGLCM_HSV	HSV	22 Grayscale plus 66 Color features
CGLCM_Lab	L*a*b	22 Grayscale plus 66 Color features
CNN-CGLCM_RGB	RGB	1920 DenseNet-201 plus 22 Grayscale plus 66 Color features
CNN-CGLCM_HSV	HSV	1920 DenseNet-201 plus 22 Grayscale plus 66 Color features
CNN-CGLCM_Lab	L*a*b	1920 DenseNet-201 plus 22 Grayscale plus 66 Color features

texture features, and six of them, again three for each classifier, were trained using both textural and deep features. Table 8 shows the datasets' names and corresponding color spaces.

Among models trained using texture features, the SVM classifier with an accuracy of 73.1% performed best via CGLCM_RGB feature set, however, the model using SVM classifier trained on the feature set CNN-CGLCM_HSV had the best accuracy with 92.4%. Table 9 presents the methods, corresponding datasets, and their accuracies. Observe from Table 9 that models trained using both textural and deep features outperformed models trained using textural features alone, moreover, the models trained by the dataset CNN-CGLCM_HSV for both SVM and KNN classifiers performed better.

On the other hand, Fig. 5 shows the confusion matrices for the models trained on the dataset CNN-CGLCM_HSV, with the classifier SVM on the left and KNN on the right. The confusion matrix reveals how a classifiers performed in each class. 0 represents "Immune", number 1 represents "R", 2 represents "MR", 3 represents "MR-MS", 4 represents "MS" and 5 represents "S" class. A close look at confusion matrices reveals that prediction to MS class was poor by both models.

The test set confusion matrix across all runs was recorded, and the following three metrics were also calculated: recall, precision and F1-score are demonstrated in Table 10. Among the textural features, observe that the models trained on CGLCM_RGB dataset perform better than others. Indeed, models trained on combined features considerably outperform models trained on textural features alone. The model trained on CNN-CGLCM_HSV dataset, with the SVM classifier achieves the best results in all performance criteria.

Table 11 presents the class performance values of the best model trained on the CNN-CGLCM_HSV dataset and using the SVM classifier. A closer look at the table reveals that the model performs poorly in distinguishing MR-MS and MS classes compared to the others.

Table 9 The accuracy rates (%)

Dataset name	Accuracy by SVM	Accuracy KNN
CGLCM_RGB	73.1	72.5
CGLCM_HSV	69.5	67.8
CGLCM_Lab	61.8	67.5
CNN-CGLCM_RGB	90.7	88.6
CNN-CGLCM_HSV	92.4	88.1
CNN-CGLCM_Lab	86.9	86.6

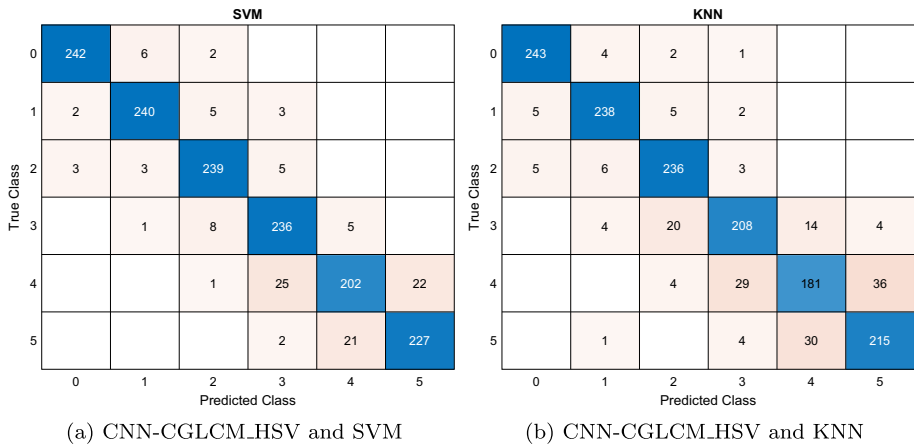(a) CNN-CGLCM_HSV and SVM   (b) CNN-CGLCM_HSV and KNN

Fig. 5 The Confusion matrix for the models trained on the dataset CNN-CGLCM_HSV, on the left by the classifier SVM and on the right by KNN

5 Discussion

From the experiment results, we observed that the combined features have better accuracy rates than traditional textural analysis methods. The dataset which represents healthy wheat leaves and diseased wheat leaves with different types of infection (Yellow-Rust-19) was used in all methods. The best performance for combined features was calculated as 92.4% for HSV color space. In addition, it was observed that average accuracy rate increase for all color spaces. The same SVM parameters were used for all classifications for fair comparison.

The popular texture feature extraction techniques are stated in the literature [29]. The GLCM is a robust method for statistical texture analysis [28, 34, 43] and is an effective texture identifier. It is also proven that it has better accuracy value and computational time than other texture analysis methods [9]. Therefore, the authors preferred the GLCM for texture feature extraction. In addition, it is stated that GLCM can be improved to be applied on different color spaces [29, 41, 45]. This paper also supports this view.

Table 10 Performance comparison of the models

Dataset	SVM			KNN		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
CGLCM_RGB	0.73	0.74	0.73	0.73	0.72	0.72
CGLCM_HSV	0.70	0.69	0.70	0.68	0.68	0.67
CGLCM_Lab	0.62	0.62	0.62	0.69	0.67	0.67
CNN-CGLCM_RGB	0.91	0.91	0.91	**0.89**	**0.89**	**0.88**
CNN-CGLCM_HSV	**0.93**	**0.93**	**0.93**	0.88	0.88	**0.88**
CNN-CGLCM_Lab	0.87	0.87	0.87	0.87	0.87	0.87

Bold fonts in Table emphasize the highest value

Table 11 Classwise performance comparison for the CNN-CGLCM_HSV dataset

Class	SVM				KNN			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Immune	99.13	0.97	0.98	0.97	98.87	0.97	0.96	0.97
R	98.67	0.96	0.96	0.96	98.20	0.95	0.94	0.95
MR	98.20	0.96	0.94	0.95	97.00	0.94	0.88	0.91
MR-MS	96.73	0.94	0.87	0.91	94.60	0.83	0.84	0.84
MS	95.07	0.81	0.89	0.85	92.47	0.72	0.80	0.76
S	97.00	0.91	0.91	0.91	95.00	0.86	0.84	0.85

Although, there is a performance improvement in all color spaces, confusion matrices showed there were differences in the performance obtained for each class. This deficiency must be taken into account since the statistics of the confusion matrices were similar in all color spaces. However, deep learning methods can be helpful for increasing the performance of textural features and creating a more stable system. Although performance tests prove the performance of combined methods, better classification methods should be tested to use the features in real conditions.

Transfer learning (TL) as a deep learning approach, allows an existing pre-trained model on a large-scale dataset for one task on a domain to be used as a starting point for another task on a similar domain with relatively less labeled data. TL-based models need less training time, compared to models trained from scratch [30].

Wang and Zhang [52], reported that DenseNet surpasses other pre-trained models based on the comparison results of popular pre-trained models. Also, in the same study, Wang and Zhang have proved that DenseNet-201 is more successful than other DenseNet versions. Therefore, it was preferred DenseNet-201 for deep feature extraction.

This paper, discusses the feasibility of combination of texture features and deep features to determine the severity level of yellow rust in wheat. Although textural and deep features were combined in this study, the expected high classification performance could not be achieved in the MS class, which contains similar samples to MR-MS and S classes. On the other hand, even though, the Yellow-Rust-19 dataset was labeled by two experts via

Table 12 Performance comparison for the combined features (CF) against the standard texture features (STF)

Work	M	TA Method	DL Method	Classifier	STF	CF
Proposed Work	A	Color-based GLCM	DenseNet-201	SVM	72.1	92.4
Zhang et al. [56]	A	LBP	VGG-19, VGG-f, Caffe-ref	Ensemble	71.29	85.47
Wei et al. [53]	A	NPPS	Deep Belief Network	LR	76.70	79.23
Nanni et al. [32]	A	Various methods	Pre-trained DenseNet-201	SVM	86.13	89.47
Luz et al. [27]	A	Various methods	LeNet	SVM	71.2	85
Zhang et al. [57]	A	Various methods	VGG-16	SVM	96.3	98.8
Naz et al. [33]	A	LBP and SFTA	Inception-V3 and VGG-16	SVM	59.2	99.3

Note: Best results of the studies were represented; M represents Metric; TA represents textural Analysis; DL represents Deep Learning; A represents Accuracy; R represents Recall; LR represents Logistic Regression; NPPS represents Nucleotide Pair Position Specificity; GMMs represents Gaussian Mixture Models.

consensus to create a gold standard (reference standard), it is still possible that mislabeling happens and the performance of the proposed method might be degraded.

Finally, the results of this paper and related works which research combined features are presented in Table 12. As a result, the contribution of deep features to classification performance is evident in other studies, as in this study.

6 Conclusion

Throughout the study, a hybrid method based on combining deep and textural features was proposed in a six-way multi-label classification task for the Yellow-Rust-19 dataset. Deep features were obtained by the pre-trained network DenseNet. Different models on different datasets with different classifiers were developed. Test accuracy, as well as other performance metrics on the test set, such as precision, recall, F1 score, indicate that combined methods are promising (Table 10). Although the textural features alone are not sufficient, it has been proved that it improves sensitivity when combined with deep features. In future studies, the results can be compared by testing different pre-trained networks other than DenseNet. In addition, other texture analysis methods other than CGLCM used in this study can be tested and their contribution to success can be discussed.

Author Contributions TH, HE and FV designed the hybrid models and analysed the data. TH, FV and NA collected images. FH and NA labeled images and supervised the experiments. All autors wrote the manuscript and read and approved the final manuscript.

Data Availability All data generated or analysed during this study are specified in this published article (for details see Section 3.2). The data have been deposited in the Kaggle database (https://www.kaggle.com/datasets/tolgahayit/yellowrust19-yellow-rust-disease-in-wheat). Requests for material should be made to the corresponding authors.

References

1. Abayomi-Alli OO, Damaševičius R, Misra S, Maskeliūnas R (2021) Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. Expert Syst 38(7):e12746
2. Aksoy S, Haralick RM (2000) Using texture in image similarity and retrieval. In: Texture analysis in machine vision, pp 129–149. World scientific
3. Almadhor A, Rauf HT, Lali MIU, Damaševičius R, Alouffi B, Alharbi A (2021) Ai-driven framework for recognition of guava plant diseases through machine learning from dslr camera sensor based high resolution imagery. Sensors 21(11):3830
4. Benco M, Hudec R, Kamencay P, Zachariasova M, Matuska S (2014) An advanced approach to extraction of colour texture features based on glcm. Int J Adv Robot Syst 11(7):104
5. Chakraborty S, Paul S, Rahat-uz Zaman M (2021) Prediction of apple leaf diseases using multiclass support vector machine. In: 2021 2Nd international conference on robotics, electrical and signal processing techniques (ICREST), pp 147–151. https://doi.org/10.1109/ICREST51555.2021.9331132
6. Chen X (2020) Pathogens which threaten food security: Puccinia striiformis, the wheat stripe rust pathogen. Food Secur 12(2):239–251
7. Choi JY, Ro YM, Plataniotis KN (2011) Color local texture features for color face recognition. IEEE Trans Image Process 21(3):1366–1380

8. Clausi DA (2002) An analysis of co-occurrence texture statistics as a function of grey level quantization. Can J Rem Sens 28(1):45–62
9. De Siqueira FR, Schwartz WR, Pedrini H (2013) Multi-scale gray level co-occurrence matrices for texture description. Neurocomputing 120:336–345
10. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on computer vision and pattern recognition, pp 248–255. IEEE
11. Devadas R, Lamb D, Simpfendorfer S, Backhouse D (2009) Evaluating ten spectral vegetation indices for identifying rust infection in individual wheat leaves. Precis Agric 10(6):459–470
12. Dubey N, Bhagat E, Rana S, Pathak K (2023) A novel approach to detect plant disease using densenet-121 neural network. In: Smart trends in computing and communications, pp 63–74. Springer
13. Dusunceli F, Cetin L, Albustan S, Beniwal S (1996) Occurrence and impact of wheat stripe rust (puccinia striiformis) in Turkey in 1994/95 crop season. In: Proceedings 9th European and mediterranean cereal rusts and powdery mildews conference,. Lunteren, the Netherlands, p 309
14. Fekriershad S, Tajeripour F (2017) Color texture classification based on proposed impulse-noise resistant color local binary patterns and significant points selection algorithm. Sensor Review
15. Gui W, Liu J, Yang C, Chen N, Liao X (2013) Color co-occurrence matrix based froth image texture extraction for mineral flotation. Miner Eng 46:60–67
16. Guru D, Sharath Y, Manjunath S (2010) Texture features and knn in classification of flower images. IJCA, Special Issue on RTIPPR (1) pp 21–29
17. Haralick RM (1979) Statistical and structural approaches to texture. Proc IEEE 67(5):786–804
18. Haralick RM, Shanmugam K, Dinstein IH (1973) textural features for image classification. IEEE Trans Syst Man Cybernet SMC 3(6):610–621
19. Hayit T, Erbay H, Varċın F, Hayit F, Akci N (2021) Determination of the severity level of yellow rust disease in wheat by using convolutional neural networks. J Plant Pathol 103(3):923–934
20. Hossain K, Parekh R (2010) Extending GLCM to include color information for texture recognition. In: AIP Conference proceedings, vol 1298, pp 583–588. American Institute of Physics
21. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
22. Johnston C, Browder L (1964) Seventh revision of the international register of physiologic races of puccinia recondita f. sp tritici
23. Khaldi B, Aiadi O, Kherfi ML (2019) Combining colour and grey-level co-occurrence matrix features: a comparative study. IET Image Process 13(9):1401–1410
24. Kim KI, Jung K, Park SH, Kim HJ (2002) Support vector machines for texture classification. IEEE Trans Pattern Anal Mach Intell 24(11):1542–1550. https://doi.org/10.1109/TPAMI.2002.1046177
25. Kundu N, Rani G, Dhaka VS, Gupta K, Nayak SC, Verma S, Ijaz MF, Woźniak M (2021) Iot and interpretable machine learning based framework for disease prediction in pearl millet. Sensors 21(16):5386
26. Lu J, Zhou M, Gao Y, Jiang H (2018) Using hyperspectral imaging to discriminate yellow leaf curl disease in tomato leaves. Precis Agric 19(3):379–394
27. Luz JS, Oliveira MC, Araujo FH, Magalhães DM (2021) Ensemble of handcrafted and deep features for urban sound classification. Appl Acoust 175:107819
28. Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 18(8):837–842
29. Metre V, Ghorpade J (2013) An overview of the research on texture based plant leaf classification. arXiv preprint arXiv: 1306.4345
30. Mohanty SP, Hughes DP, Salathé M. (2016) Using deep learning for image-based plant disease detection. Front Plant Sci 7:1419
31. Nandhini S, Ashokkumar K (2022) An automatic plant leaf disease identification using densenet-121 architecture with a mutation-based henry gas solubility optimization algorithm. Neural Comput Appl 34(7):5513–5534
32. Nanni L, De Luca E, Facin ML, Maguolo G (2020) Deep learning and handcrafted features for virus image classification. J Imaging 6(12):143
33. Naz J, Sharif M, Raza M, Shah JH, Yasmin M, Kadry S, Vimal S (2021) Recognizing gastrointestinal malignancies on wce and cce images by an ensemble of deep and handcrafted features with entropy and pca based features optimization. Neural Process Lett 55:115–140
34. Nikoo H, Talebi H, Mirzaei A (2011) A supervised method for determining displacement of gray level co-occurrence matrix. In: 2011 7Th iranian conference on machine vision and image processing, pp 1–5. IEEE
35. Ning X, Tian W, Yu Z, Li W, Bai X, Wang Y (2022) Hcfnn: high-order coverage function neural network for image classification. Pattern Recogn 108873:131

36. Ning X, Xu S, Nan F, Zeng Q, Wang C, Cai W, Li W, Jiang Y (2022) Face editing based on facial recognition features IEEE Transactions on Cognitive and Developmental Systems
37. Pardey PG, Beddow J, Kriticos D, Hurley T, Park R, Duveiller E, Sutherst R, Burdon J, Hodson D (2013) Right-sizing stem-rust research. Science 340(6129):147–148
38. Peterson RF, Campbell A, Hannah A (1948) A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. Can J Res 26(5):496–500
39. Pietikainen MK (2000) Texture analysis in machine vision, vol 40. World Scientific
40. Roelfs AP (1992) Rust diseases of wheat: concepts and methods of disease management Cimmyt
41. Selvarajah S, Kodituwakku SR (2011) Analysis and comparison of texture features forcontent based image retrieval. Int J Latest Trends Comput 2(1):108–113
42. Shearer SA (1986) Plant identification using color co-occurrence matrices derived from digitized images, texture, pattern recognition. The Ohio State University
43. Soh LK, Tsatsoulis C (1999) Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. IEEE Trans Geosci Remote Sens 37(2):780–795
44. Srinidhi V, Sahay A, Deeba K (2021) Plant pathology disease detection in apple leaves using deep convolutional neural networks: Apple leaves disease detection using efficientnet and densenet. In: 2021 5Th international conference on computing methodologies and communication (ICCMC), pp 1119–1127. IEEE
45. Tou JY, Tay YH, Lau PY (2009) Recent trends in texture classification: a review. In: Symposium on progress in information & communication technology, vol 3, pp 56–59. Citeseer
46. Vapnik V (1999) The nature of statistical learning theory Springer science & business media
47. Varcin F, Erbay H, Cetin E, Cetin I, Kultur T (2021) End-to-end computerized diagnosis of spondylolisthesis using only lumbar X-rays. J Digit Imaging 34(1):85–95
48. Vardhan MH, Rao SV (2014) Glcm architecture for image extraction. Int J Adv Res Electron Commun Eng 3:75–82
49. Verma R, Singh V (2021) Leaf disease identification using densenet. In: International conference on artificial intelligence and speech technology, pp 500–511. Springer
50. Wang C, Ning X, Sun L, Zhang L, Li W, Bai X (2022) Learning discriminative features by covering local geometric space for point cloud analysis. IEEE Trans Geosci Remote Sens 60:1–15
51. Wang C, Wang X, Zhang J, Zhang L, Bai X, Ning X, Zhou J, Hancock E (2022) Uncertainty estimation for stereo matching based on evidential deep learning. Pattern Recogn 108498:124
52. Wang SH, Zhang YD (2020) Densenet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16(2s):1–19
53. Wei L, Su R, Wang B, Li X, Zou Q, Gao X (2019) Integration of deep feature representations and handcrafted features to improve the prediction of n6-methyladenosine sites. Neurocomputing 324:3–9
54. Wellings CR (2011) Global status of stripe rust: a review of historical and current threats. Euphytica 179(1):129–141
55. Yurttakal AH, Erbay H, İkizceli T, Karacavus S, Çinarer G (2018) A comparative study on segmentation and classification in breast mri imaging. IIOAB J 9(5):23–33
56. Zhang J, Xia Y, Xie Y, Fulham M, Feng DD (2017) Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. IEEE J Biomed Health Inform 22(5):1521–1530
57. Zhang W, Pogorelsky B, Loveland M, Wolf T (2021) Classification of covid-19 x-ray images using a combination of deep and handcrafted features. arXiv preprint arXiv: 2101.07866
58. Zhou W, Newsam S, Li C, Shao Z (2018) Patternnet: a benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS J Photogramm Remote Sens 145:197–209

Affiliations

Tolga Hayıt[1] · Hasan Erbay[2,3] · Fatih Varçın[4] · Fatma Hayıt[5] 🆔 **· Nilüfer Akci[6]**

Tolga Hayıt
tolga.hayit@bozok.edu.tr

Hasan Erbay
hasan_erbay@yahoo.com;herbay@thk.edu.tr

Fatih Varçın
fatihvarcin@subu.edu.tr

Nilüfer Akci
nilufer.akci@tarimorman.gov.tr

[1] Department of Computer Engineering, Faculty of Engineering and Architecture, Yozgat Bozok University, Yozgat, 66900, Türkiye

[2] Computer Engineering Department, Engineering Faculty, University of Turkish Aeronautical Association, 06790 Etimesgut Ankara, Türkiye

[3] Computer Engineering Department, Engineering Faculty, Ostim Technical University, 06374 Ostim Ankara, Türkiye

[4] Department of Computer Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya, 54187, Türkiye

[5] Department of Gastronomy and Culinary Arts, Tourism Faculty, Yozgat Bozok University, Yozgat, 66900, Türkiye

[6] Directorate of Plant Protection Central Research Institute, Republic of Türkiye Ministry of Agriculture and Forestry, Ankara, 06172, Türkiye