Journal of Agriculture and Food Research 14 (2023) 100658

Contents lists available at ScienceDirect

Journal of Agriculture and Food Research

journal homepage: www.sciencedirect.com/journal/journal-of-agriculture-and-food-research

Species classification of brassica napus based on flowers, leaves, and packets using deep neural networks

Munjur Alom [a], Md. Yeasin Ali [a], Md. Tarequl Islam [a], Abdul Hasib Uddin [a], Wahidur Rahman [b,c,*]

[a] *Department of Computer Science and Engineering, Khwaja Yunus Ali University, Enayetpur, Sirajganj, 6751, Bangladesh*
[b] *Department of Computer Science and Engineering, Uttara University, Dhaka, Bangladesh*
[c] *Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Tangail, 1902, Bangladesh*

A R T I C L E I N F O

Keywords:
Brassica napus
Machine learning
Classification
Flower
Leaves
Convolutional neural network

A B S T R A C T

Deep learning (DL) has gradually taken the lead as the most effective approach in the agricultural fields due to the early identification and classification of plant species and diseases for improving the quality of crop production because of recent technological breakthroughs, which have had a significant impact on agriculture. Plenty of complicated problems in farming, including species classification, plant disorder identification, yield approximation, and weather and soil moisture prediction, are made simple using deep neural networks. Thus, this proposed study aims to classify Brassica Napus (B. Napus) rapeseed species based on their most significant features, like flowers, leaves, and packets. The study has adopted two types of rapeseed such as B. Rapa and B. Alba. Five contemporary deep learning-based Convolutional Neural Network (CNN) models have also been assessed for distinguishing rapeseed species. These models are DenseNet201, VGG19, InceptionV3, Xception, and ResNet50. Initially, the researchers collected data from the agricultural field, and then image pre-processing is performed to create our dataset. After that, CNN models were applied to this dataset and enumerated the experimental data accordingly. Our DenseNet201 model successfully classified both species with the highest accuracy of 100% for flowers and 97% for both packets and leaves. A comprehensive analysis with companion studies confirmed the efficacy of our preferred paradigm for the near future. Nevertheless, future studies will compare these methodologies to data from a separate metabolomics dataset from comparable crops.

1. Introduction

B. Napus is the most crucial oilseed grain globally, called canola, rapeseed, and mustard. More than three hundred genera and three thousand species are present [1]. B. Rapa and B. Alba are the two essential variations cultivated in Bangladesh [2]. In recent years, agriculture manufacturing has become a significant industry for introducing large-scale products with food, health, and commercial goods. Due to the growing volume in this industry, gathering knowledge and including processing techniques for diverse agricultural implementations is necessary. Professionals in the agricultural industry are essential in decision-making and delivering useful information, yet they are frequently unreachable when help is required. The architectural aspects of rapeseed, like those of other crops, are crucial for designing a device for managing, harvesting, searing, assembling, aerating, producing oil, and purifying [3]. Expert schemes give a hopeful potential experience in agricultural growth. Because seeds are used in many places for rapeseed processing and insemination, even specialists can mistakenly classify variants. As a result, it is important to identify varieties quickly and to make a correct choice. To expedite the strategies, a CNN-based expert machine for classifying diversity could desire to ensure a difference.

Rapeseed is produced worldwide in more than 75 million tons annually [6] for animal feed production, ingestible herbs for human consumption, and the industrial component of biodiesel. Among them, 38% of canola oil is used for commercial purposes, 3% for feed, and roughly 60% for dietary purposes [1]. Several important secondary components include free fatty acids, phosphatides (gum), enzymes (particularly myrosinase), and glucosinolates. Erucic acid is essential for the quality of the oil, and glucosinolate is vital for providing the meal's perfection [3]. Rapeseed is the second-largest oil grain globally. It

* Corresponding Author. Department of Computer Science and Engineering, Uttara University, Dhaka, Bangladesh.
E-mail addresses: munjuralom.cse@gmail.com (M. Alom), mdyeasinalicse@gmail.com (Md.Y. Ali), tareq.cse@gmail.com (Md.T. Islam), abdulhasibuddin@gmail.com (A.H. Uddin), wahidtuhin0@gmail.com (W. Rahman).

https://doi.org/10.1016/j.jafr.2023.100658
Received 30 November 2022; Received in revised form 20 May 2023; Accepted 27 May 2023
Available online 10 June 2023

belongs to the *Brassicaceae* family. B. Rapa and B. Alba are the most important variations. These two variations are cultivated in the agriculture field in Bangladesh. These species are classified based on their visible characteristics, such as flowers, leaves, and packets. It is challenging for ordinary people to identify the diversity of rapeseed in the agriculture sector. Identifying B. Napus diversities from images established on visual traits using computer vision and deep learning technology is an exciting problem for researchers. However, the classifier's overall performance depends on the quality of the trained image dataset. An image dataset is essential for testing the identification and recognition techniques.

In previous studies, deep neural network methods were used to classify B. Napus rapeseed species. To distinguish between transgenic B. Napus and their hybrids, deep convolutional neural network [1] techniques are applied in conjunction with Vis-NIR spectroscopy. This study used a deep convolutional neural network, a neural network commonly used in image classification tasks, and visible-near infrared (Vis-NIR) spectroscopy to differentiate between transgenic B. Napus and their hybrids. The advantage of this approach is that it can analyze large datasets rapidly and accurately, making it a valuable tool for screening and identifying genetically modified plants. However, the disadvantage is that it requires a significant amount of training data to achieve high accuracy, and it may only be suitable for detecting some types of genetic modifications. Non-GM and GM (genetically modified) hybrids may be distinguished by combining machine-learning methods with Vis-NIR spectroscopy; B. Juncea, B. Napus, and F1 hybrids may be determined [7]. This approach applied machine-learning methods in conjunction with Vis-NIR spectroscopy to distinguish between B. Juncea and B. Napus, as well as their F1 hybrids. The advantage of this approach is that it can be used to identify hybrid plants, which are often difficult to distinguish using traditional morphological or biochemical methods. However, the disadvantage is that it may be less accurate than other methods, such as DNA sequencing, and it requires access to specialized equipment and expertise in spectroscopic analysis. Support Vector Machine (SVM) and convolutional neural network models are used with expert system applications through computer vision to differentiate between different types of rapeseed [3]. Prediction of single and multiple-characteristics in rapeseed using multi-layer perceptron, random forests, and SVM [4]. During the flowering season, a deep convolutional neural network has been employed to identify male [8] oilseed rape plants. For the species categorization of B. Napus, this study used the five most efficient deep CNN models, including DenseNet201, InceptionV3, Xception, VGG19, and ResNet50.

Thus, the research aims to develop an intelligent system to classify the species of B. Napus based on a pre-built image dataset. The early stages of this research were to collect image data from agricultural fields to create a dataset. This research mainly focuses on the features of flowers, leaves, and packets from B. Rapa and Rapa to build the dataset. Then, the study utilized Deep CNN architectures for training and testing to extract each induvial image's features. After that, a set of Machine Learning (ML) classifiers was applied to the model and compared models' performance based on their assessment matrices.

2. Literature review

This section provides a background analysis of earlier research relevant to our proposed strategy. The farming industry and the B. Napus farming method have already gained tremendous contributions from many great contributors. Shon et al. [1] In their study, the authors suggested a way for differentiating between B. Napus and genetically modified (GM) B. Napus, as well as between B. Rapa and the F1 hybrids (crosses between GM B. Napus and B. Rapa). This method used Vis-NIR spectroscopy in conjunction with machine learning algorithms. With normalization, the previously mentioned classification technique with CNN produced 98.9% accuracy. Sun et al. [8] developed a methodology to recognize male oilseed rape (OSR) plants of B. Napus based on the CNN. During the flowering period of OSR, the researchers collected the images. The male and female OSR plants have few differences in morphology, color, texture, etc. The authors achieved 93.54% accuracy using the mentioned method.

Sohn et al. [7] introduced methodologies such as Vis-NIR spectroscopy associated with ML algorithms, including chemo-metrics. They separated genetically modified from non-genetically modified B. Juncea and B. Napus, as well as the F1 hybrids (crosses between B. Juncea and genetically modified B. Napus), and they kept track of every aspect of the safety of the crops and the products made from them. They established the most effective technique, Support Vector Machine (SVM), with an efficiency rate of 99.6%, while Savitzky-Golay, combined with deep learning, offered a 99.1% classification accuracy. Da Jiang et al. [9] pre-processed images with the help of image processing methods such as threshold segmentation. They merged the morphological procedures to categorize the B. Napus root disease and created a paired hidden layer of backpropagation (BP). Their approach achieved 93% accuracy, and the predicted value of the mean squared error (MSE) was 0.049.

Rahman Khan et al. [4], Lasso Regression (LR), Random Forest (RF), SVM, and Multi-Layer Perceptron are four different types of machine learning models that were used to predict and specify single or multiple features of OSR's (B. Napus) established RNA-Sequence (MLP). They included information on twelve different rapeseed phenotypic features. The MSE scores for RF and SVM were 0.045 and 0.016, respectively. The Decision Tree-based Segmentation Model (DTSM) was used to extract the flowering area by Xie et al. [10]. ML methods, including SVM, naive Bayes, RF, and K-nearest neighbors, and an unmanned aerial vehicle platform are used to assess the flowering dynamic for different genotypes (KNN). Each relative root mean square error (RMSE) score was 0.051, and their best accuracy was $R2 = 0.97$. The authors gathered information on 299 different rapeseed cultivars. K-means clustering analysis was used to group all kinds into four groupings.

Ou et al. [5] initiated YOLO-V4, an automatic object detection algorithm, to remove the problem, which created interferences and inconveniences to studying for detecting the flowering stage. To categorize the distribution range of B. Napus using CNN-based methods. The target identification technique's mean average precision (MAP) improves accuracy by 96.24%, and the average loss is only 0.5843%. Potato leaf disease was detected and classified using ML methods proposed by Singh, Kaur [11]. Segmented images to extract the features through the K-means methodology. Multi-class support vector machine used for detection and classification purposes. Their proposed model achieved 95.99% accuracy in the popularly known plant village dataset.

Features extracted via HSV color spaces and Haralick textures Xian et al. [12] The Extreme Learning Machine (ELM) classifies diseases and includes a single-layer feed-forward neural network. The proposed methodology inputs extracted features and identifies conditions from tomato leaves. ELM shows 84.94% accuracy compared to other models, such as SVM and decision trees. Naeem et al. [13] introduced ML methods for classifying medicinal plants. The processed dataset uses ML classifiers such as MLP, logit-boost, bagging, basic logistics, and RF. MLP outperforms the other classifiers in terms of accuracy (99.01%). The information was limited to six different varieties of medicinal plant leaves.

Classification of plant diseases using early disease symptoms and monitoring plant health using artificial methods in conjunction with machine learning models like RF, SVM, and neural networks, as suggested by Kanade, Prasad [14]. Their proposed plans achieved 96.22%, 95.91%, and 94.99% accuracy. Again, Abdalla et al. [15] used several pre-trained CNNs to extract features from acquired images. Long Short-Term Memory (LSTM) uses the retrieved features as input to categorize the OSR nutritional statuses into nine types. Inceptionv3 and LSTM produced excellent observations while utilizing a cross-dataset validation, with a performance of 92% accuracy, and also achieved the highest overall classification accuracy of 95%. Fusarium Oxysporum

(FO) disorder in tomato leaves employing image processing operations and ML algorithms by Dhaya., [16]. The suggested methodology uses a picture recognition system with a naive Bayes association. That performed 96% effectively in classifying FO illness. The 87,500 photos in the public collection include 60% images of afflicted leaves and 40% images of healthy plant leaves.

Sclerotinia disease in the B. Napus plant was detected by Kong et al. [17] using hyperspectral imaging with chemometrics. The machine learning techniques utilized were SVM, Soft Independent Modeling of Class Analogies (SIMCA), Partial Least Squares-Discriminant Analysis (PLS-DA), and KNN. The experiment occurred in three stages, and their predicted accuracy was 82.61%, 96.67%, and 100.00%. Yu et al. [18] Using data from fully-connected neural networks and stacked auto-encoders (SAE), a deep learning-based technique for grounded regression was used to identify and measure the nitrogen concentration in rapeseed leaves (B. Napus) (FNN). Visible near-infrared hyperspectral imaging and SAE were used to extract the in-depth feature in the OSR leaf. Their fair performance was (R2 P = 0.903, RMSEP = 0.307%, and RPDP = 3.238). The author of the paper [19] developed a neural model based on CNN, MLP, and the Radial Basis Function (RBF) network used to identify fungal contamination. There was a small % identification error of 14% for the test data, 18% for MLP, and 21% for the RBF network. CNN is used, among other methods, to detect mold in oilseeds. Their achieved performance was 0.89% with a deep CNN network and 0.85% through MLP and RBF networks Varieties of rapeseed was discriminated by Kurtulmus, Ünal [3]. Using computer vision, including grid-search, K-fold cross-validation, and SVM and CNN models. Using color images connected to the SVM classifier, 29 features effectively identified seven oilseed diversities. They achieved a 99.24% performance rate. Table 1 shows the summary of the literature reviews discussed in this section.

3. Methodology of the proposed solution

This section is classified into three significant subsections: data type specification and collection, pre-processing of collected data, and finally, the working principle of a deep learning-based framework to classify crops.

3.1. Dataset specification and collection

3.1.1. Data details

In our archive, B. Napus: an image dataset of two various oilseed rape diversities gathered by a smartphone from particular agriculture fields in rural areas of Bangladesh. The datasets by B. Napus are a handful for observing different components of identifiers, namely momentum,

Table 1
Summary of the literature reviews.

SL	References	Contribution	Methodology	Accuracy %	Limitation
1	Sohn et al. [7],	Differentiating transgenic B. Napus and their B. Juncea hybrids	Vis-NIR spectroscopy, SVM, Savitzky-Golay	99.6% and 99.1%	Rapeseed plants 20–25 days old were employed for this experiment in a lab setting.
2	Da Jiang et al. [9],	BP neural network-based prediction model	Image processing methods, BP	93% and MSE 0.049.	Enhancing the system by creating a program that automatically grades root-swelling illness as it progresses
2	Rahman Khan et al. [4],	Prediction of rapeseed traits from RNA-seq data	SVM, RF, Lasso Regression (LR), and MLP	MSE were 0.045 and 0.016	This study did not evaluate trait heritability.
4	Xie et al. [10],	Assessing rapeseed flowering dynamics for several genotypes	naive Bayes, SVM, KNN, and DTSM	R2 = 0.97% and RMSE was 0.051%	The examination of germplasm resources will require OSR flowering dates and higher-throughput phenotyping.
5	Singh, Kaur. [11],	Disease detection and classification of potato plant leaves	K-means methodology, Multi-class SVM	95.99%	Using artificial neural networks for early studies can be extended, especially CNN.
6	Xian et al. [12],	Classification of plant diseases	HSV color spaces and Haralick Textures, ELM and FNN	84.94%	It would be hopeful to think about disease identification for more plants, namely rice, corn, and wheat.
7	Naeem et al. [13],	Multispectral and texture-based classification for medicinal plant leaves	Bagging, RF, Logit-boost, MLP, and simple logistics	99.01%	Future applications of this methodology will include other therapeutic plants.
8	Kanade, Prasad. [14],	Classification and observation of plant conditions.	ANN, SVM, and RF	96.22%, 95.91%, and 94.99%	Incorporating the classifier of the neural networks with DL techniques employing more plant leaves to gather information may be beneficial.
9	Dhaya. [16],	Fusarium oxysporum detection in tomato plant leaves	Image processing operations, naive Bayes	96%	A large plant site will reduce the time needed to classify illnesses effectively.
10	Kong et al. [17],	Sclerotinia stems rot on oilseed rape leaves was discovered	Chemometrics, Soft Independent Component Analysis (SVM), PLS-DA, SIMCA	82.61%, 96.67, and 100.00%.	Precisely selecting optimal wavelengths and observation techniques should be strictly studied to achieve a quick, accurate result.
11	Kurtulmus, Ünal. [3],	Classification of rapeseed variants	SVM, CNN, and grid-search with K-fold cross-validation	99.24%	Future research can enhance the suggested approach by combining multiple classification methods.
12	Shon et al. [1],	Classification of transgenic canola (B. Napus l.) and its hybrids with b. Rapa	CNN, Vis-NIR spectroscopy.	98.9%	This research included only leaves as a feature.
13	Sun et al. [8],	Male oilseed rape plants for UAAS navigation applications on supplementary pollination and aerial spraying.	CNN	93.54%	More varieties of male rapeseed need further demonstrations.
14	Ou et al. [5],	Automatic marking using object detection in Aerial Images.	YOLO-V4, CNN	96.24% and The loss was 0.5843%.	To accumulate these tagged aerial images may assemble for a later experiment on flowering stages identification.
15	Abdalla et al. [15],	Nutrient status diagnosis of infield oilseed rape	Pre-trained CNN, Inceptionv3 and LSTM	95% and 92%	The recommended method is flexible enough to be used for different implementations.
16	Przybył et al. [19],	Rapeseed contamination with fungi detection	CNN, MLP, and RBF	0.89% and 0.85%	Similar models will be implemented, for instance, in expert systems, which would enable, among other things.
17	Yu et al. [18],	Quickly detecting nitrogen concentrations in oilseed rape leaf	SAE, FNN	(R2 P = 0.903, RMSEP = 0.307%, and RPDP = 3.238).	The SAE-FNN method's results will spur more studies into employing DL as a novel chemometric technique.

learning rate, and their influence on overall achievement. CNN is the most broadly used architecture for image identification based on visible characteristics. B. Rapa and B. Alba rapeseed are two data classes.

- This dataset belongs to the exact B. Napus species identification based on two distinct varieties (B. Rapa and B. Alba).
- Supports testing and evaluating several classifiers built on the distinctive characteristics of B. Napus oilseed rape species.
- Information enabling algorithm performance to be estimated under various species features, such as the size and shape of the flower, leaves, and packet.
- A dataset of binary class images that the researchers may use to test various feature extractors and track how well they perform.
- Data applied against a constant white background in a regulated domain.
- Researchers may use the datasets to assess the environmental risks associated with transgenic B. Napus.

Identifying B. Napus can be valuable for several purposes, including classifying various B. Napus varieties. The type of oil seed rape is beneficial for the study of the agricultural sector. Additionally, species recognition from images is a promising research area in categorizing B. Napus using machine learning and computer vision. Identification of the oil seed rape (B. Napus), which aids in tracking the species' activity in agriculture, is crucial for agricultural research. In this study, a dataset is generated of two different B. Napus variations, i.e., B. Rapa and B. Alba, contain 1200 images, which we have collected using cameras.

3.1.2. Data collection

We used digital cameras such as Realme 6i, Vivo Y12, and OPPO A53 with sensors bearing the resolution of 48 MP quad camera, 12.98 MP (4160 × 3120 pixels), and 13 MP primary camera OV13B sensor with 2 MP Macro Lens utilized for all image collection. We collected the data from January to February, and we managed this data from different rural areas in Bangladesh. A rural area named Enayetpur is part of Chowhali Upazilla in the district of Sirajganj. On sunny days from 11:00 a.m. - 4:00 p.m., we captured the B. Rapa images from an agriculture field in Enayetpur. And, on foggy days at 10:00 a.m., the pictures of B. Alba were collected from an agriculture field in a rural area called Bhangabari, a part of Belkuchi Upazilla in the district of Sirajganj. From the various agricultural regions, we were able to collect a total of 1500 photos of these two variations. Then, we keep the clear images and exclude the unclear pictures from the data for preprocessing to create an ultimate dataset. Fig. 1 shows the sample images of the collected dataset, and Fig. 2 shows the corresponding data collection regions of the proposed research.

3.2. Data pre-processing

Before the actual dataset is prepared, individual raw images are ready. We have preprocessed those images one by one using image processing techniques. Fig. 3 shows the corresponding mechanism for the data pre-processing. This figure demonstrates that we have first applied image processing, then the research involves image enhancement and resizing techniques. After that, background removal was performed, and Contrast Limited Adaptive Histogram Equalization

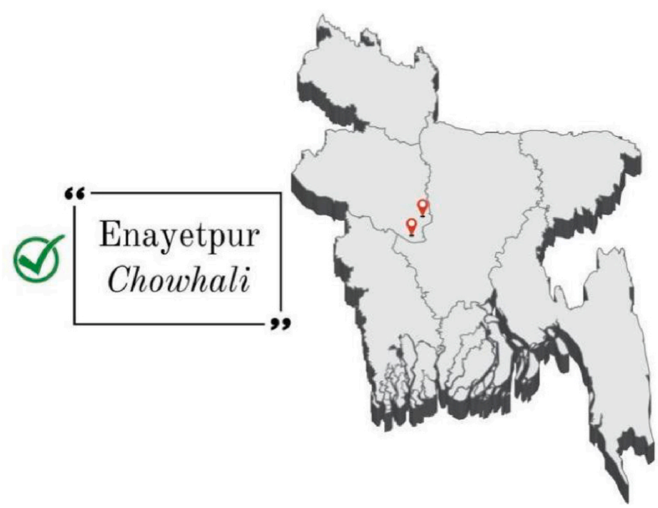

Fig. 2. Data collection areas in Bangladesh.

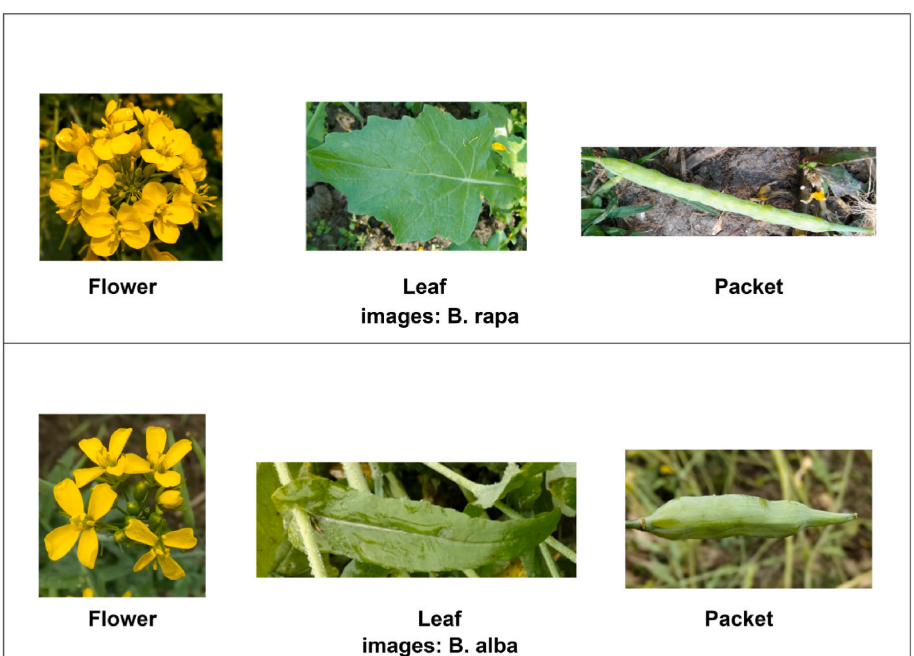

Fig. 1. Sample images of the collected dataset.

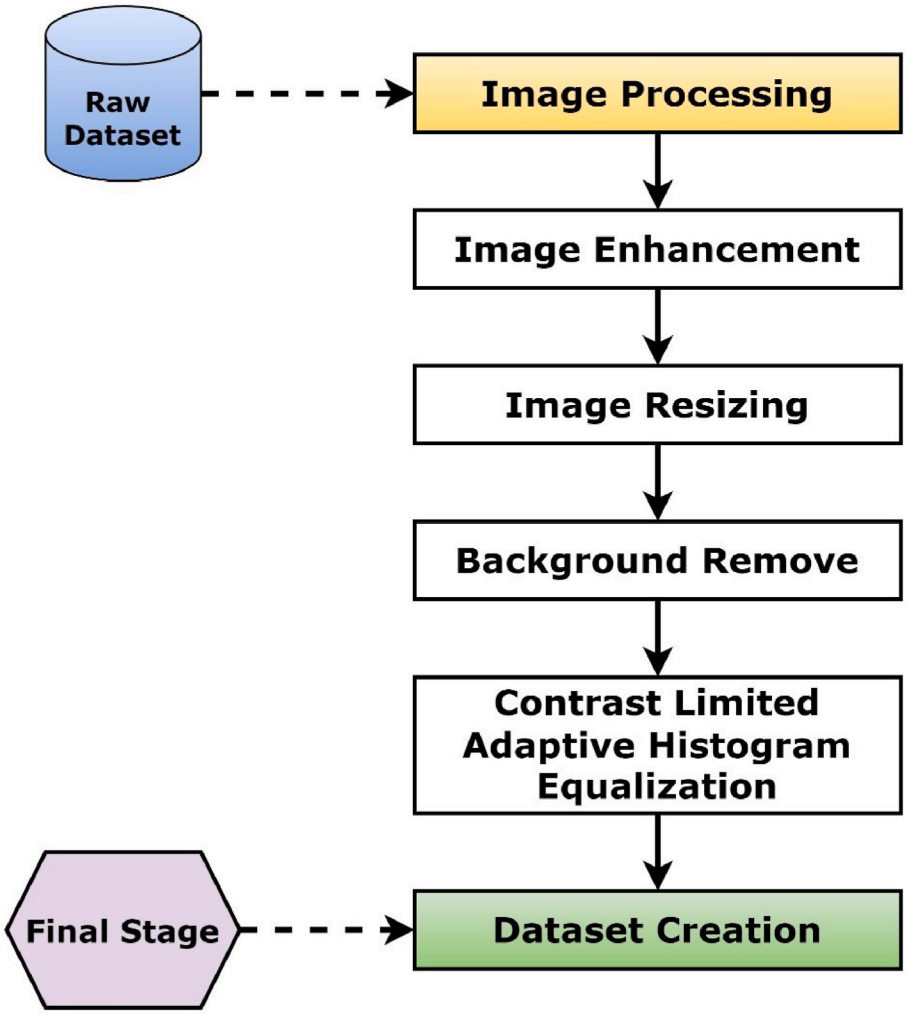

Fig. 3. Pipeline of data pre-processing.

(CLAHE) was finally applied. Fig. 4 represents the corresponding flowchart of the data pre-processing in dataset creation.

3.2.1. *Removing background*

Quick selection tools in Adobe Photoshop CC 2020 were used to extract pixel values for image analysis. The performance of the object ratio extracted using machine learning techniques and the manually removed object ratio by Adobe Photoshop has then been compared by R2 and RMSE. The calculation is shown in the following way:

$$\mathrm{RMSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \widehat{y}_i)^2 \tag{1}$$

$$R^2 = \frac{SS_{residual}}{SS_{total}} \tag{2}$$

3.2.2. *Average weight*

Red, Green, and Blue are the three colors that make up an RGB image, which can be converted to grayscale by adding R to G and B and dividing by 3. Fig. 5 shows the sample images after applying greyscale.

$$Grayscale = (R + G + B\ /\ 3) \tag{3}$$

$$New\ grayscale = ((0.3\ *\ R) + (0.59\ *\ G) + (0.11\ *\ B)) \tag{4}$$

- Enhance its appearance for visual interpretation
- Facilitate the performance of subsequent tasks

$$\mathrm{N_{avg}} = \frac{\mathrm{N_{CR-X_p}} * \mathrm{N_{CR-Y_p}}}{\mathrm{N_{gray}}} \tag{5}$$

Where,

N_{avg} = number of average pixels.
N_{gray} = The number of gray levels in the context.
N_{CR-Xp} = A contextual region's number of pixels in the X direction.
N_{CR-Yp} = A contextual region's number of pixels in the Y direction
Following the accurate clip-limit calculation.
$^{N}CL{=}^{N}CLIP * {}^{N}avg$

3.2.3. *Contrast Limited Adaptive Histogram Equalization (CLAHE)*

Limited Adaptive Histogram Equalization (CLAHE) [20] modifies adaptive histogram equalization. Before calculating the transformation function, this technique applies the enhancement function to all neighboring pixels. This differs from Adaptive Histogram Equalization (AHE) because it has a contrast constraint. The CLAHE method for enhancement redistributes and trims the image's grayscale using the most significant value of the histogram and creates enhanced images by interpolating gray-level mapping. The mapping process is applied using four-pixel clusters. Each mapping tile partially overlaps in the image region before a single pixel is extracted, and the four mapping processes are done to that pixel. Repeat over an image, interpolating between

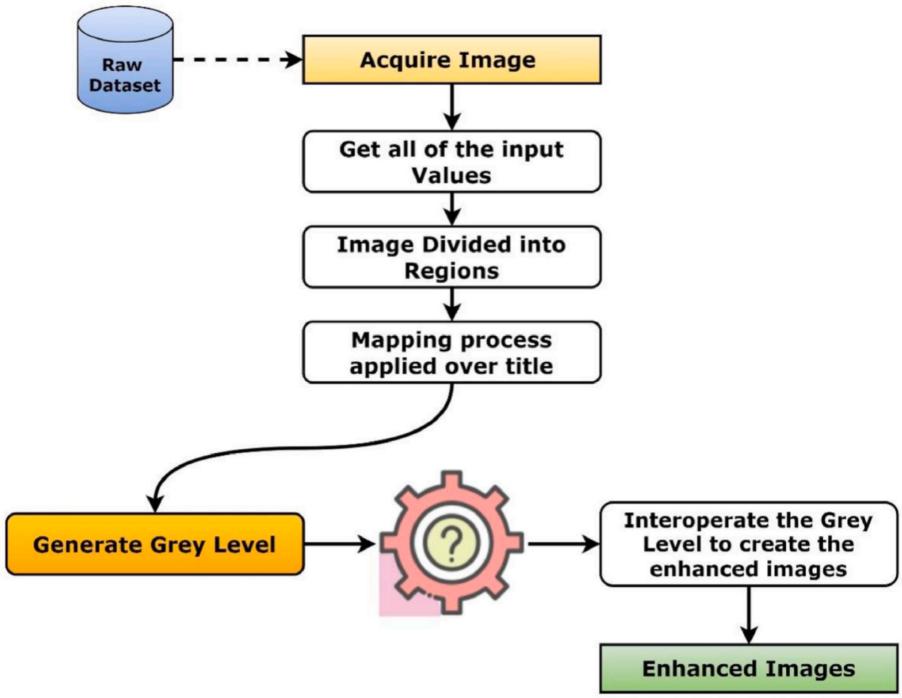

Fig. 4. Flowchart for data pre-processing in dataset creation.

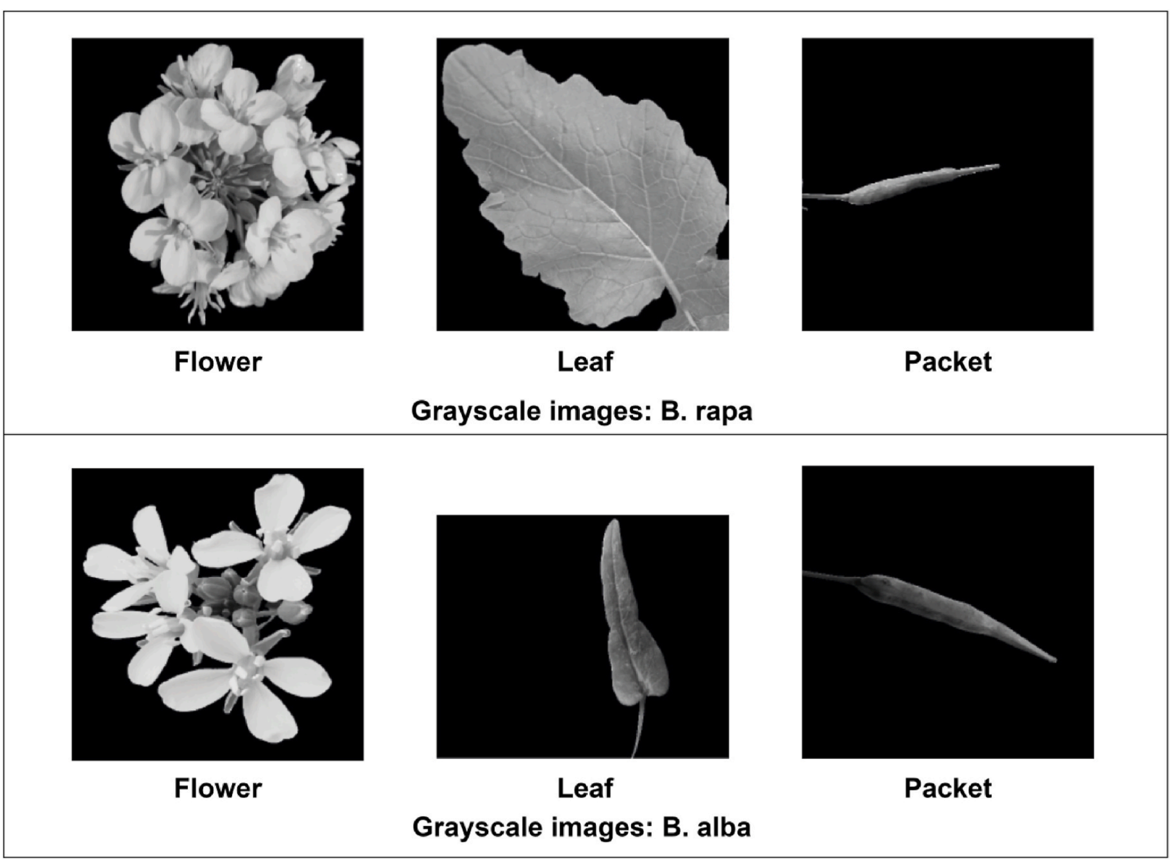

Fig. 5. Images after applying grayscale.

those results to achieve improved pixels.

CLAHE method relies on two essential hyper-parameters such as the number of tiles (8 × 8), which is the default in OpenCV) and the clip limit (2.0). CLAHE enhanced the features of background-removed photos by converting RGB images into grayscale images. CLAHE has two primary intentions while improving the contrast of the images. This makes image analysis, object detection, and image segmentation easy. Fig. 6 shows the pictures with CLAHE.

3.3. Deep learning pipeline of the proposed solution

3.3.1. Dataset preparation

The binary class dataset contains three dominant features of B. Napus variations (flowers, leaves, and packets). We have created two datasets for our study.

❖ No Background Dataset.
❖ Contrast Limited Adaptive Histogram Equalization Applied Dataset

Each dataset contains 1200 images, 600 from B. Rapa and 600 from B. Alba. We have considered images from three different parts (200 images of flowers, 200 images of leaves, and 200 images of packets). Every portion contains three parts: training, testing, and validation. The training, testing, and validation consist of 80%, 10%, and 10% images for the classification. Then, we upload the created dataset to the Kaggle account for coding purposes; we chose the Kaggle platform because of its easy usability.

3.3.2. Main methodology

There were several factors due to which we had to limit our collection to only 1500 authentic images, such as the availability of images, and the cost and time required to collect and annotate each image manually. To address the issue of limited data, we utilized transfer learning, which involves using a pre-trained CNN as a starting point for training on a smaller dataset, and can also help improve the model's performance with limited data.

In our research, different successive stages are used for classifying the most important variations of B. Napus, such as B. Rapa and B. Alba, cultivated in Bangladesh's agricultural field. We used the most effective deep CNN models to identify them. The used models are DenseNet201, ResNet50, InceptionV3, Xception, and VGG19. Fig. 7 shows the overall system illustration of this research.

3.3.3. Adopted models

For classification purposes, five state-of-the-art neural network models were adopted, namely DenseNet [21], Xception [22], VGG19 [23], InceptionV3 [24], and ResNet50. Each of these models was pre-trained on the ImageNet dataset, and the weights were transferred while initiating training. Before training the models, the top layer of every model was discarded, and a new dense layer containing two neurons for two image classes was introduced.

While training, the batch size was set to 16. The learning rate was 0.0001, and the learning rate decay was 0.00001. For convergence to the global minimum loss, a categorical cross-entropy loss function was utilized along with the RMSProp optimizer. The Softmax classifier was applied to the final dense layer for classification purposes.

While training, the corresponding validation loss after each epoch was monitored. The training was continued until no improvement in validation loss was encountered in 10 consecutive epochs. The learning trends for the DenseNet model on the flowers, leaves, and packets datasets are shown in Fig. 8. The entire training and evaluation procedure was performed on the Kaggle GPU platform. The models were applied with Python and Keras with TensorFlow on the backend.

4. Results and discussion

In this experiment, our dataset contains two types of B. Napus species, such as B. Rapa and B. Alba. These variations are commonly observable in their flower, leaf, and packet features. We defined those two variations, B. Rapa as "M" and B. Alba as "S" in our dataset. The

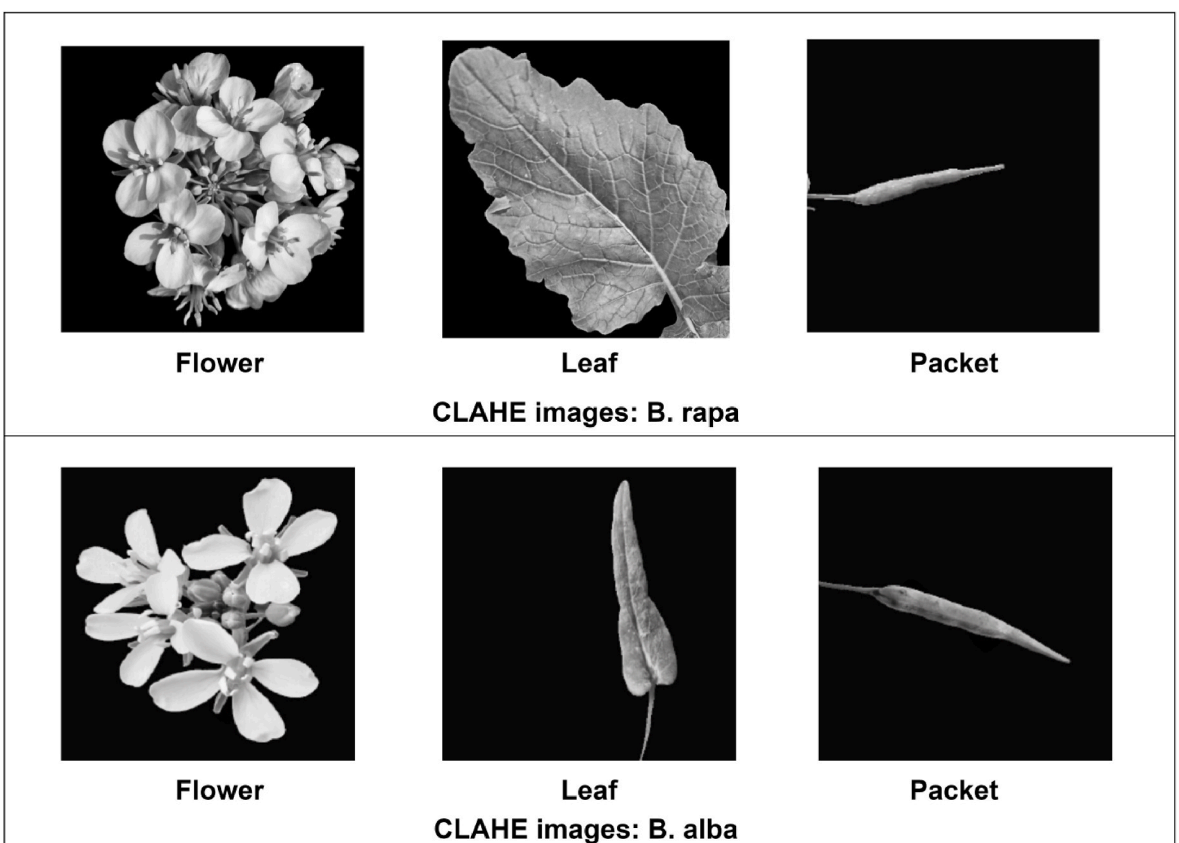

Fig. 6. Images after applying CLAHE.

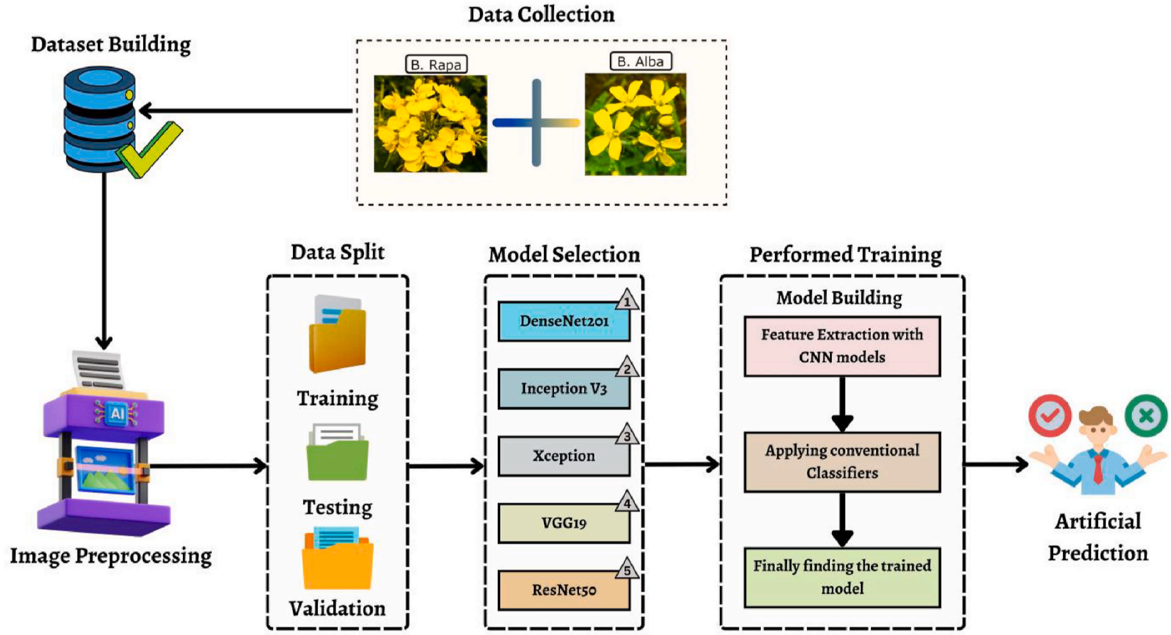

Fig. 7. Architecture of proposed rapeseed species identification system.

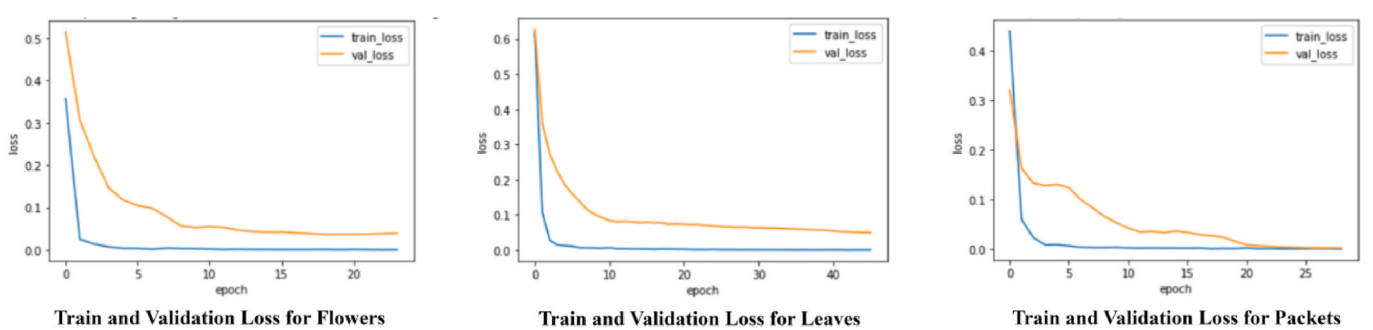

Fig. 8. Train and Validation loss for the DenseNet model on flowers, leaves, and packets.

employed dataset was split into training, testing, and validation groups. The background of the images is removed through Adobe Photoshop, and then an image processing method called CLAHE is performed for better characteristics of the species. We created two datasets for this study: a without-background images dataset and a CLAHE-applied images dataset. Our experiment occurred in two stages. First, we applied the adopted five-deep CNN models (DenseNet201, ResNet50, VGG19, InceptionV3, and Xception) to the without-background dataset. We individually used these models on flowers, leaves, and packets and achieved 100% high accuracy for flowers and packets and 97% accuracy for leaves with the DenseNet201. And then, we applied DenseNet201 to the CLAHE leaves dataset, and finally, we achieved 100% higher leaf accuracy. Tables 2–4 show the classification report of the flowers, packets, and leaves with different CNN architectures and their performance evaluation matrix. In contrast, Fig. 8 shows the training and validation losses.

4.1. With DenseNet

We evaluated the flowers for both variations, such as B. Rapa and B.

Table 2
Classification report for the flowers of B. Rapa and B. Alba.

Classification report for the flowers						
Model	Species	Precision	Recall	F1- score	Support	Accuracy (%)
DenseNet201	*B. Rapa*	1.00	1.00	1.00	20	1.00
	B. Alba	1.00	1.00	1.00	20	
VGG19	*B. Rapa*	1.00	1.00	1.00	20	1.00
	B. Alba	1.00	1.00	1.00	20	
Inceptionv3	*B. Rapa*	1.00	1.00	1.00	20	1.00
	B. Alba	1.00	1.00	1.00	20	
Xception	*B. Rapa*	1.00	1.00	1.00	20	1.00
	B. Alba	1.00	1.00	1.00	20	
ResNet50	*B. Rapa*	0.53	1.00	0.69	20	0.55
	B. Alba	1.00	0.10	0.18	20	

Table 3
Classification report for the leaves of B. Rapa and B. Alba.

Classification report for the Leaves						
Model	Species	Precision	Recall	F1-score	Support	Accuracy (%)
DenseNet201	*B. Rapa*	0.95	1.00	0.98	20	0.97
	B. Alba	1.00	0.95	0.97	20	
VGG19	*B. Rapa*	0.87	1.00	0.93	20	0.93
	B. Alba	1.00	0.85	0.92	20	
Inceptionv3	*B. Rapa*	0.91	1.00	0.95	20	0.95
	B. Alba	1.00	0.90	0.95	20	
Xception	*B. Rapa*	0.95	1.00	0.98	20	0.97
	B. Alba	1.00	0.95	0.97	20	
ResNet50	*B. Rapa*	0.60	0.90	0.72	20	0.65
	B. Alba	0.80	0.40	0.53	20	

Table 4
Classification report for the packets of B. Rapa and B. Alba.

Classification report for the packets						
Model	Species	Precision	Recall	F1-score	Support	Accuracy (%)
DenseNet201	*B. Rapa*	1.00	1.00	1.00	20	1.00
	B. Alba	1.00	1.00	1.00	20	
VGG19	*B. Rapa*	1.00	0.90	0.95	20	0.95
	B. Alba	0.91	1.00	0.95	20	
Inceptionv3	*B. Rapa*	0.95	0.95	0.95	20	0.95
	B. Alba	0.95	0.95	0.95	20	
Xception	*B. Rapa*	0.95	0.95	0.95	20	0.95
	B. Alba	0.95	0.95	0.95	20	
ResNet50	*B. Rapa*	0.52	0.70	0.60	20	0.53
	B. Alba	0.54	0.35	0.42	20	

Alba, and computed their accuracy, precision, recall, f1-score, and support. We achieved outstanding achievements with DenseNet201 based on the characteristics of the flowers for the two varieties (B. Rapa and B. Alba). Our acquired accuracy is 100%, and the packets were categorized with a high accuracy and precision of 100% for both versions. Contrarily, we achieved 0.97% accuracy when analyzing the leaves, with 95% precision for B. Rapa and 100% precision for B. Alba.

4.2. With VGG19

Several evaluation criteria, including recall, precision, F1-score, and accuracy, were assessed using the suggested approach of VGG19. As mentioned above, the model attained a precision of about 100% for both types of flowers, a precision of around 87% for B. Rapa and 100% for B. Alba, and an accuracy of 93% for the leaves. Overall accuracy was 100% for flowers. For the packets with 100% precision for B. Rapa and 91% for B. Alba, the accuracy was 95%.

4.3. With InceptionV3

The calculated accuracy and precision of both flowers were 100%. And the accuracy we achieved for the leaves was 95%, with 91% precision for B. Rapa and 100% for B. Alba. The accuracy and precision of the packets were 95% for both variations.

Following the InceptionV3, with the applied method Xception, the attained performance of the flowers for two variations was the same as the InceptionV3. The achieved accuracy for the leaves was 97%, with 95% precision for B. Rapa and 100% for B. Alba. On the other hand, the performance we achieved with the packets was the same as that of Inception V3.

4.4. With ResNet50

The suggested method, ResNet50, evaluated several evaluation parameters, including recall, precision, F1-score, and accuracy. The model above achieved an accuracy of 65% for the leaves, a precision of about 60% for B. Rapa and 80% for B. Alba, a precision of nearly 100% for B. Alba and 53% for B. Rapa for flowers, and an overall accuracy of 55%. The achieved accuracy was 53% for the packets, with 52% precision for B. Rapa and 54% for B. Alba.

By analyzing the experimental results, we can observe that DenseNet performs better than other networks. DenseNet performs well primarily due to its densely connected architecture, which helps address several challenges faced in deep neural networks. For Instances:

Dense Connectivity: DenseNet connects every layer to every other layer in a feed-forward fashion. This dense connectivity allows for a better flow of information and gradients throughout the network, avoiding the vanishing gradient problem [13].

Parameter Efficiency: DenseNet uses a dense block structure where the output of one layer is concatenated with the input of the next layer, leading to efficient use of parameters. This helps reduce the number of parameters compared to traditional Convolutional Neural Networks (CNNs), making them easier to train [14].

Reuse of Features: DenseNet enables the reuse of features learned by the network, which can be helpful in tasks where the same features are needed at multiple scales. This can improve performance as the network can use and recombine features learned at earlier stages [15].

Reduction in Overfitting: By connecting every layer to every other layer, DenseNet can achieve feature reuse and reduce overfitting, which is a common issue in deep neural networks [16].

5. Conclusion

A deep neural network expert structure was designed to use scanner pictures to distinguish between the two B. Napus rapeseed types. The best predictive strategies for classifying oilseed rape were observed using a variety of character sets, feature models, and deep-learning classifiers. The test and training sets were used to fine-tune the method parameters during training epochs by altering the prediction method's difficulty level. After obtaining the best-optimized model for each degree of complexity, a dedicated validation set is then utilized to approximate the performance for that degree of complexity. The most accurately classified prediction model for classifying B. Napus types was DenseNet201, which offers 100% accuracy. Three characteristics were sufficient to distinguish between two rapeseed types using deep learning classifiers. For the classification of two types of rapeseed, the performances in this study ranged from 90% to 100%, and high performance was also attained using the research described here. However, the main advantage of the suggested approach is that it may be used as a test instrument for rapeseed variety classification without the need for any expensive equipment. There are countless varieties of region-specific oilseed rape species throughout the world. Two rapeseed types in our investigation were identified by looking over them. A more comprehensive expert architecture must be developed to accommodate a much wider range of oilseed rape variants to benefit the oilseed sector and research community. Future research must include developing an automation expert system to identify distinctive characteristics of rapeseed samples.

This experiment is limited to two types of B. Napus species classification, while there are thousands of kinds of rapeseed globally. It's a binary classification with a statistical approach, so the system could not specify multi-class species. In future studies, this proposed approach can be extended to other rapeseed species and include an expert application combined with IoT.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] S.I. Sohn, et al., Discrimination of transgenic canola (Brassica Napus l.) and their hybrids with b. Rapa using vis-nir spectroscopy and machine learning methods, Int. J. Mol. Sci. 23 (1) (2022), https://doi.org/10.3390/ijms23010220.

[2] M.M. Karim, et al., Production of high yield short duration Brassica Napus by interspecific hybridization between B. oleracea and B. Rapa, Breed Sci. 63 (5) (2013) 495–502, https://doi.org/10.1270/jsbbs.63.495.

[3] H. Ünal, Expert Systems with Applications Discriminating Rapeseed Varieties Using Computer Vision and Machine Learning, October, 2014, https://doi.org/10.1016/j.eswa.2014.10.003.

[4] A. Rahman, K. Asif, Machine Learning to Predict Rapeseed Traits from RNA-Seq Data, 2022, pp. 136–145, https://doi.org/10.3233/AISE220031.

[5] X. Shan, X. Liu, Q. Zhang, C. Liu, Z. Ou, An automatic marking method based on object detection algorithm in aerial images an automatic marking method based on object detection, Algorithm. Aerial.Images (2022), https://doi.org/10.1088/1742-6596/2218/1/012041.

[6] R. Chen, A. Shimono, M. Aono, N. Nakajima, R. Ohsawa, Y. Yoshioka, Genetic diversity and population structure of feral rapeseed (Brassica Napus L.) in Japan, PLoS One 15 (1) (2020) 1–14, https://doi.org/10.1371/journal.pone.0227990.

[7] S.I. Sohn, et al., Vis-NIR spectroscopy and machine learning methods for the discrimination of transgenic Brassica Napus L. And their hybrids with B. Juncea, Processes 10 (2) (2022) 1–10, https://doi.org/10.3390/pr10020240.

[8] Z. Sun, et al., Image recognition of male oilseed rape (Brassica Napus) plants based on convolutional neural network for UAAS navigation applications on supplementary pollination and aerial spraying, Agric. For. 12 (1) (2022) 1–15, https://doi.org/10.3390/agriculture12010062.

[9] I. Dhaliwal, A. Mason, S.A. Semerenko, N.A. Bushneva, Classification Prediction Model Based on BP Neural Network Classification Prediction Model Based on BP Neural Network, 2022, https://doi.org/10.1088/1742-6596/2218/1/012040.

[10] Z. Xie, S. Chen, G. Gao, H. Li, X. Wu, L. Meng, Evaluation of Rapeseed Flowering Dynamics for Different Genotypes with UAV Platform and Machine Learning Algorithm, 2022.

[11] T.Y. Polyakova, O. V Gorbacheva, Potato Plant Leaves Disease Detection and Classification Using Machine Learning Methodologies Potato Plant Leaves Disease Detection and Classification Using Machine Learning Methodologies, 2021, https://doi.org/10.1088/1757-899X/1022/1/012121.

[12] F.M. Laggner, E. Wolfrum, Plant Diseases Classification Using Machine Learning Plant Diseases Classification Using Machine Learning, 2021, https://doi.org/10.1088/1742-6596/1962/1/012024.

[13] R. Ahmad, K. Sherwani, M.U. Hassan, The Classification of Medicinal Plant Leaves Based on Multispectral and Texture Feature Using Machine Learning Approach, 2021.

[14] P. Kanade, Machine Learning Techniques in Plant Conditions Classification and Observation, 2021, https://doi.org/10.1109/ICCMC51019.2021.9418386. April.

[15] A. Abdalla, H. Cen, L. Wan, K. Mehmood, Y. He, Nutrient status diagnosis of infield oilseed rape via deep learning-enabled dynamic model, IEEE Trans. Ind. Inf. 17 (6) (2021) 4379–4389, https://doi.org/10.1109/TII.2020.3009736.

[16] R. Dhaya, Flawless Identification of Fusarium Oxysporum in Tomato Plant Leaves by Machine Learning Algorithm 2, 2020, pp. 194–201.

[17] W. Kong, et al., Detection of Sclerotinia stem rot on oilseed rape (Brassica Napus L.) leaves using hyperspectral imaging, Sensors 18 (6) (2018) 1–15, https://doi.org/10.3390/s18061764.

[18] X. Yu, H. Lu, Q. Liu, Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (Brassica Napus L.) leaf, Chemometr. Intell. Lab. Syst. 172 (2018) 188–193, https://doi.org/10.1016/j.chemolab.2017.12.010.

[19] K. Przybył, J. Wawrzyniak, K. Koszela, F. Adamski, M. Gawrysiak-Witulska, Application of deep and machine learning using image analysis to detect fungal contamination of rapeseed, Sensors 20 (24) (2020) 1–11, https://doi.org/10.3390/s20247305.

[20] G. Yadav, S. Maheshwari, A. Agarwal, Contrast limited adaptive histogram equalization based enhancement for real time video system, in: Proc. 2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI, 2014, pp. 2392–2397, https://doi.org/10.1109/ICACCI.2014.6968381, 2014.

[21] G. Huang and K. Q. Weinberger, "Densely Connected Convolutional Networks".

[22] C. Google, Xception: Deep Learning with Depthwise Separable Convolutions, 2014.

[23] F.O.R.L. Arge, C.I. Mage, V d c n l -s i r, 2015, pp. 1–14.

[24] C. Szegedy, V. Vanhoucke, J. Shlens, Z. Wojna, Rethinking the Inception Architecture for Computer Vision, 2014.