ORIGINAL ARTICLE

Identification of Green Gram *(Vigna radiata)* Grains Infested by *Callosobruchus maculatus* Through X-ray Imaging and GAN-Based Image Augmentation

L.G. Divyanth[1] · V. Chelladurai[2] · M. Loganathan[3] · Digvir S. Jayas[4] · Peeyush Soni[1]

Abstract

Purpose Green gram (*Vigna radiata*) is an important food legume of the world. However, post-harvest losses due to pulse beetle, *Callosobruchus maculatus* (F.), are significant due to improper storage management practices and undetected internal infestations. The detection of early stages of infestation could help in implementing suitable control practices for insect disinfestation. This study determined the potential of detecting internal infestations caused by *C. maculatus* using the soft X-ray method and deep learning. Furthermore, this study aims to reduce the time and effort needed to prepare a huge amount of image data for this highly data-driven process by using generative adversarial networks (GANs).

Methods A three-class classification method was implemented to identify the infestation stages, namely, uninfested kernel, larva stage, and pupa stage. The approach was based on features extraction from the deepest pooling layer of a state-of-the-art Convolutional Neural Network architecture—the *Xception*, and using support vector machine as the classifier. Moreover, a GAN model was proposed to synthesize artificial X-ray images.

Results The overall $F1$-score produced by the model was improved from 0.86 to 0.91 when the GAN-synthesized dataset additionally supported the training data. Also, the classification accuracy for detecting the stage of internal infestation improved by 5.5%.

Conclusion The experiment showed that X-ray imaging and deep learning–based automatic features extraction could identify internal infestation in green gram grains. The results determine that augmentation using GANs can enhance the status of learning-based grain quality assessment models with reduced manual effort.

Keywords Grain quality · X-ray imaging · Green gram · Convolutional neural network · Support vector machine

Nomenclature

ANN	Artificial neural network
CNN	Convolutional neural network
GAN	Generative adversarial network
GG	Green gram
SVM	Support vector machine

✉ Digvir S. Jayas
   digvir.jayas@umanitoba.ca

1 Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India

2 Department of Agricultural Engineering, Bannari Amman Institute of Technology, Coimbatore, Tamil Nadu, India

3 National Institute of Food Technology, Entrepreneurship and Management, Thanjavur, Tamil Nadu, India

4 Department of Biosystems Engineering, University of Manitoba, Winnipeg, Canada

Introduction

Food legumes are an important source of protein in the human diet (Purohit et al., 2013). Unfortunately, storage losses of food legumes account for a high percentage of total post-harvest losses (Nanda et al., 2012). The reasons behind these losses have been attributed to poor storage structures and lack of regular monitoring (Kedia et al., 2015). In food legumes, insects belonging to the Bruchidae family, such as *Callosobruchus maculatus* (F.), are a major threat, responsible for up to 90% of infestation in green gram, chickpea, black gram, pigeon pea, and lentil (Kedia et al., 2015; Sivakumar

et al., 2010). The eggs are laid on the grain, the larvae enter inside and consume the kernel, pupates inside, and the adult comes out by creating a hole in the seed coat, resulting in complete loss of the grain (Gbaye et al., 2011). Regular and proper monitoring of stored grains will help in early detection of internal infestation and ascertain the trend in the rate of infestation. Numerous conventional methods, such as visual inspection, insect traps, probe sampling, Berlese funnel, electronic nose, thermal imaging, and acoustic detection, have been reported to be useful in the detection of insects in stored grains (Banga et al., 2018; Kumar et al., 2017). However, most of these methods are subjective, inefficient, time-consuming, and can fail to identify the presence of insects at the early stage during storage and transportation. Early detection, cost-effectiveness, and reliability are the major factors considered when adopting any detection method.

Among the aforementioned methods, it has been found that soft X-rays can provide an automated detection system for monitoring insect infestation. This method is capable of detecting various stages of both internal and external grain feeding insects infesting the food products in a non-destructive way quickly (Neethirajan et al., 2007). There are several areas where the application of soft X-ray for assessing the grain quality is likely to be expanded to meet the needs of grain storage researchers and storage authority regulators. The soft X-ray was adopted in numerous studies to detect several stored-product insects with identification accuracies of 84–98% (Karunakaran et al., 2003; Karunakaran et al., 2004a; Narvankar et al., 2009). Detection of *C. maculatus* infestation by soft X-ray imaging and hyperspectral imaging has been attempted in soybean, with limited success (Chelladurai et al., 2014). However, more than 95% accuracy was achieved while detecting the infestation by *Sitophilus oryzae* (L.) in wheat kernels infested by larval stages and more than 99% for pupa- and adult-infested wheat grains (Karunakaran et al., 2003). The classification accuracy of a soft X-ray method to detect the internal infestation of *Rhyzopertha dominica* (F.) and external feeder, *Tribolium castaneum* (Herbst) larva, in wheat was reported as 98% and 86%, respectively (Karunakaran et al., 2004a; Karunakaran et al., 2004b). The approach was also used to detect *Sitophilus granarius* (L.) eggs and internal infestation in wheat grains (Fornal et al., 2007). The stage of infestation was roughly estimated using conventional image processing and segmentation techniques. Similarly, Nawrocka et al. (2010) developed a two-stage algorithm for the X-ray image analysis to estimate the mass loss caused by granary weevil in wheat. Al-Mezeini et al. (2016) investigated the capability of X-ray imaging in detecting infestations caused by saw-toothed beetle (*Oryzaephilus surinamensis* L.) in stored dates. As an improvement, they extracted 44 features from the X-ray images and used linear discriminant analysis (LDA) to discriminate infested dates from un-infested dates. It was also reported that the soft X-

ray method has the ability to detect fungal infections in stored wheat kernels (Narvankar et al., 2009).

On the other hand, deep learning and computer vision algorithms have demonstrated their potential in many stages of agricultural activity, which pave the way for effective and non-destructive evaluation and handling of fruits, vegetables, and food grains (Chen et al., 2002; Chlingaryan et al., 2018; Cubero et al., 2011; Naik & Patel, 2017; Syal et al., 2013). Detection and identification of insect infestations in stored food grains is one such domain these algorithms have demonstrated their usefulness (Junior & Rieder, 2020; Li et al., 2020; Neethirajan et al., 2007; Shen et al., 2018). An artificial neural network (ANN)–based classification model was proposed by Boniecki et al. (2014) for detecting wheat kernel damages due to granary weevil from digital X-ray images. Their three-layer ANN model identified 98.4% of healthy kernels and 100% of infested kernels correctly. However, the features required for constructing this network had to be manually identified and extracted. Similarly, manual feature extraction was adopted in many other studies to investigate the ability of soft X-ray imaging technique to identify internal infestation in grains. For instance, Chelladurai et al. (2014) extracted 33 features based on the texture and histogram of the soft X-ray images of soybean. The use of deep learning–based image analysis through convolutional neural networks (CNNs) presents a quick way to automatically extract the features during the prediction process while simultaneously being a non-destructive method of monitoring infestations in stored grains. Numerous deep learning models using CNNs have been developed for classification purposes in the agricultural domain since, theoretically, they can cope with several challenges such as inter-class similarities and large intra-class variations in background, colour, illumination, and occlusion. Henceforth, a CNN-based classification approach was used in this study to classify the green gram grain X-ray images according to the infestation stage.

The application of deep learning networks has improved the authenticity of many machine vision tasks. However, it is a highly data-driven process. Thus, one of the major challenges for vision tasks is its demand for huge image data. Acquisition of massive amounts of data, labelling, and annotating them is a tedious and tiresome task, especially for agricultural applications (Dyrmann et al., 2016). To overcome these difficulties, image data augmentation through geometric- and intensity-based transformations have been adopted, where the images can be cropped, rotated, scaled, blurred, noised, or translated along axes. In recent years, another advancement in the field of deep neural networks, the generative adversarial networks (GANs) (Goodfellow et al., 2014; Radford et al., 2015; Wiatowski & Bölcskei, 2008), endorses data augmentation and image enhancement. Through GAN models, artificial-realistic images are generated with the help of existing image data and combined with the original training

set to guide the model. GANs have been previously used in some agricultural applications, such as for creating artificial images of specific plants (Giuffrida et al., 2017), weed identification (Espejo-Garcia et al., 2021), plant disease recognition (Arsenovic et al., 2019), and synthesizing images of plant seedlings (Madsen et al., 2019). Giuffrida et al. (2017) proposed a deep-convolutional GAN to synthesize artificial images of Arabidopsis plants with a specified number of leaf count for augmentation purpose. The feasibility and benefits of GAN-based image augmentation over traditional image augmentation techniques for the task of multiple-disease identification were assessed by Arsenovic et al. (2019). In a study on tomato disease detection, a deep learning model was used in conjunction with GANs (Abbas et al., 2021) for generating synthetic images of tomato plants to increase the amount of image data. The model resulted in an accuracy of 97.1% and concluded that augmentation through GANs increases model's generalizability and prevents it from over-fitting problem. Nevertheless, the benefit of GAN-based image augmentation approach over other augmentation approaches for improving the performance of grain infestation detection needs to be further evaluated prior to its application.

Therefore, the major objectives of this study were to (i) determine the potential of soft X-ray imaging method to identify infestations caused by an internal grain feeder, the pulse beetle, *C. maculatus* during its development in green gram grain, and classification of different developmental stages; (ii) design a two-stage methodology—combining GANs and CNN-based feature extraction technique to improve the classification performance; and (iii) compare conventional image data augmentation methods and GAN-based image data augmentation for deep learning–based identification of internal infestation in stored grains.

Materials and Method

In this study, a CNN-based model coupled with support vector machine (SVM) (Schölkopf et al., 2000) classifier was designed to classify three infestation stages of green gram caused by *C. maculatus*. After experimenting with traditional image augmentation methods, the validity of combining GAN-based image augmentation with CNN-based feature extraction was assessed for the classification purpose. A pre-defined CNN architecture called the *Xception* (Chollet, 2017) was employed for the above tasks. The activations (features) from a deep layer of *Xception* network were derived to build a machine learning model based on SVM. The proposed method was evaluated on a test dataset, different from the dataset used for training the GAN and classification models. The MATLAB (The MathWorks Inc., Natick, MA, USA) application was run on Acer Nitro 5 Intel Core i5 9th

Generation Laptop (32GB/1 TB HDD/Windows 10 Home/GTX 1650 Graphics).

Sample Preparation, Image Acquisition, and Dataset

The sample preparation and image acquisition for this study were based on the procedure followed by Chelladurai et al. (2014). Green gram grains purchased from a local supermarket in Winnipeg, Manitoba, was used for this study. Sound kernels were separated from foreign materials, stones, and other impurities manually using sieves. One kilogram of green gram sample was divided into three groups and filled in glass containers for acquiring images for three different datasets: uninfested (sound), larval stage, and pupal stage. For getting larvae- and pupae stage–infested samples, green gram grains in the two glass containers were spread separately in a single layer on a steel plate inside a gage box, which was specifically designed for insects rearing at the Cereal Research Centre, Agri & Agri-Food Canada, Winnipeg, and newly emerged *C. maculatus* adult insects were introduced inside the gage box for artificial rearing. After allowing 24 h for laying eggs on the kernels, green gram kernels with a single egg were collected and placed in two separate glass jars (for developing larvae and pupal stage). These glass containers were properly labelled and placed inside an environmental chamber (CONVIRON, Controlled Environments Limited, Winnipeg, MB, Canada) set at 30°C and 70% humidity. A soft X-ray imaging system (Model: LX-85708, Manufacturer: Lixi Inc., Downers Grove, IL, USA) was used to acquire images of uninfested and infested green gram kernels. The soft X-ray imaging system was turned on 30 min prior to imaging for warm-up and then samples were placed on a sample holder of the radiation window. Preliminary data collection with the uninfested green gram grains was performed to select the best tube current and voltage for acquiring images of green gram without any background noise. Based on the results of this preliminary study, tube current and voltage potential were set at 120 μA and 14.5 kV, respectively. For acquiring images of larval and pupal stages, green gram kernels were imaged after 15 days and 21 days from the egg stage, during which the eggs would have grown into larvae and pupae, respectively.

A total of 326 sound and insect-damaged green gram grains were imaged to determine the feasibility of the soft X-ray method to identify infested kernels. The initial dataset consisted of 117 images of green gram grains without any infestation, 126 images with larval stage, and 83 images with the pupal stage of infestation inside the grain. The three-class classification corresponds to the infestation stages, namely, "Grain GG" (uninfested) (GG refers to green gram kernel), "Larva GG", and "Pupa GG". The images had 640 × 480 pixels and 96 dpi

horizontal and vertical resolution. Sample X-ray images belonging to the three classes are shown in Fig. 1.

Pre-processing

All analyses in this study, including pre-processing of images, were performed using *MATLAB R2021b* (The Mathworks Inc., Natick, MA, USA) software. Since the *Xception* CNN architecture accepts images with three channels as input, the greyscale X-ray images were converted into three channel images using the *gray2rgb* function (Fig. 2a). The grey level of the background was not uniform throughout the image; hence, it was difficult to separate out the kernel from the background by adopting simple segmentation techniques. Thus, to extract the kernel without losing significant information, the following steps were executed: (i) a base image was created using a green gram X-ray image by manually removing the kernel present in it (Fig. 2b), (ii) all the X-ray images and the base image were converted into single-channel greyscale images, (iii) *imcomplement* function was used to derive the greyscale complements of all images (Fig. 2c and d present the complement of actual green gram X-ray image and the base image, respectively), (iv) the base image complement (i.e. Fig. 2d) was subtracted from the complement of green gram X-ray images using *imsubtract* function (to obtain the background eliminated image as shown in Fig. 2e), and finally, (v) the resulting images were again inverted back (Fig. 2f). Now, the background-eliminated images were enhanced (contrast increment) by saturating 1% top and bottom of all pixel values in the image (Fig. 3). From each class, 30 images were separated out for testing purpose. Hence, the distribution of the training set was as follows: 87 images in Grain GG class, 96 images in Larva GG class, and 53 images in Pupa GG class.

Image Synthesis Through GAN

Typically, a huge image dataset is employed in vision tasks for obtaining better results. Here, image augmentation through a GAN was performed to increase the number of images for training. In this section, the main goal was to develop a GAN model that could learn how to generate synthetic X-ray images of green gram grains infested by *C. maculatus*. The resulting images were combined with the original images to train the *Xception* (Chollet, 2017) model for extracting the features. A GAN consists of a generator for producing new images, and a discriminator that differentiates the fake images synthesized by the generator from the real ones (Goodfellow et al., 2014; Radford et al., 2015). Both the components are trained in an adversarial process where the generator tries to deceive the discriminator through its artificial images, and the discriminator tries to diagnose the generated images. The features of the output image from the generator network are conditioned by the features of real images used for training the GAN model.

The generator was designed to take a random valued vector z and transform it to a sound/infested green gram pseudo-X-ray image $G(z)$ of size 256×256×3 through a set of transposed

Fig. 1 Sample X-ray images from the dataset (**a**) uninfested (sound) kernel, (**b**) larval stage, and (**c**) pupal stage

(a)

(b)

(c)
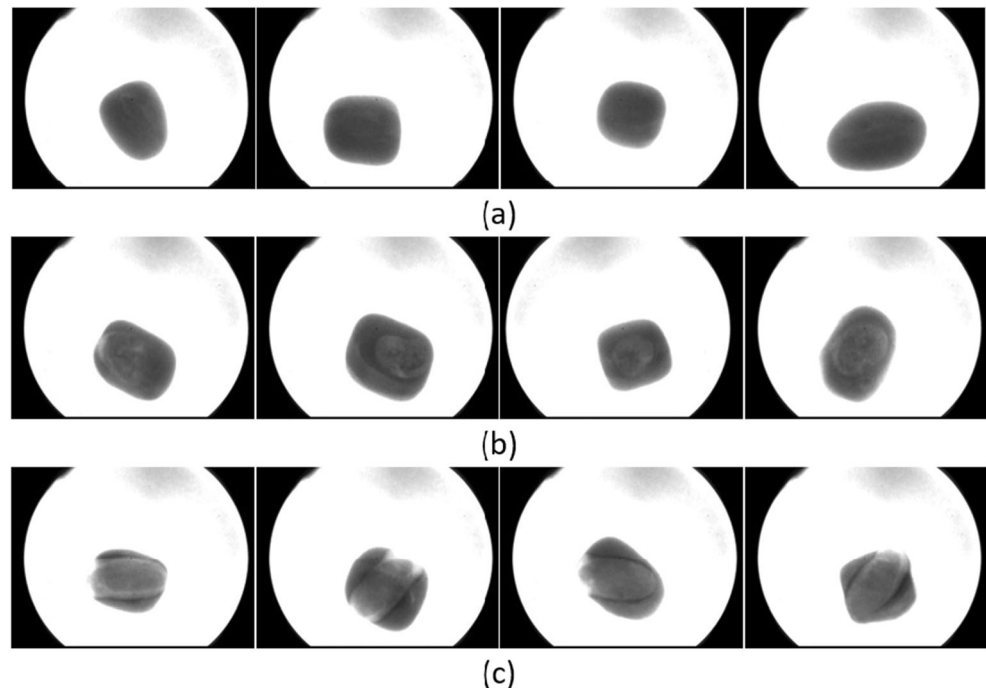
Fig. 2 X-ray image of infested green gram: (**a**) actual, (**b**) base image, (**c**) greyscale complement of the actual image, (**d**) base image complement, (**e**) image obtained after background removal, and (**f**) final image
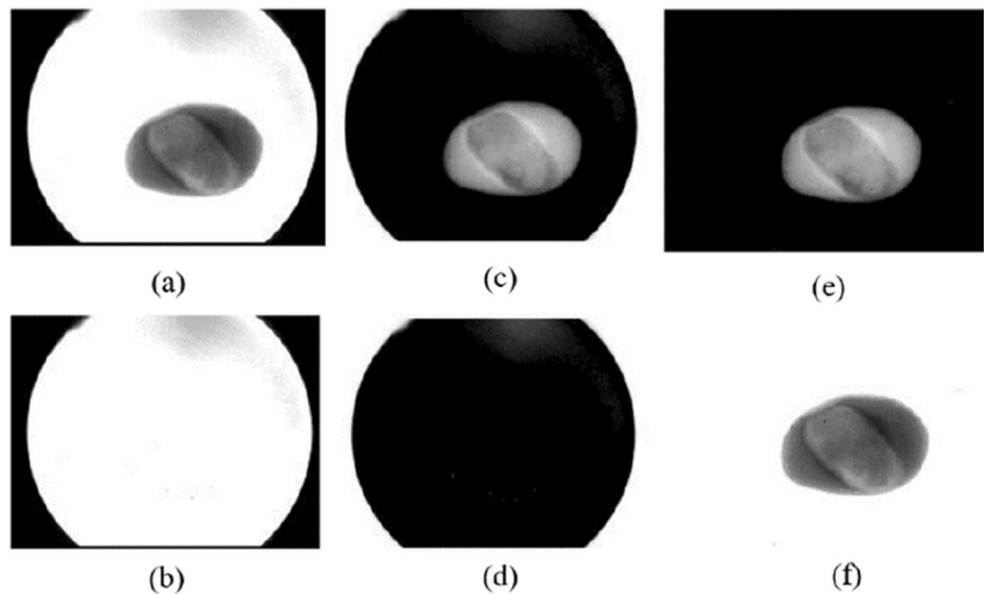
convolutional (t-Conv) layers. A vector of size 1×1×100 was fed into the first layer and upscaled to a matrix of size 28×28×512 using a fully connected operation. Furthermore, this array passed through a set of four t-Conv layers, which used 5×5 filters and 2×2 strides to perform transposed convolutions. For the last t-Conv layer, three 5×5 filters were specified, which corresponded to three channels of the generated RGB image (Fig. 4).

The discriminator network intakes the fake images $G(z)$ synthesized by the generator and the real images x, and returns a prediction score (whether the image is recognized as real or fake) using a series of strided convolutions for down-sampling (Fig. 5a). The convolutional layers used 5×5 filters with 2×2 strides. The number of filters layer-wise is shown in Fig. 5a, which results in a 1×1×1 numerical prediction score at the discriminator's output. The generator's goal was to produce artificial images $G(z)$ very similar to x, whereas the discriminator optimized its weights to classify the real and fake images correctly. Other architectural implementations made in the discriminator were the use of leaky ReLU (identity function when the input is positive; however, it has a small slope for negative

values instead of a flat slope) activation layers (with a slope of 0.2) and dropout layers (dropout rate of 0.3) after each convolutional layer to avoid the potential problem of overfitting. Furthermore, the weights were Gaussian weight initialised, and hyperbolic tangent function was enforced in the final output layers of generator and discriminator networks.

The loss function scores of the generator and discriminator would reach an equilibrium after training some iterations, meaning that the generator has learnt the features of real images to its maximum ability. Once this equilibrium is reached, the generator can be expected to convert the random vectors into conceivable images. Equation (1) shows the objective function of the GAN model.

$$G^* = arg \overset{min}{G} \overset{max}{D} [E_x \log(D(x)) + E_z \log(1-D(G(z)))] \qquad (1)$$

where $(D(x))$ denotes the output from discriminator for real data x, and $(1 - D(G(z)))$ denotes the discriminator output for generated synthetic data $G(z)$. The discriminator tries to maximize this function against the generator that tries to minimize it (Fig. 5b).

Fig. 3 (**a**) Image after background removal; (**b**) enhanced image
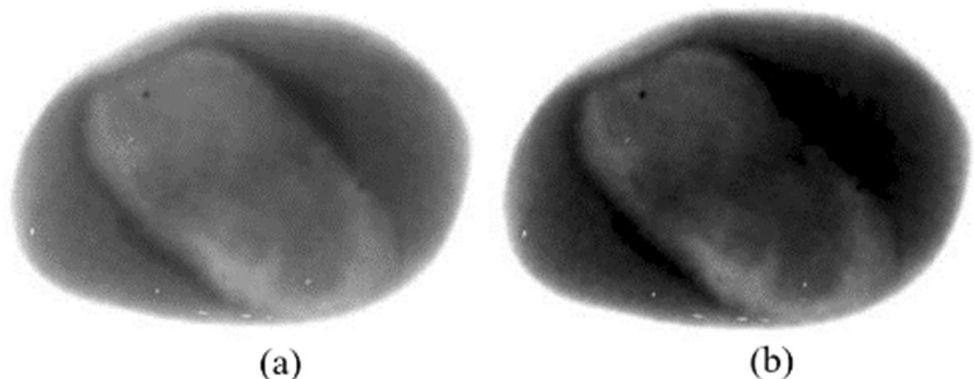
Fig. 4 Network architecture of GAN-generator model
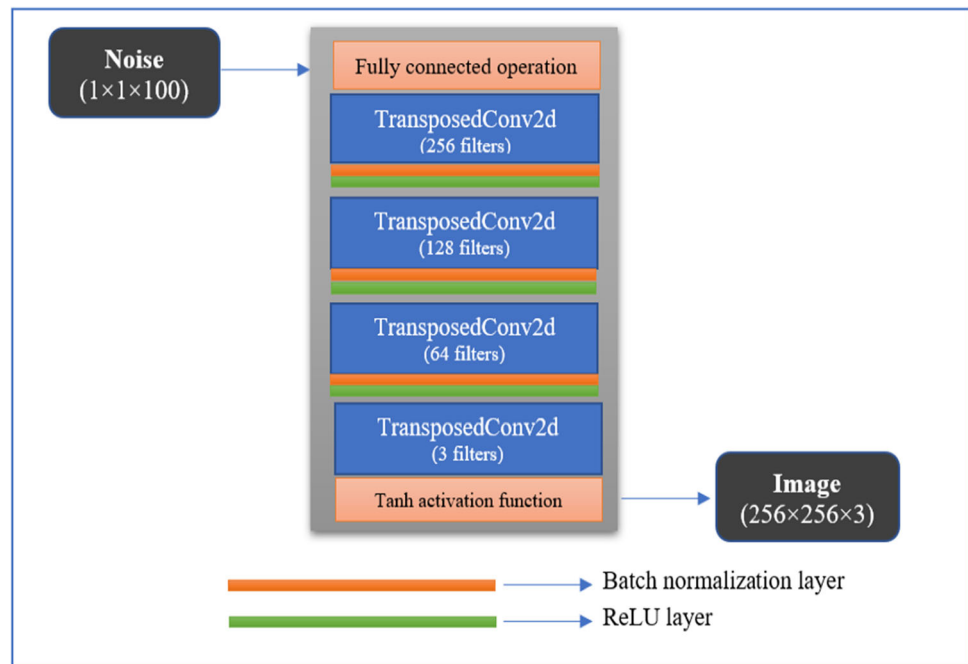
The adversarial training process was carried out for 500 epochs for each class (the network was trained thrice separately (for Grain GG, Larva GG, and Pupa GG) on the real images to synthesize images of the corresponding class) through the Adam optimisation algorithm (Kingma & Ba, 2017), with a mini-batch size of 16, learn rate of 0.0005, and a gradient decay factor of 0.5. The t-distributed stochastic neighbour embedding (t-SNE) visualization (van der Maaten & Hinton, 2008) was adopted to check the fidelity of the generated images. The distribution of data points of generated and original images in a t-SNE plot gives an insight into the reliability of the achieved data. The details and the intuition behind these plots have been reported in van der Maaten and Hinton (2008).

Classification Through Feature Extraction

Deep Learning Architecture

In this study, a popular CNN architecture—*Xception* (Chollet, 2017)—was adopted, which achieved the third-best performance on the ImageNet dataset (Deng et al., 2009). The model is based on depth-wise separable convolutions, where two simultaneous processes are involved: spatial and point-wise convolutions. The spatial convolution is performed independently on each input channel from the previous layer's output. Furthermore, the point-wise convolutions use 1×1 kernels to convolve point by point (pixel by pixel). This drastically reduces the number of parameters and calculations, thus increasing the computational performance. The model consisted of 14 residual blocks, which included 3 convolutional layers and

33 depth-wise convolutions (Fig. 6). The separable convolutions were placed throughout the architecture.

The convolutional layers of the CNN contain a set of filters with weights associated with them, bringing forth the features of input to the successive layers (Szegedy et al., 2015). In feature extraction, the activations from the deep layers of a CNN model are derived to fit a machine learning model with an adopted classifier. Feature extraction (Wiatowski & Bölcskei, 2008) also helps in reducing the computational complexity since relative importance can be given to the most useful variables, alongside being memory efficient compared to transfer learning techniques. In this study, features were extracted from the global pooling layer of *Xception* network after training the model from scratch with the green gram X-ray dataset, giving rise to a vector of 2048 features. A machine learning model based on the SVM classifier was employed as a substitution to the fully connected and classification layers in the *Xception* model to perform the classification task, with these features as the predictor variables.

Support Vector Machine Classification

The final classification output layer in the original *Xception* network is based on a logistic regression classifier. However, studies have preferred other non-linear kernel-based classifiers like SVM over a simple logistic regression and have demonstrated its advantages over others, including better results (Habib et al., 2019; Hlaing & Zaw, 2017; Padol & Sawant, 2016). SVM, a kernel-based algorithm applied for both classification and regression tasks, shows better performance when the data is not regularly distributed, and in cases

Fig. 5 (**a**) Network architecture of GAN-discriminator model; (**b**) image synthesis flowchart (the prediction score in turn is used to update the weights of the network)

(a)

(b)
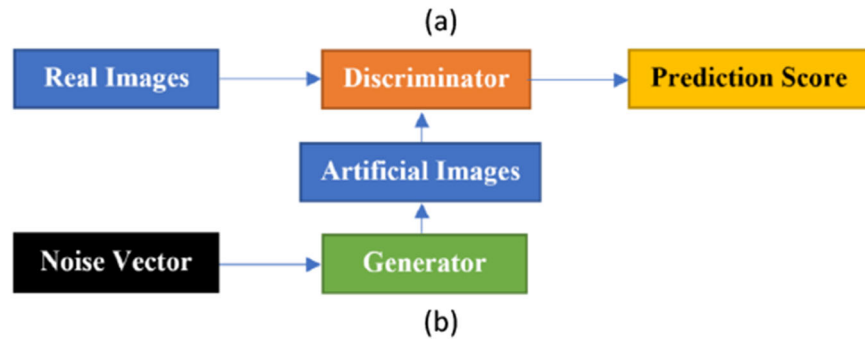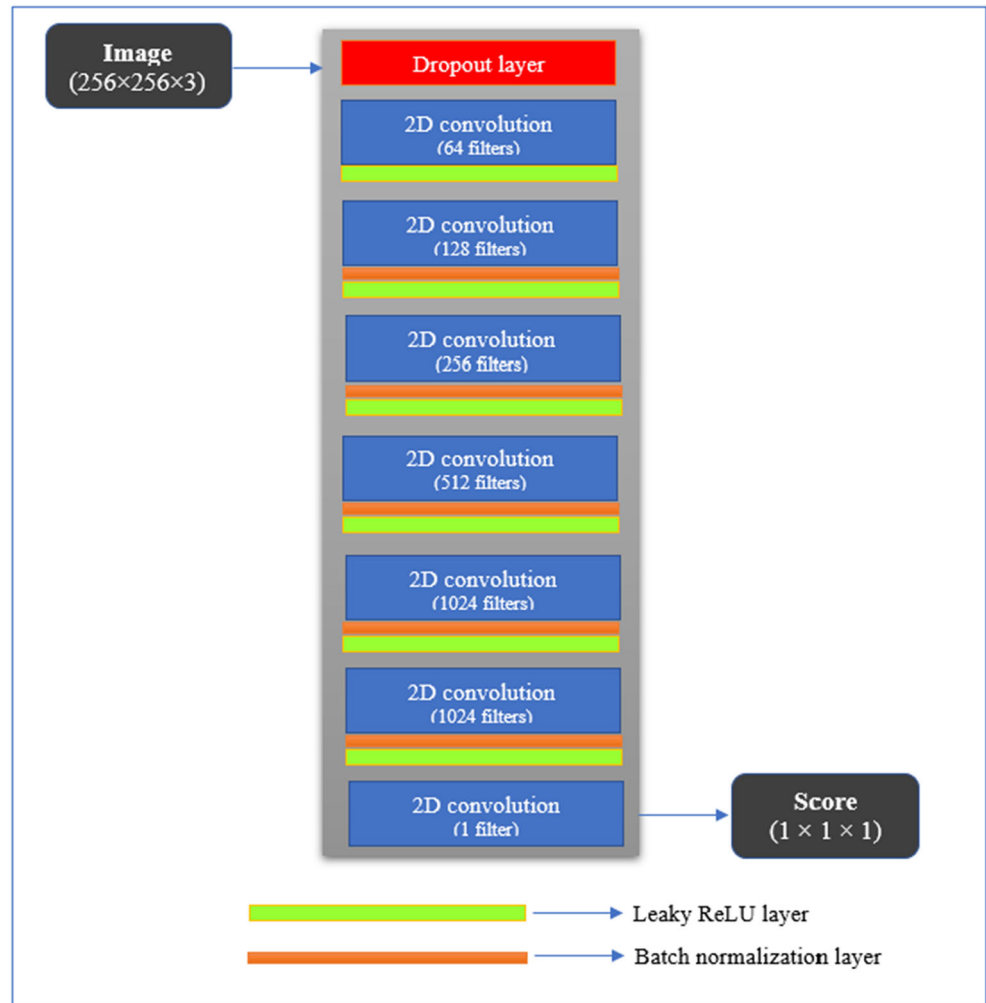
when the number of training samples is less than the number of features (Habib et al., 2019). Furthermore, SVMs are known for their memory efficiency, faster prediction (less execution time), and better computation complexity (Huang et al., 2006; Schölkopf et al., 2000; Vapnik, 1995). Classification is executed by finding a hyperplane that differentiates the classes very well. Suppose x represents the feature space of n dimensions, the separating plane H can be represented as in Eq. (2):

$$H : \left(w_{1 \times n}^{T}\right)\left(x_{n \times 1}\right) + b = 0 \tag{2}$$

where b is the intercept/bias term of the plane.

Given a set of data points in the space, the ultimate goal of the algorithm is to maximize the minimum distance between the separating hyper-plane and a point (Eq. (3)) (support vector of the point at a minimum distance from the plane).

$$w^{*} = arg^{max}{}_{w}[min_{n}d_{H}(\phi(x_{n}))] \tag{3}$$
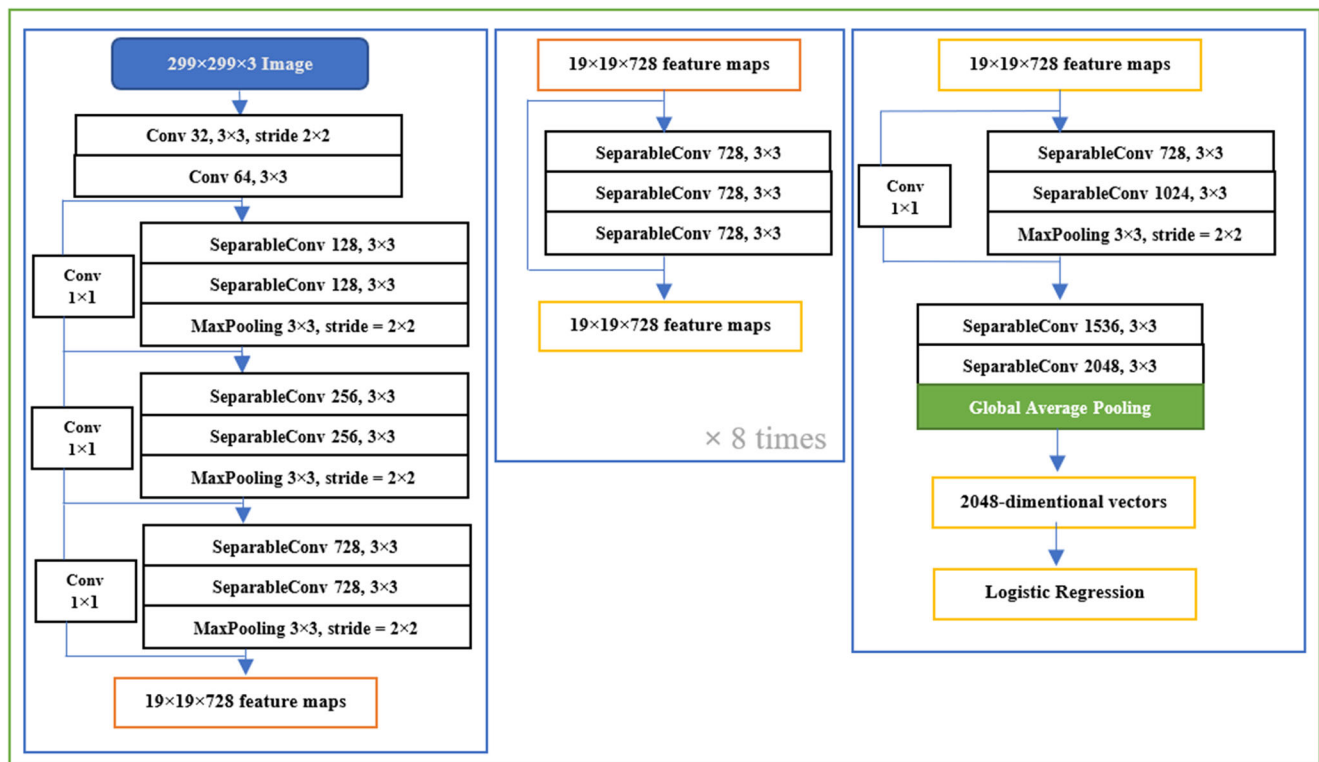
Fig. 6 Illustration of the convolutional and depth-wise separable convolutional layers present in the *Xception* architecture

where $d_H(\phi(x_n))$ denotes the distance between the plane and data point. The features were transformed using the Radial (Gaussian) basis function (RBF) in this work. Transforming the data using a kernel function maps data to a higher dimension, expecting that the resulting points are easier to separate into classes. Kernels make the SVM more flexible to handle non-linear problems. The RBF kernel computes the element (i, j) of Gram matrix G given by Eq. (4),

$$G(x_i, x_j) = exp\left(-|x_i - x_j|^2\right) \tag{4}$$

where x_i and x_j are the *i*th and *j*th observations of the training set, respectively.

Training Options and Evaluation Metrics

All the images were resized to 299×299×3 pixels to fit the input size of *Xception*. The images were resized using the bicubic interpolation method, i.e. the output pixel values were the weighted mean of the 4×4 neighbourhood pixels. A two-fold evaluation was performed based on the training image dataset—one with the real/original green gram X-ray images, and then with combined real and artificial images. Conventional augmentation approaches based on image geometry and intensity, such as random rotations (+90 to −90),

translation along x- and y-axes, x- and y-axes mirroring, and brightness variations with scales of 0.8, 0.9, 1.1, and 1.2 were specified during the training process for both the cases to check the competence of synthetic dataset. Regarding training options for the CNN, the chosen optimizer was stochastic gradient descent with momentum, with the following hyperparameters—initial learn rate set to 0.001, momentum of 0.9, and a weight decay factor of 0.0001. A mini-batch size of 16 was set, and the maximum number of epochs was limited to 200. The two-stage methodology proposed for combining GAN and feature extraction technique is depicted in Fig. 7. The performance of the models was statistically analyzed through metrics such as classification accuracy, Cohen's Kappa coefficient (Pérez et al., 2020), precision, recall, and *F*1 scores (Eqs. (5–8); where, TP, TN, FP, FN represent true positives, true negatives, false positives, and false negatives, respectively).

$$Precision = \frac{True\ Positives}{Number\ of\ predicted\ samples} \tag{5}$$

$$Recall = \frac{True\ Positives}{Actual\ number\ of\ samples} \tag{6}$$

$$F1-score = 2\ \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$
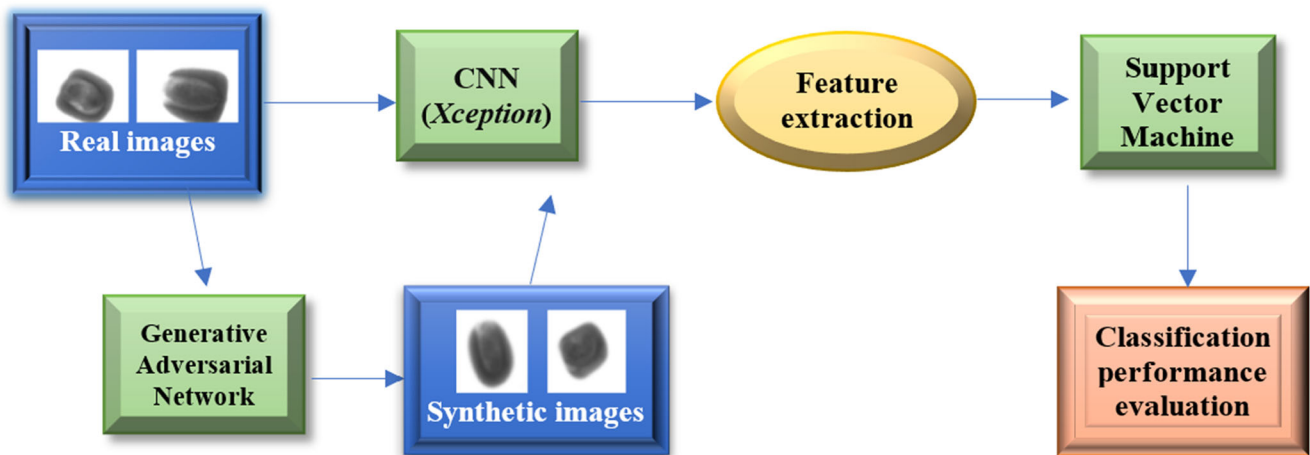
Fig. 7 General workflow of artificial image generation through GAN and classification of infestation stages in green gram grains through feature extraction from *Xception* network

Fig. 8 (**a**) Synthetic X-ray images of green gram kernels produced by GAN (left: uninfested; middle: larval stage; right: pupal stage); (**b**) plot illustrating the generator and discriminator scores during training (4500 epochs)


(a)

(b)
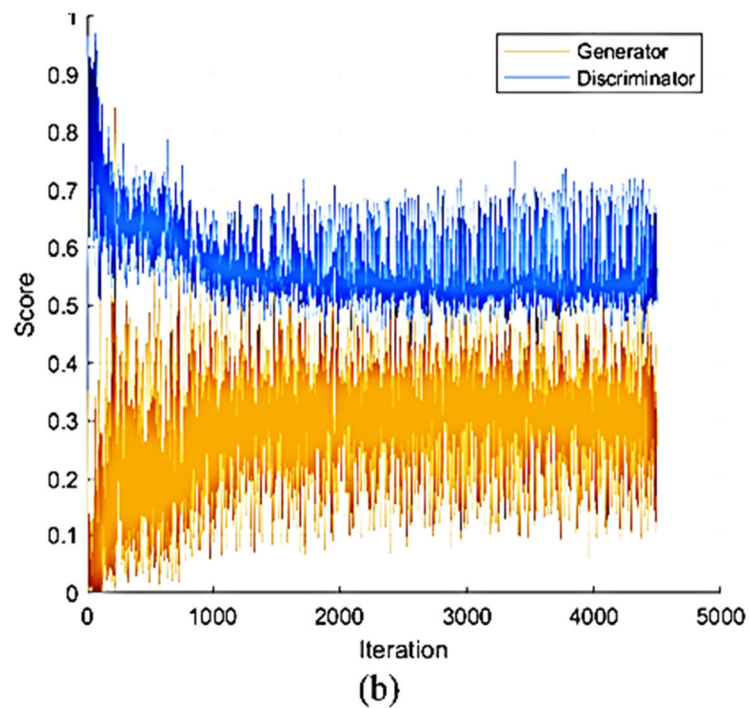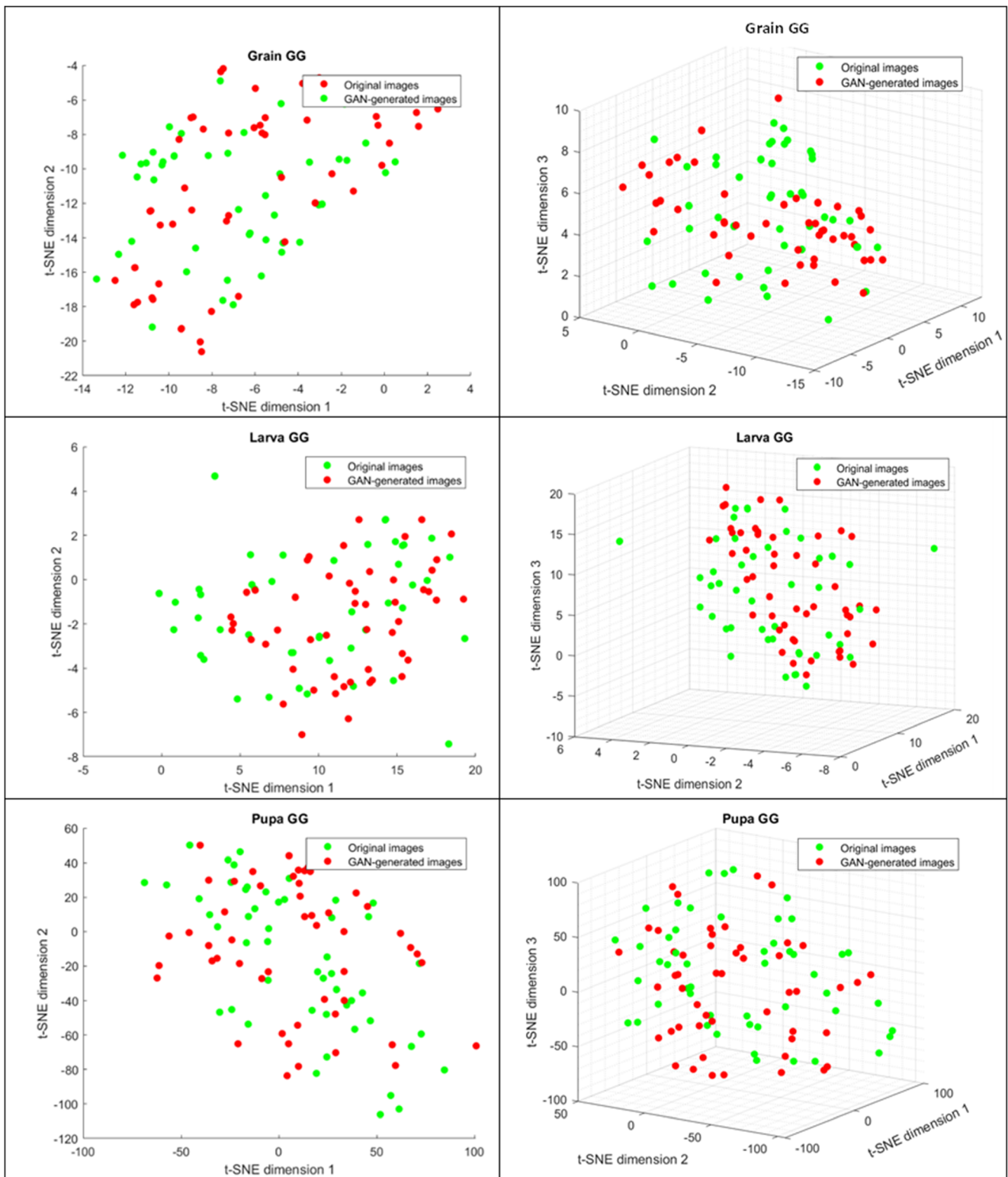
Fig. 9 t-SNE visualization plots of real and synthetic images belonging to the three classes. The features of the images were reduced to two dimensions (left column) and three dimensions (right column) using the t-SNE algorithm. The Grain GG class represent sound kernels, Larva GG and Pupa GG represents the kernels infested at the larval and pupal stages, respectively. The green and red points correspond to original and generated images, respectively

Fig. 10 Classification results ($F1$-score) with different number of artificial images in the training dataset
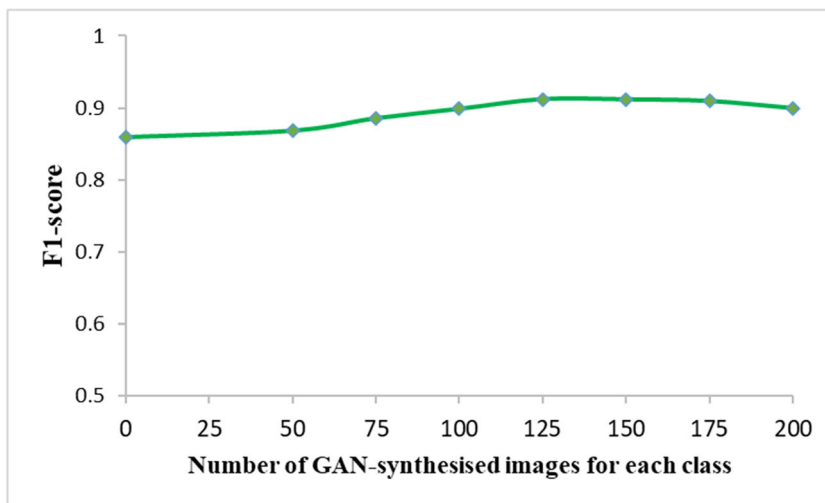
Cohen's Kappa coefficient

$$= \frac{2 \times (TP \times TN - FN \times FN)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$$

$$(8)$$

Results and Discussion

Firstly, the artificial X-ray images generated by the GAN model were examined through visual inspection and t-SNE plots. Second, the model's performance was evaluated for grain infestation stage classification. Further, the validity of artificially synthesized X-ray images of green gram grains was assessed by comparing the classification performances with and without data augmentation through GAN.

Visual Inspection and Assessment of Generated Images

The fidelity of the generated images was assessed before evaluating the classification results. The GAN model was run

Table 1 Summary of image distribution for each class.

Dataset	Grain GG	Larva GG	Pupa GG
Real/original images	117	126	83
Artificial images	125	125	125
Total	242	251	208
Training images	212	221	178
Testing images	30	30	30

three times to generate artificial images for each class with the real images of three classes separately. From the training plot (Fig. 8), it can be observed that equilibrium was attained between the discriminator and generator scores after nearly 1800 iterations. Henceforth, the generator should have learned an appreciable number of visual features from real images and reflected in the artificial images. The t-SNE method of visualization was adopted for plotting the features of 50 real and 50 generated images belonging to the three classes separately to get a fair insight into the visual impression created by artificial images (Fig. 9). Features of these images were obtained from the final layer of the trained discriminator network. The t-SNE algorithm plots the high-dimensional feature vectors of the images into a low-dimensional (two and three dimensions in this work) plot. A similar distribution of the points corresponding to real and generated images demonstrated that desired features were learned and reproduced by the GAN model. The generated images widen the feature coverage of classes and preserve the key features of the real images. However, unlike in the case of sound (uninfested) grain class, the distribution was not homogenous in the plots for larval and pupal stages. In these two classes, the points corresponding to GAN-generated images were mostly concentrated in the centre of the distribution of real images. This indicates that the greater the number of real images available, the better plausible artificial images could be synthesized.

Classification of Grain Infestation Stage

The larvae consumption of green gram grain was noticeable as a bright region in the germ and endoplasm area of the X-ray images. A significant portion of the endoplasm and complete

Table 2   Analysis of classification results of SVM classifier on deep features of *Xception*.

Class	Precision		Recall		F1-score	
	GAN$_0$-SVM	GAN$_{125}$-SVM	GAN$_0$-SVM	GAN$_{125}$-SVM	GAN$_0$-SVM	**GAN$_{125}$-SVM**
Grain GG	0.813	0.931	0.867	0.900	0.839	**0.915**
Larva GG	0.806	0.867	0.833	0.867	0.819	**0.871**
Pupa GG	0.963	0.935	0.867	0.967	0.912	**0.951**
Overall	0.861	0.911	0.856	0.911	0.857	**0.912**

germ part was eaten up when the insect reached the pupal stage. The grey value of a sound kernel was higher than that of an infested kernel class, due to less grain density in the infested portion. As a means of comparing the classification performance with and without data augmentation through GAN, two *Xception* models were trained with the same hyperparameters, one with the original green gram X-ray image dataset with conventional data augmentations, another with the GAN-generated images in addition to the former. The test dataset comprised only of original images—the GAN-generated images were only used to support the training set. The number of synthetic images used to support the training process was increased from 50 to 200, in increments of 25 images. For each case, the features from the global pooling layer of the trained network were used to fit an SVM model. The feature vector obtained from this layer had a dimension of 2048, hence, offering a wide range of features for learning. Dimensionality reduction using principal component analysis (PCA) was not incorporated because, to make the classification robust, PCA features must be modified frequently based

on the changes in data distribution. Also, during inference (in real-time), each data point must be passed through an extra computation step. The classification F1-scores on the test dataset are presented as Fig. 10, showing that the addition of GAN-generated images in the training set improved the model's performance. After training with only the initial original dataset with conventional image augmentations, it was found that the model exhibited an overall F1-score of around 0.857. After adding GAN-synthesized images, the average F1-score improved significantly and reached a maximum value of 0.912, corresponding to the addition of 125 artificial images to each class. However, as more synthetic images were added, there was no further improvement in the performance. The F1-score started to decrease when the number of artificial images in the training set was more than 125.

The class-wise image distribution used by the best model (GAN$_{125}$-SVM, subscript number denoting the number of GAN-generated images) for training and testing purposes (the test data contains only independent real images) is summarized in Table 1. The detailed classification result of this
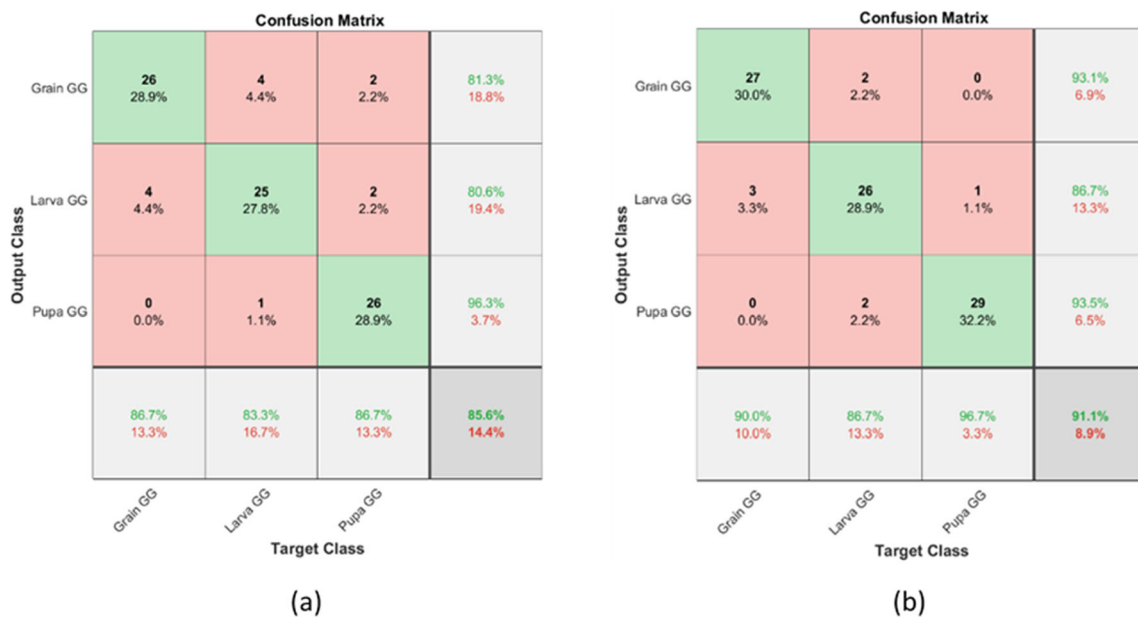
Fig. 11 Classification confusion matrices produced by the model when trained (**a**) without GAN synthesized dataset, (**b**) using both real and GAN-synthesized datasets

model during testing is compared with the original model (GAN$_0$-SVM) in Table 2, in terms of mean precision, mean recall, and average F1-scores for five independent runs. Once the model was trained with images augmented through the adversarial network, the accuracy improved to 91.1% from the previously attained 85.6% on the test data (Fig. 11). Moreover, the Kappa coefficient (which is a measure of how the classification results compare to the values/classes assigned by chance) was noted to have been increased from 0.78 to 0.87. Figure 11 presents the confusion matrices of the model trained using only the original images (GAN$_0$-SVM) and with both original and artificial images (GAN$_{125}$-SVM). The precision and recall rates for the classes except Larva GG were ≥ 0.9. The most misprediction was observed for sound kernels that were wrongly detected as infested by the larval stage. With the addition of artificial images, the recall rate of classifying kernels infested in the pupal stage improved from 0.867 to 0.967. Overall, it is evident from the results that synthetic images have contributed to the increase in the performance of the SVM-based classification model for the identification of infested green gram grains. Moreover, Table S1 in Online Resource 1 provides the classification results when the classification layer of the *Xception* network was used to identify the infestation stage instead of following the feature extraction approach. Comparing this result with that of Table 2, it can be observed that using SVM classifier on the intermediate features accord slightly better results.

The activation maps derived from the intermediate layers of the *Xception* model are illustrated in Fig. 12. It can be noticed that the model learns finer details present in the images as it passes through deeper layers. The activations in the initial layers just present the outlines of the shapes and edges, whereas, as we go to deeper layers, the feature maps seem to be more abstract and gradually fade, meaning that the infested portions of the grain get more attention.

Identification of *C. Maculatus* in stored grains through acoustic sensors was studied by Banga et al. (2019) and Bittner et al. (2018). Though acoustic sensing provides a non-destructive approach for detecting the presence of insects and identifying its developmental stage, it is time-consuming, needs appropriate environment to effectively capture the sound, and demands decisive signal processing techniques. Also, most of the commonly used conventional methods of insect detection such as probe sampling and trap method (Mohan & Rajesh, 2016) and Berlese funnel (Minkevich et al., 2002) take much time (4–8 h) to produce the result, and there is no possibility of individual grain-level monitoring. This study demonstrated the potential of soft X-ray method to detect infestation caused by *C. maculatus* in green gram grains. The results have shown that internal infestations in green gram can be identified non-destructively with an appreciable accuracy by applying a conventional classifier, which was SVM in this study, on the intermediate features of a CNN. The *Xception* CNN architecture obtained a performance accuracy of 91.1% on the independent test data. The method proves to be less tedious than developing a grain infestation model through manual features extraction (Boniecki et al., 2014; Chelladurai et al., 2014; Karunakaran et al., 2004a).

Furthermore, it can be concluded that data augmentation through GANs, apart from conventional methods, proves to be efficient in training classification models with reduced labour effort. Artificial images with good quality were synthesized, and visual inspection and t-SNE plot showed that enough features were learned and reproduced by the model to support the training of the classification model. The improvement in the performance of the SVM-based

Fig. 12 Input image and activation maps derived from the first, third, fifth, and seventh depth-wise convolution blocks (from left to right) of the *Xception* model
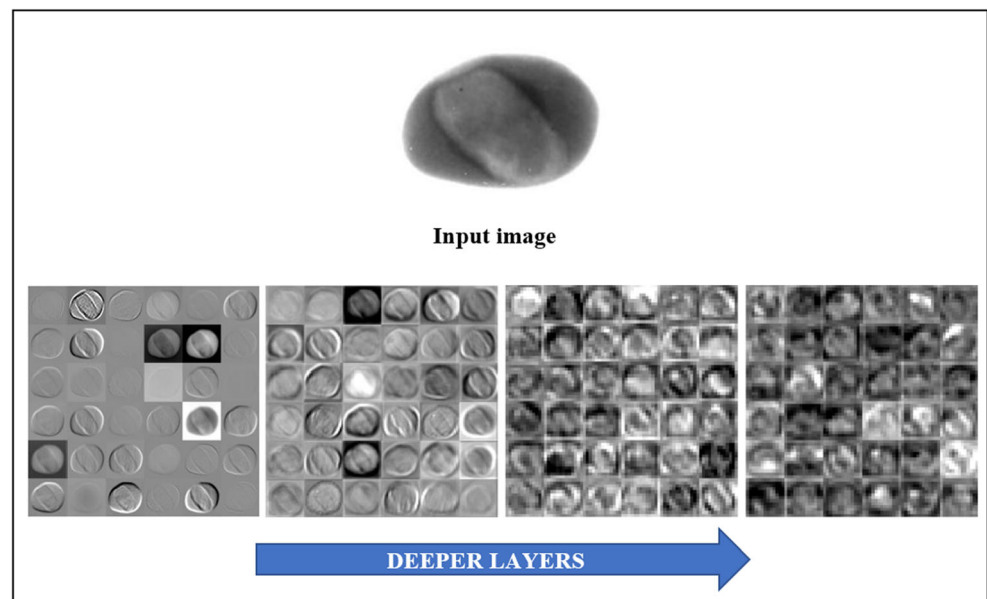
classification model by using the artificial X-ray images of green gram demonstrated that the study's objectives were achieved. The greatest advantage of such augmentation is that the model trained with this data can generalize much better, especially on shape- and texture-based features. This study has supported well that GANs could be a useful tool in assessing grain quality through machine vision technology. In the future, work could be carried out to validate the best architecture for the generator and discriminator of GAN, which could generate more favourable images that could plausibly improve the performance.

In this study, we have hypothesized initially that X-ray imaging can be used to detect the internal infestation in green gram, since many previous studies have made use of this technology for monitoring infestations in other food grains and achieved good results. The key improvement is the use of deep learning–based image analysis, which does not involve complex and time-consuming manual feature extraction tasks and conventional image processing. This study revealed that the achieved performance is promising, and could be further improved, considering the potential of various deep learning architectures. The arguments that motivated the adoption of *Xception* architecture for features extraction were its network size, memory efficiency, and convolution parameters such as the kernel size and strides (Chollet, 2017). For instance, although CNN architectures such as AlexNet and GoogleNet perform better than Xception on the ImageNet dataset, they must not be preferred for feature extraction because they use large filter sizes and strides that are not ideal for detecting fine-grain features. In future, more analysis can be performed to reduce the complexity of the CNN architecture further and assess the performances of other conventional classifiers. Furthermore, the level of infestation estimated by this method can give an estimation of the potential contamination in green gram kernels when fumigated. Also, during the experiment, green gram kernels were placed manually on the sample platform to take X-ray images; therefore, further testing is required for insect infestation detection using X-ray method with a single layer of grain samples on a line-scan X-ray system.

Conclusion

In this study, a deep learning–based vision system for identifying internal infestation of *C. maculatus* in green gram grains through X-ray imaging was evaluated. The features were extracted from the pooling layer of a *Xception* model, and an SVM classifier was trained for the classification task to identify the correct stage of infestation. Moreover, a GAN-based image data augmentation technique was presented to support the identification of infested grains of green gram under a limited dataset. The amount of training data for each class

was increased by using synthetic images synthesized by the GAN model. The generated images were manually inspected, and the t-SNE method was adopted to check their fidelity. The plot showed that the feature distributions of real and fake images resemble each other. Furthermore, the generated images and the original dataset were combined to train a classification model under the same methodology, i.e. through CNN-based feature extraction method, and with SVM as classifier. With the help of generated images, the accuracy and F1-scores of the model improved from 85.6% and 0.86 to 91.1% and 0.91, respectively. The study demonstrated that classification models could yield an improved performance with the help of GAN-based augmentation as compared to traditional augmentation methods.

Declarations

Competing interests The authors declare no competing interests.

References

Abbas, A., Jain, S., Gour, M., & Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with C-GAN synthetic images. *Computers and Electronics in Agriculture, 187*, 106279. https://doi.org/10.1016/j.compag.2021.106279

Al-Mezeini, N., Manickavasagan, A., Al-Yahyai, R., Al-Wahaibi, A. K., Al-Raeesi, A. A., & Khriji, L. (2016). X-ray imaging of stored dates to detect infestation by saw-toothed beetles. *International Journal of Fruit Science, 16*(1), 42–56. https://doi.org/10.1080/15538362.2015.1044692

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., & Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry, 11*(7), 939. https://doi.org/10.3390/sym11070939

Banga, K. S., Kotwaliwale, N., Mohapatra, D., & Giri, S. K. (2018). Techniques for insect detection in stored food grains: An overview. *Food Control, 94*, 167–176. https://doi.org/10.1016/j.foodcont.2018.07.008

Banga, K. S., Kotwaliwale, N., Mohapatra, D., Giri, S. K., & Babu, V. B. (2019). Bioacoustic detection of Callosobruchus chinensis and Callosobruchus maculatus in bulk stored chickpea (Cicer arietinum) and green gram (Vigna radiata). *Food Control, 104*, 278–287. https://doi.org/10.1016/j.foodcont.2019.02.026

Bittner, J. A., Balfe, S., Pittendrigh, B. R., & Popovics, J. S. (2018). Monitoring of the cowpea bruchid, *Callosobruchus maculatus* (Coleoptera: Bruchidae), feeding activity in cowpea seeds: Advances in sensing technologies reveals new insights. *Journal of Economic Entomology, 111*(3), 1469–1475. https://doi.org/10.1093/jee/toy086

Boniecki, P., Piekarska-Boniecka, H., Świerczyński, K., Koszela, K., Zaborowicz, M., & Przybył, J. (2014). Detection of the granary weevil based on X-ray images of damaged wheat kernels. *Journal of Stored Products Research, 56*, 38–42. https://doi.org/10.1016/j.jspr.2013.11.001

Chelladurai, V., Karuppiah, K., Jayas, D. S., Fields, P. G., & White, N. D. G. (2014). Detection of *Callosobruchus maculatus* (F.) infestation in soybean using soft X-ray and NIR hyperspectral imaging techniques. *Journal of Stored Products Research, 57*, 43–48. https://doi.org/10.1016/j.jspr.2013.12.005

Chen, Y. R., Chao, K., & Kim, M. S. (2002). Machine vision technology for agricultural applications. *Computers and Electronics in Agriculture, 36*(2), 173–191. https://doi.org/10.1016/S0168-1699(02)00100-X

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Computers and Electronics in Agriculture, 151*, 61–69. https://doi.org/10.1016/j.compag.2018.05.012

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, 7 (3), 1251-1258.

Cubero, S., Aleixos, N., Molto, E., Gmez-Sanchis, J., & Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food and Bioprocess Technology, 4*, 487–504. https://doi.org/10.1007/s11947-010-0411-8

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248-255. https://doi.org/10.1109/CVPR.2009.5206848.

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering, 151*, 72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024

Espejo-Garcia, B., Mylonas, N., Athanasakos, L., Vali, E., & Fountas, S. (2021). Combining generative adversarial networks and agricultural transfer learning for weeds identification. *Biosystems Engineering, 204*, 79–89. https://doi.org/10.1016/j.biosystemseng.2021.01.014

Fornal, J., Jeliński, T., Sadowska, J., Grundas, S., Nawrot, J., Niewiada, A., Warchalewski, J. R., & Błaszczak, W. (2007). Detection of granary weevil *Sitophilus granarius* (L.) eggs and internal stages in wheat grain using soft X-ray and image analysis. *Journal of Stored Products Research, 43*(2), 142–148. https://doi.org/10.1016/j.jspr.2006.02.003

Gbaye, O. A., Millard, J. C., & Holloway, G. J. (2011). Legume type and temperature effects on the toxicity of insecticide to the genus *Callosobruchus* (Coleoptera: Bruchidae). *Journal of Stored Products Research, 47*(1), 8–12. https://doi.org/10.1016/j.jspr.2010.08.001

Giuffrida, M. V., Scharr, H., & Tsaftaris, S. A. (2017). ARIGAN: Synthetic arabidopsis plants using generative adversarial network. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2064-2071.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde Farley, D., Ozair, S., Courville, A. C., & Bengio, Y. (2014). Generative adversarial nets. In: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K.Q. Weinberger (Eds.), NeurIPS Proceedings on Advances in Neural Information Processing Systems 27 (NIPS 2014).

Habib, M. T., Majumder, A., Nandi, R. N., Ahmed, F., & Uddin, M. S. (2019). A comparative study of classifiers in the context of papaya disease recognition. In M. Uddin & J. Bansal (Eds.), *Proceedings of International Joint Conference on Computational Intelligence* (pp. 417–429). https://doi.org/10.1007/978-981-13-7564-4_36

Hlaing, C.S., & Zaw, S.M.M. (2017). Model-based statistical features for mobile phone image of tomato plant disease classification. In: 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies, 223–229. https://doi.org/10.1109/PDCAT.2017.00044

Huang, T. M., Kecman, V., & Kopriva, I. (2006). *Kernel based algorithms for mining huge data sets: Supervised, semi-supervised, and unsupervised learning*. Springer.

Junior, T. D. C., & Rieder, R. (2020). Automatic identification of insects from digital images: A survey. *Computers and Electronics in Agriculture, 178*, 105784. https://doi.org/10.1016/j.compag.2020.105784

Karunakaran, C., Jayas, D. S., & White, N. D. G. (2003). Soft X-ray inspection of wheat kernels infested by *Sitophilus oryzae*. *Transactions of ASAE, 46*(3), 739–745. https://doi.org/10.13031/2013.13576

Karunakaran, C., Jayas, D. S., & White, N. D. G. (2004a). Detection of internal wheat seed infestation by *Rhyzopertha Dominica* using X-ray imaging. *Journal of Stored Products Research, 40*(5), 507–516. https://doi.org/10.1016/j.jspr.2003.06.003

Karunakaran, C., Jayas, D. S., & White, N. D. G. (2004b). Identification of wheat kernels damaged by the red flour beetle using X-ray images. *Biosystems Engineering, 87*, 267–274. https://doi.org/10.1016/j.biosystemseng.2003.12.002

Kedia, A., Prakash, B., Mishra, P. K., Singh, P., & Dubey, N. K. (2015). Botanicals as eco-friendly biorational alternatives of synthetic pesticides against *Callosobruchus* spp. (Coleoptera: Bruchidae)-a review. *Journal of Food Science and Technology, 52*(3), 1239–1257. https://doi.org/10.1007/s13197-013-1167-8

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *CoRR*, https://arxiv.org/abs/1412.6980. Accessed 14 Mar 2022.

Kumar, S., Mohapatra, D., Kotwaliwale, N., & Singh, K. K. (2017). Vacuum hermetic fumigation: A review. *Journal of Stored Products Research, 71*, 47–56. https://doi.org/10.1016/j.jspr.2017.01.002

Li, J., Zhou, H., Wang, Z., & Jia, Q. (2020). Multiscale detection of stored-grain insects for intelligent monitoring. *Computers and Electronics in Agriculture, 168*, 105114. https://doi.org/10.1016/j.compag.2019.105114

Madsen, S. L., Dyrmann, M., Jørgensen, R. N., & Karstoft, H. (2019). Generating artificial images of plant seedlings using generative adversarial networks. *Biosystems Engineering, 187*, 147–159. https://doi.org/10.1016/j.biosystemseng.2019.09.005

Minkevich, J. M., Demianyk, C. J., White, N. D. G., Jayas, D. S., & Timlick, B. (2002). A rapid method to detect Cryptolestes ferrugineus (Coleoptera: Cucujidae) larvae in stored grain. *Canadian Journal of Plant Science, 82*(3), 591–597. https://doi.org/10.4141/P01-106

Mohan, S., & Rajesh, A. (2016). Tools for stored product insects management and technology transfer. *Indian Journal of Entomology, 78*, 59–63. https://doi.org/10.5958/0974-8172.2016.00026.2

Naik, S., & Patel, B. (2017). Machine vision based fruit classification and grading – a review. *International Journal of Computer Applications, 170*(9), 22–34.

Nanda, S. K., Vishwakarma, R. K., Bathla, H. V. L., & Chandra, P. (2012). Harvest and post-harvest losses of major crops and livestock produce in India. All India Coordinated Research Project on Post Harvest Technology, ICAR, Ludhiana.

Narvankar, D. S., Singh, C. B., Jayas, D. S., & White, N. D. G. (2009). Assessment of soft X-ray imaging for detection of fungal infection in wheat. *Biosystems Engineering, 103*, 49–56. https://doi.org/10.1016/j.biosystemseng.2009.01.016

Nawrocka, A., Grundas, S., & Grodek, J. (2010). Losses caused by granary weevil larva in wheat grain using digital analysis of X-ray images. *International Agrophysics, 24*(1), 63–68.

Neethirajan, S., Karunakaran, C., Jayas, D. S., & White, N. D. G. (2007). Detection techniques for stored product pests in grain. *Food Control, 18*(2), 157–162.

Padol, P. B., & Sawant, S. D. (2016). Fusion classification technique used to detect downy and powdery mildew grape leaf diseases. In: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 298–301. https://doi.org/10.1109/ICGTSPICC.2016.7955315.

Pérez, J., Díaz, J., Garcia-Martin, J., & Tabuenca, B. (2020). Systematic literature reviews in software engineering—enhancement of the study selection process using Cohen's Kappa statistic. *Journal of Systems and Software, 168*, 110657. https://doi.org/10.1016/j.jss.2020.110657

Purohit, P., Jayas, D. S., Yadav, B. K., Chelladurai, V., Fields, P. G., & White, N. D. G. (2013). Microwaves to control *Callosobruchus maculatus* in stored mung bean (*Vigna radiata*). *Journal of Stored Products Research, 53*, 19–22. https://doi.org/10.1016/j.jspr.2013.01.002

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR. arXiv:1511.06434

Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation, 12*(5), 1207–1245. https://doi.org/10.1162/089976600300015565

Shen, Y., Zhou, H., Li, J., Jian, F., & Jayas, D. S. (2018). Detection of stored-grain insects using deep learning. *Computers and Electronics in Agriculture, 145*, 319–325. https://doi.org/10.1016/j.compag.2017.11.039

Sivakumar, C., Chandrasekaran, S., Vijayaraghavan, C., & Selvaraj, S. (2010). Fumigant toxicity of essential oils against pulse beetle, *Callosobruchus maculatus* (F.) (Coleoptera: Bruchidae). *Journal of Biopesticides, 3*(1), 317–319.

Syal, A., Garg, D., & Sharma, S. (2013). A survey of computer vision methods for counting fruits and yield prediction. *International Journal of Computer Science Engineering, 2*, 346–350.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

van der Maaten, L., & Hinton, G. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research, 9*, 2579–2605.

Vapnik, V. (1995). *The nature of statistical learning theory*. Springer.

Wiatowski, T., & Bölcskei, H. (2008). A mathematical theory of deep convolutional neural networks for feature extraction. *IEEE Transactions on Information Theory, 64*(3), 1845–1866. https://doi.org/10.1109/TIT.2017.2776228