JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

MACHINE LEARNING ENSEMBLE WITH IMAGE PROCESSING FOR PEST IDENTIFICATION AND CLASSIFICATION IN FIELD CROPS

[1]**Dr B Ravi Prasad**, Professor, CSE Department, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, rprasad.boddu@gmail.com

[2.]**K Abdul Basith**, Associate professor, CSE Department, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, Email: khateebabdulbasith2020@gmail.com

Abstract

Reduction of agricultural, production is a serious issue in the agricultural sector, largely because of insect attacks on field's plants. Identifying & categorization of insects have historically been labour-intensive processes that have necessitated the services of trained entomologists. Earlier warning of an insect assault aids farmers in mitigating crop injury, which in turn increases crop yield and decreases pesticide use. To use a variety of features extracted such as texture, colour, form, histogram of oriented gradients-HOG, & global image descriptor, this study classifies crop insects through the application of machine vision & knowledge-based methodologies with image processing (GIST). Insects were organised according to a system that took into account all of these characteristics. In this study, 3 separate insect datasets was subjected to a variety of machine learning-ML methods, such as basic classifiers as well as ensemble classifications, with the results of these classifications being ranked according to a majority vote. Several different types of base classifiers was utilised, including naive bayes-NB, support vector machine-SVM, K-nearest-neighbour-KNN, & multi-layer perceptron-MLP. In order to improve the classification & identifying of insects, they used a combination of ensemble classifications, including random forest-RF, bagging, & XGBoost, as well as we ran a 10-fold cross-validation test. Empirical outcomes demonstrated that using majority voting with ensembles classifications to include texture, colour, shape, HOG, & GIST characteristics enhanced classifications performance.

Keywords: Crops, Ensemble classification, Image processing, insect classification, Machine learning algorithm, Majority voting.

1 Introduction

There have been a rise of the use of computer vision & image processing in agricultural domains like plant disease recognition, fruits recognitions, as well as insect identifying in crops fields. Insects that eat your crops & make you sick raise the cost of food by eating up a larger portion of your harvest.

The annual crop failure rate. When tending to a big crop area, early detection & identifying of insect pests is a significant difficulty for farmers. Manually observations make it challenging to collect reliable data including such insect type, insect features, & insect population density for various plants. As a result, manually methods are problematic due to their slowness, lack of precision, & high rate of human mistake. The present research gets beyond those restrictions by employing computer vision techniques like image processing, learning, & knowledge-based algorithms to identify insect attacks. Using an effective machine vision system, image processing has been demonstrated in the agricultural sector for the detection & identifying of insects in crops like wheat, soybeans, as well as paddy.

Multiple studies have achieved success with automatic insect identification & classification in crops via employing feature extraction and classification methods in image processing. Extraction of relevant features, such as texture, colour, and other attributes, is a crucial stage in the process of machine learning.

To identify bug species based on its appearance. Hassan et al. Suggested an automatic insect identifier that uses form & colour cues to distinguish between grasshoppers & butterflies. Pedestrian detection, facial recognition, & insect detection all benefitted from the application of HOG characteristics. Liu et al. reported combining the maximally stable extremal regions-MSER method with HOG to detect aphid insects in wheat fields with varying aphid colours & densities. Extraction of HOG features from both positive (+ve) & negative (-ve) training examples of aphids improves the accuracy of aphid identification. Scenes identification & target detection in satellite pictures have benefited greatly from GIST's ability to assess the spatial frequency as well as orientation of images while being robust to change in perspective, translation, as well as magnification.

In several areas of study, such as multi-view gender classifications, hyper-spectral image classification, as well as automatic road-sign identification, machine learning-ML methods were applied for constructing base & ensembles classifications. By merging numerous base classification techniques, ensembles learner classification algorithm helps to enhance machine learning-ML outcomes. Although Wang et al.found that Support Vector Machine performed better, the stability of their method is enhanced by using an Artificial Neural Network to classify insects based on their various attributes & ordering levels. Santana et al. created an automatic system for identifying bee species from wing photos; their MLP classifications outperformed linear discriminant analysis by 2.7 percentage points (LDA).

AdaBoost classifier is used for data variability & segmentation algorithms due to its superior performances in the pecan flaw categorization scheme despite employing a smaller number of characteristics. When identifying the presence of codling moth infestation in Gold Rush apples, the majority voting classification had the highest accuracy at 80%.

To better utilise available resources & increase classifications precision for visually identical field crop insects, machine learning-ML is utilised to manage information from several insect species. As part of our study, they used image processing to extract information from photos of crop-damaging insects, & then we used machine learning-ML methods to create classifications systems. It was investigated whether or not it would be possible to classify crop field insects using a mix of characteristics such as texture, colour, shape, HOG, & GIST properties. The image processing Toolkit in MATLAB 2017a was used to construct all the features extract methods. Using SKLEARN, we trained the insect classifications systems with base & ensemble classifiers, as we evaluate our hypotheses using a 10-fold cross-validation technique.

For enhance classification performance, majority voting is included for both base & ensembles classifications. Pest detection and identification in agricultural plants.

2 Methodology

Insects in various agricultural plants was categorised & identified using computer vision techniques.

Insect picture capture & pre-processing, feature extraction, classification base & ensemble classifiers, majority vote, & evaluation of classifications outcomes are the 5 phases shown in Figure 1. Machine learning-ML techniques, including both base & ensembles classifications, was used independently for each unique combinations of characteristics, such included texture, colour, shape, HOG, & GIST, as well as all obtained in a simultaneous sequence. It was decided to use a majority polling technique to increase the reliability of the classifications. Ensembles classifications were used to improve the classification accuracy, allowing farmers to better manage damages to crops & boost yield.

2.1 Image acquisition and pre-processing

The 3 insect datasets, including the Xie insect dataset, the Wang dataset, as well as the butterflyimagedataset(http://museumvictoria.com.au/bioinformatics/butter/images/bthumbliv. htm), were used to classify pests in plants. Every insect-related datasets is described in detail below. The identities of the classes & insects used in each datasets were listed in Tables S1, S2, & S3 of the supplementary materials.

2.1.1 Dataset 1

Researchers used the Xie insect dataset for categorize insects seen in photographs of wheat, corn, soybeans, & canola. A next step, prior to extracting features, is to pre - process insect photos by resizing them to 227 by 227 pixels. As you can see in Figure. 2, they picked insects from all 24 different insect families.

2.1.2 Dataset 2

The pre-processed Wang datasets features 9 orders of insects with such a single pure backdrop colour. Visualised in the form of an insect (Figure. 3). The 227 9 227 .The photos of insects are utilised to extract characteristics.

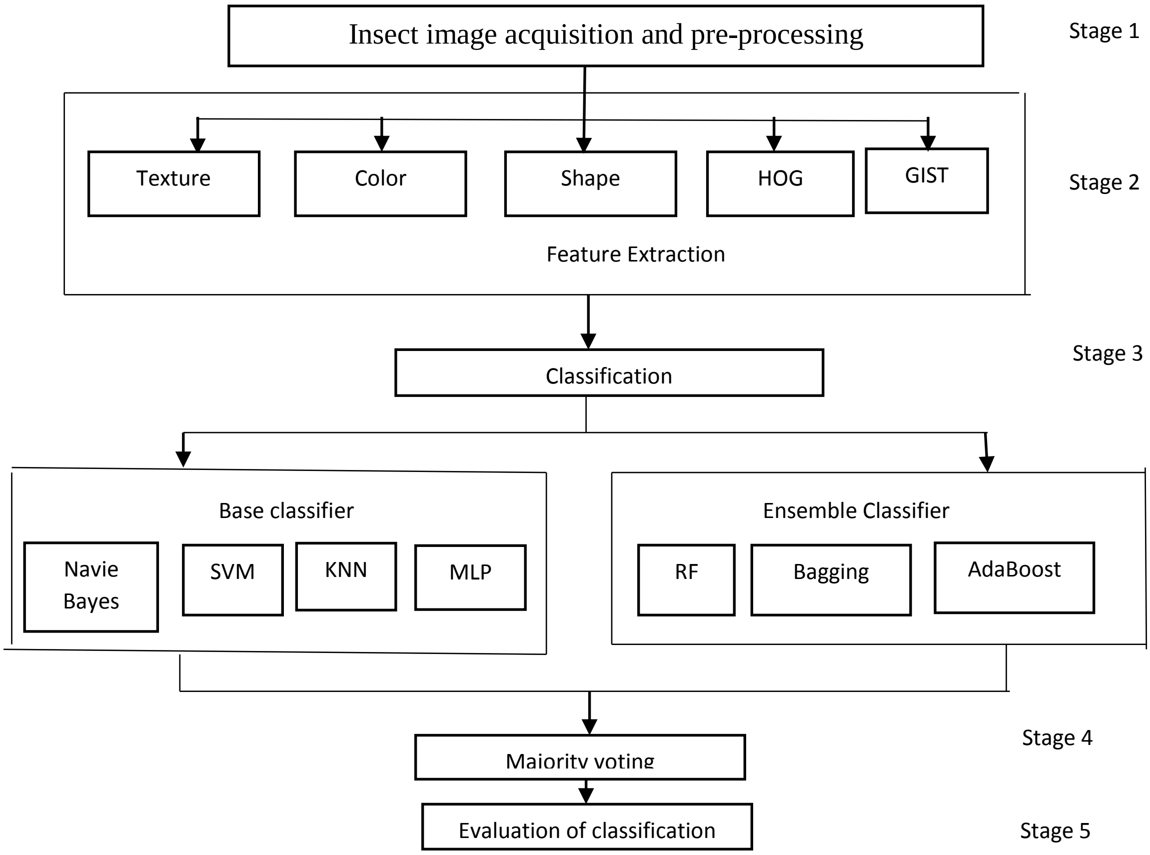


Figure. 1 Flow diagram of field crop insects classification system

Figure. 2 Xie insect images

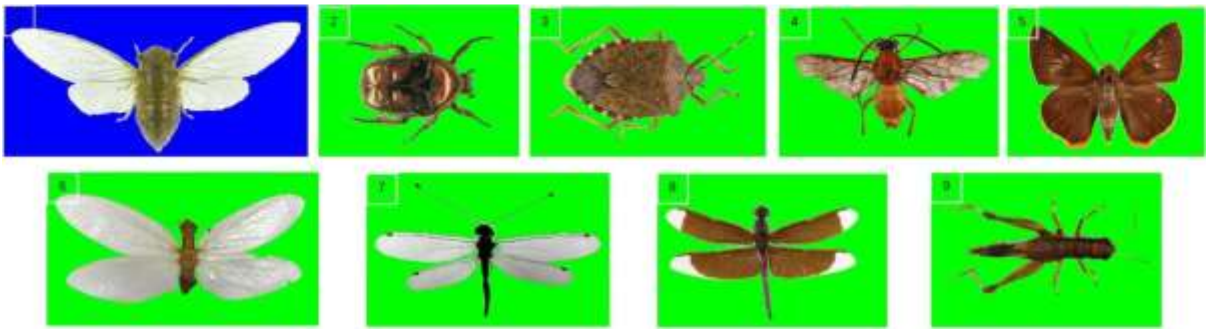


Figure. 3 Wang insect images

Figure. 4 Sample butterfly images

2.1.3 Dataset 3

To facilitate feature extraction, we choose 24 butterflies photos & downscale them to 227 by 227 pixels (Figure. 4).

2.2 Feature extraction

The relative data regarding the insect's form, colour, & texture is contained in its features. Pictures of butterflies had being identified by combining colour, form, as well as texture, whereas whiteflies had been seen in greenhouses using a combination of colour & shape-based detection methods. According to Qing et al., plant hoppers in rice paddy fields were automatically detected & counted using a combination of HOG features, colour, shape, & Haar properties. For leaf classification, a noise filtering technique is employed to the recovered invariant characteristics, such as moment invariants, convexity, perimeter ratio, multiscale distance matrix, average margin distance, as well as margin statistics. Moment of Hu theory, Legendre theory, Zernike theory, and the Walsh transform isolated printed Tifinagh characters can be recognised using a combination of texture & GIST as descriptors. This study involves extracting a main important characteristics from an insect image, including texture, colour, shape, HOG, as well as GIST features, & then combining them into extracted features to improve classifications accuracy. We'll go into the specifics of the extracting features strategies after the jump.

2.2.1 Texture

Insects can be identified & categorised in large part thanks to their textures. Several other textural extracting features strategies, including local binary pattern-LBP, grey level co-occurrence matrix-GLCM, & Gabor filters, have been suggested for insect identification. The study makes use of GLCM for texture feature extraction in insect photos. As a statistical technique, GLCM can be used to retrieve texture information from such an insects images by taking into account the spatial connection of pixels. Measuring the frequency that a pixel with grey level intensity values i occur in proportion to a neighbour image pixel 'j' yields a GLCM, this is used for texture feature analysis. Many statistical measurements that reveal an insect's texture was developed using the GLCM. Starting with the GLCM of each image, we extract three first-order histogram-based characteristics, comprising variance (average contrast), skewness, and kurtosis, as well as 5-second statistical measures, including contrast (inertia), correlation, energy, homogeneity, & entropy.

2.2.2 Color

In order to retrieve colour characteristics, several different approaches are used, such as colour histogram, colour moments, colour coherence vector (CCV) descriptor, as well as local colour contrastive descriptor (LCCD). The histogram of the insects images was used to extract colour characteristics. From the processing RGB colour insects images, we retrieve the separate red, green, & blue colour channel, yielding 3 separate 2-dimensional arrays, one for each colour components. The Tettigella viridis insect's RGB colour channel from the Xie insect dataset are depicted in Figure. 5. To use the image function in MATLAB, the histogram count value for the red, green, & blue channels were obtained, representing the 3 primary colour attributes for the bug images.

2.2.3 Shape

Determining the insect's form measurements using photographs is a popular application of shape characteristics for insects identification.

Both Qing et al. and Singh et al. used a total of 51 morphological characteristics, such as radial Fourier descriptors, border Fourier descriptors, & form moments, to detect insects in wheat kernals. In this study, we used the Sobel filter to recognize the edge of the insects images, & then further used morphological operations in MATLAB, such as the dilation, closing, & filling operations, initially extract the geometrical shape information. For the 4 insects class in the Xie insect datasets, including Tettigella viridis, Sogatella furcifera, Pieris rapae, & Eurydema gebleri, the outcomes of applying the Sobel filter & performing a morphological analysis are displayed in Figures 6. Ten geometric shape parameters, including perimeter, area, form factor, main axis length, eccentricity, minor axis length, solidity, compactness, circularity, and extent, were extracted to examine the insects' shapes.

They used the method detailed in our previous work to analyse & compute the first 9 shape characteristics. The Extent characteristic is specified as the fraction of the image's overall bounding box that falls within the insect's region of interest.

It's a present from,

$$\text{Extent} = \frac{\text{Area}}{\text{Bounding box area}}$$

While its size of the shortest rectangles that completely encloses the region is given as the value 1 in the vector 1 by Q 2, where Q 14 2 is the case for a two dimensional bug images.

2.2.4 HOG

When it comes to human detection, Dalal et al. recommended utilising overlapping local contrast normalizations that are made possible by properties of HOG. Three steps are required to complete HOG feature extractions.

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

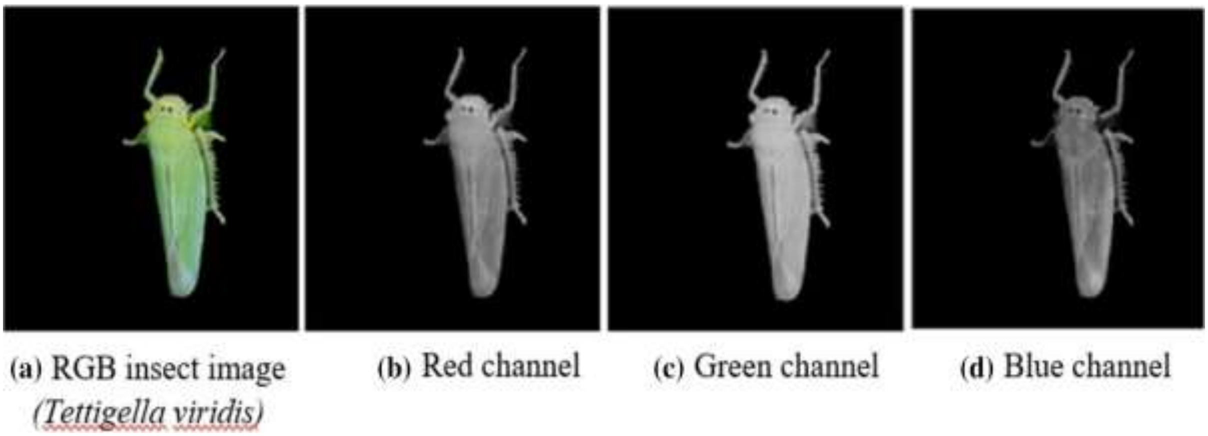


Figure. 5 RGB color channels

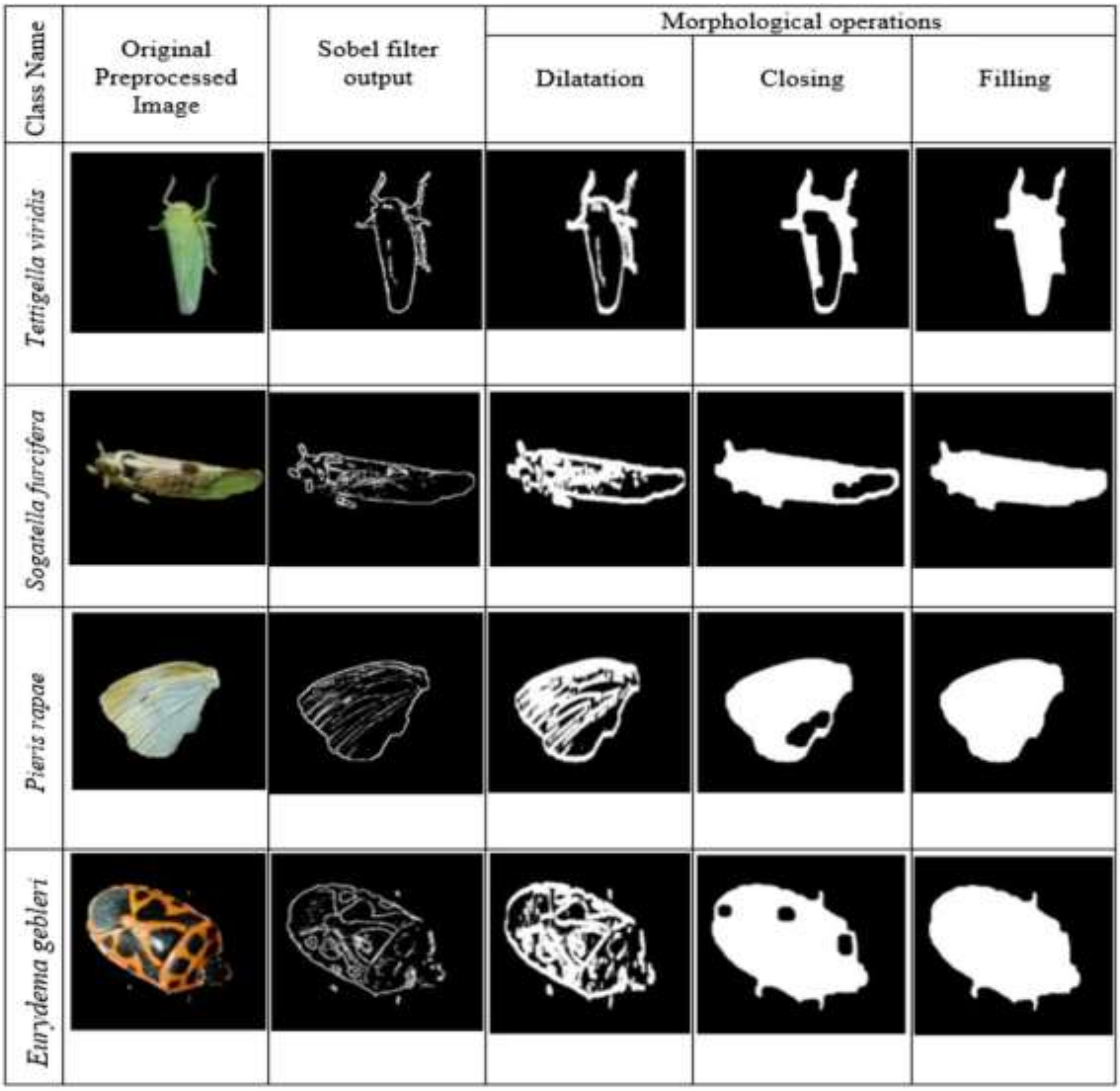


Figure. 6 Results from Sobel filter and morphological operations

the normalizing of blocks, the creation of a histogram, as well as the calculating of gradients. Before declaring an images to include insects, it is reduced in length to 64 by 128 pixels. This resulting images of the bug is then split into 16 9 16 pixels with such a 50% consisting of 105 chunks (7915=105) that overlay one another blocks. Every one of the 8 9 8 cells in a blocks includes two 9 2 of them. Pixels. Where the gradients are pointing & how big they are figured out on a block-by-block basis. Slope directions of a gradient were distributed evenly over nine angular bins, as well as their histogram. Every block's orientation is determined by a mathematical formula. The data points in the histogram are subsequently aggregated after being normalised. There was total of 3780 cells in the dataset (9 blocks x 4 cells x 9 bins).characteristics (per histogram = 105949).

2.2.5 GIST

Originally introduced by Oliva et al. for activity recognition using low-level aspects of the scene without using any segmentation methods, the GIST description has now seen widespread adoption. A primary spatial architecture of a scene can be described by a collection of perceptual dimensions including roughness, openness, ruggedness, & expansion. The GIST descriptions were universally applicable picture searching terms because the classifying of traffic scenes & improved precision when employing a variety of scaling, cropping, & compression techniques. When computing GIST characteristics, 32 Gabor filters were convolved with the insect images at 4 scalings & 8 orientations to produce 32 feature maps. The average features value was calculated across 16 regions (494 grid) in each features map. At last, 512 (16 x 32 = 512) GIST features are produced by concatenating the averages values of all 32 featuring maps.

2.2.6 Feature reduction by principal component analysis (PCA)

In addition to improve classification performance & data storage, Principal component analysis was used to minimise the dimensionality of HOG & GIST characteristics. Principal component analysis is an effective method for keeping a more important variants while decreasing the dimensionality of a data set featuring many connected variables. To prevent over fitting when developing classifier method in a learners, it is helpful to reduce the dimensions of the characteristics being used. Reduced-dimensional HOG-PCA & GIST-PCA characteristics are generated in MATLAB by the Principal component analysis operations being performed to both HOG & GIST characteristics. Usually the elements, account for 95percent of the variance are retained via Principal component analysis in MATLAB.

2.3 Classification

The main popular insect classifications machine learning-ML methods include linear discriminant analysis, support vector machines, decision trees-DT, radial basis functions, neural networks-NN, & closest neighbours. Insects were classified in this study using retrieved characteristics, such as texture, colour, form, HOG, & GIST. This collection of characteristics is then used to train classifications on the insects' phenotypes. They used 3 distinct bug dataset to test out base & ensembles classifications.Navie Bayes, Support Vector Machine, K - nearest neighbors, & MLP were the primary classification methods employed. Ensembles classifications like Random Forest, bagging, & XGBoost were employed. As comparison to a singular models, the classifying accuracy of ensembles classifications is significantly higher because they include numerous base classifier.

2.4 Majority voting

Last but not least, to further enhance the classifications efficiency, a majority polling combinations rules were implemented to the base & ensembles classifications. Think about the 'n' classification methods, h1(X), h2(X),..., hn (X). The overall voting classification is the result of combining all the separate classifications together,

$$C(x) = mode\ \{h_1(X), h_2(X) \dots \dots . h_{n(x)}\}$$

They receive the more vote and classifications and hence give results superior to the particular classifications.

3 Results and discussion

3.1 Insect datasets

Tests of classifications efficacy was conducted using information from 3 different insects' datasets: the Xie datasets, the Wang datasets, & the butterflies picture datasets. In order to increases the size of insects datasets, image augmentation technologies were applied to insects images, including left rotation (at 90 degrees), right rotation (at 90 degrees), left-to-right flipping regarding the vertical axis, top-to-bottom flipping about the horizontal axis, and scaling (90%, 75%, as well as 60percent of original insects image). Optimal efficiency is achieved in classifications situations when the train-test ratios is 70-30%. All insects datasets used in the suggested research was divided into a 70percentage training dataset as well as a 30percentage testing datasets. Extensive information for all 3 insects datasets are given in Tables 1.

3.2 Classification accuracy of base and ensemble classifiers for different combination of features

The following characteristics have been retrieved from the insects images: 8 texture characteristics contrast, correlation, energy, homogeneity, entropy, variance, skewness, & kurtosis, 3 color characteristics maximum histogram counts for red, green, & blue channels, 10 shape characteristics area, perimeter, major axis length, minor axis length, eccentricity, circularity, solidity, form factor, compactness, and extent, HOG-PCA & GIST MATLAB 2017a was used for the execution of all of the features extractions methods, as well as the SKLEARN machine learning-ML framework was selected for classifications.

The classification of the insect was accomplished by separately employing 4 basic classifications (Navie Bayes, Support Vector Machine, K - nearest neighbors, & MLP) as well as 3 ensembles classifications (Random Forest, Bagging, & XGBoost). In order to validate the samples, a k-fold cross-validation is performed, & the value of k is set to 10 so that the predicted accuracy can be improved.

The datasets including insects is partitioned onto k subgroups, of which 1 of the k subgroup was selected at random to serve as the testing set, while the remaining subgroups, k minus one, serve as the training set. This process is done as many times as necessary until all folding has been put through its paces. An evaluation of the classification accuracy can be obtained by taking the sum of the accuracies that were obtained from the k separate instances of cross-validation.

Insect dataset	No. of classes	No. of insect images	No. of training images	No. of testing images
Xie insect dataset	24	10,344	6144	6144
Wang insect dataset	9	2855	1680	1175
Butterfly image dataset	24	1604	944	660

Table 1 Details of insect datasets

3.2.1 Performance studies of base classifiers

The suggested insect identifying method was initially tested by applying base classifiers, which includes Navie Bayes, Support Vector Machine, and K - nearest neighbors, as well as MLP for the combinations of textures, colour, & shapes, as well as HOGPCA & GIST-PCA for each of the 3 insects dataset independently. Figure 7a–c illustrates the classification accuracy produced from these base classifications by using 5 distinct features combinations for the Xie insects datasets, the Wang insect dataset, as well as the butterflies picture datasets, respectively. Every specific use of a solitary features during the classifications stage (texture, colour, shape, HOG-PCA & GIST-PCA) could be challenging due to the tiny difference among its results [5]. Once taking into account texture, colour, & shape features combinations, it has been found that the performance of texture? colour feature features combinations is lower while comparison to an effectiveness of colour? Shape & texture? Shape features combinations in Naive Bayes, Support Vector Machine, K - nearest neighbors, & MLP base classification methods for all 3 insect datasets. This is the case regardless of which insect dataset is being used.

As can be seen in Figure 7a & 7b, the combining of colour & shapes features results in superior performance in MLP and KNN for the Xie insects datasets and the Wang insects datasets, correspondingly. It should also be mentioned that the MLP base classification achieves satisfactory results for butterflies' insects datasets which contain textures and form information (Figure. 7c).

In the MLP analysis of the Xie insects' datasets and the butterflies picture datasets, the accuracy of the textures, colour, as well as form characteristics was found to be the greatest. On the others hand, the Support Vector Machine analysis of the Wang insects datasets produced the best results when comparing to the other classifications. The MLP neural network-NN classifier was developed specifically for use in multi-class classifiers situations, as it implements the cross-entropy loss function. The datasets compiled by Xie has 24 distinct classes of insect, each of which has a unique textures, colour, & size (Shape). Therefore, the textures, colour, and shape features combination seems to be successfully more useful in identifying insects with good categorization results using MLP classification model for Xie's dataset comparison to others.

This is because it minimises the loss function, which ensures that the classifying accuracy is high. When comparison to textures, colour, & form attributes, the improvements brought about by HOG-PCA & GIST-PCA is far highly efficient. The comparison outcomes, which can be seen in Figure 7a–c, revealed that the performance of the texture, colour, & shape characteristics significantly improved by combination of HOG-PCA and GIST-PCA characteristics, in comparison to the individual performances. The effectiveness of GIST features combinations yielded higher results in terms of classifications accuracy. Therefore, the GIST-PCA characteristic could be useful when combined for the classification of insects. Its accuracy is obviously high, as its discovery verified that utilising low-level features provides useful data on spatial scales without the application of segmentation from the datasets. As a result, the GIST is capable to rapidly restrict local characteristics & improve insects recognitions. It can be inferred that higher classifying accuracy can be reached in MLP classifiers for Xie insect dataset (86.37%) & butterflies picture datasets (80.12%), &in Support Vector Machine classifiers for Wang insect dataset (86.52%), by incorporating all 5 variables, namely textures, colour, form, & size the HOG-PCA as well as the GIST-Principal Component Analysis

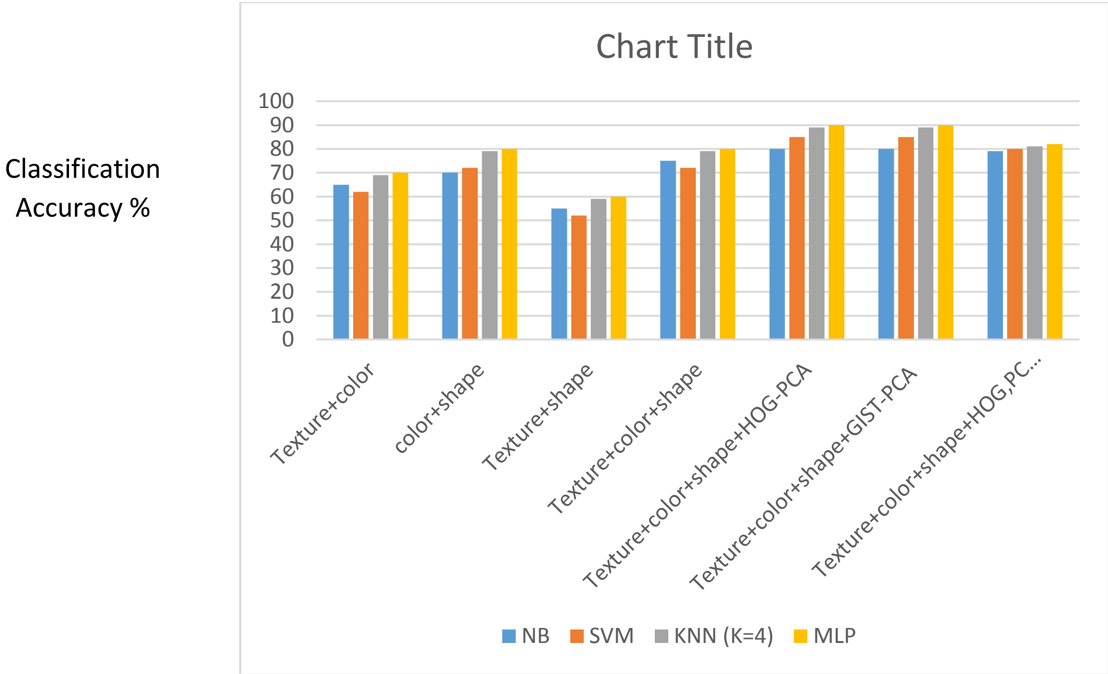
3.2.2 Performance studies of ensemble classifies

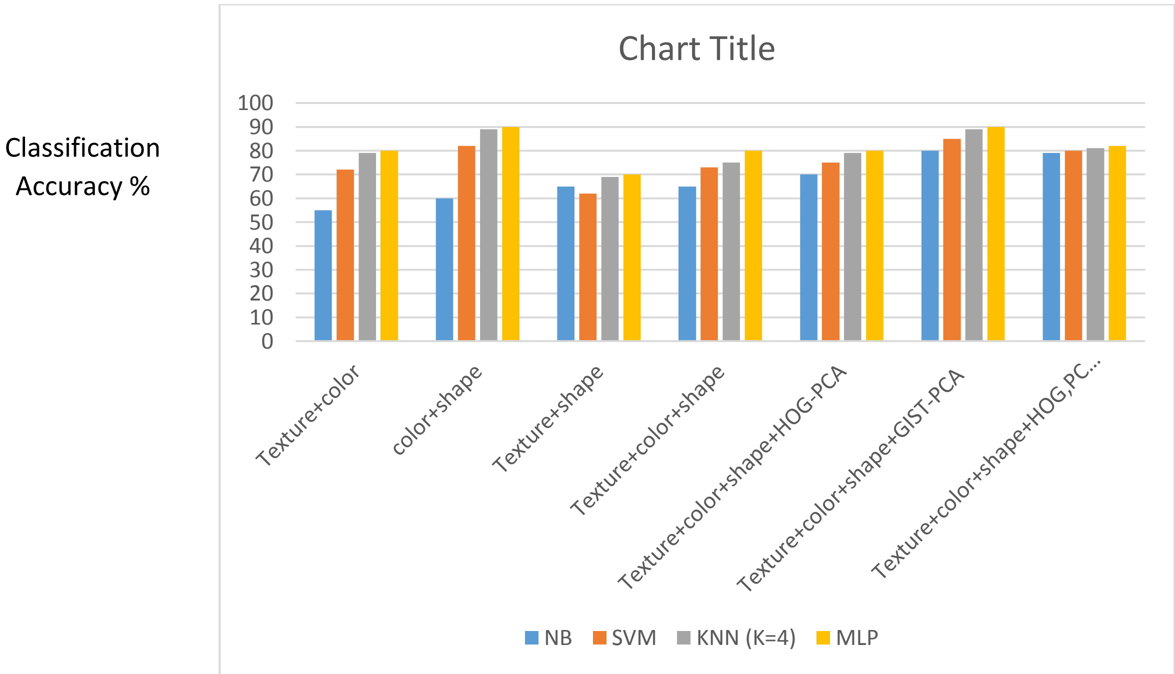
3 distinct ensembles classifications, including Random Forest, Bagging, & XGBoost method, was evaluated one-by-one with 10-fold cross-validation for 3 insects datasets, each of which contained a unique combinations of features. The goal of this exercise was to improve the classifications accuracy. Figure 8a–c illustrates the obtained outcomes of the ensembles classification for the Xie insects datasets, the Wang insects datasets, & the butterflies picture dataset, respectively. The amount of trees used in the Random Forest strategy is set at 100. The fast decision tree learners, also known as Retire, as well as the J48 decision tree was used in bagging & XGBoost, correspondingly, as their respective basic learning algorithms. It can be shown from these figures that the Random Forest classifier performs better than the remaining 2 ensembles classifiers (Bagging & XGBoost) in all 3 datasets for a variety of features combination. In addition, Random Forest classifiers can handle a very huge number of input characteristics while simultaneously reducing the amount of time required for the procedure .

When compared with Random Forest and XGBoost, the outcome of the Bagging classifiers was much lower across all 3 datasets. With the combinations of all 5 features, the Random Forest classifier for the Xie datasets achieved an accuracy of 89.57%, while the butterflies image datasets achieved an accuracy of 91.96%. The XGBoost classifications for the Wang dataset achieved an accuracy of 95.89%. The outcomes of the study showed these results.

There has been a significant enhancement in classification accuracy whenever these ensembles classifier outcomes were comparison with the base classification models that are described in Section 3.2.1. This advancement was 3.2% for the Xie image datasets, 9.37% for the Wang image datasets, & 11.84% for the butterflies image datasets, respectively.

Figure. 7 Base classifiers results for a Xie insect dataset, b Wang insect dataset and c butterfly image dataset





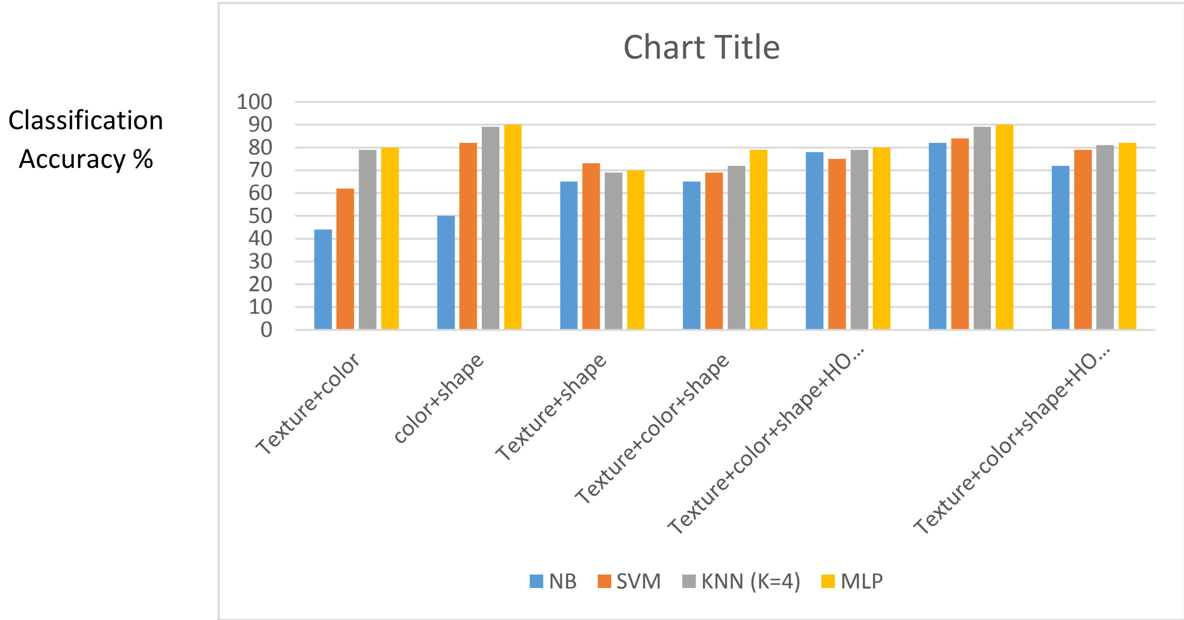
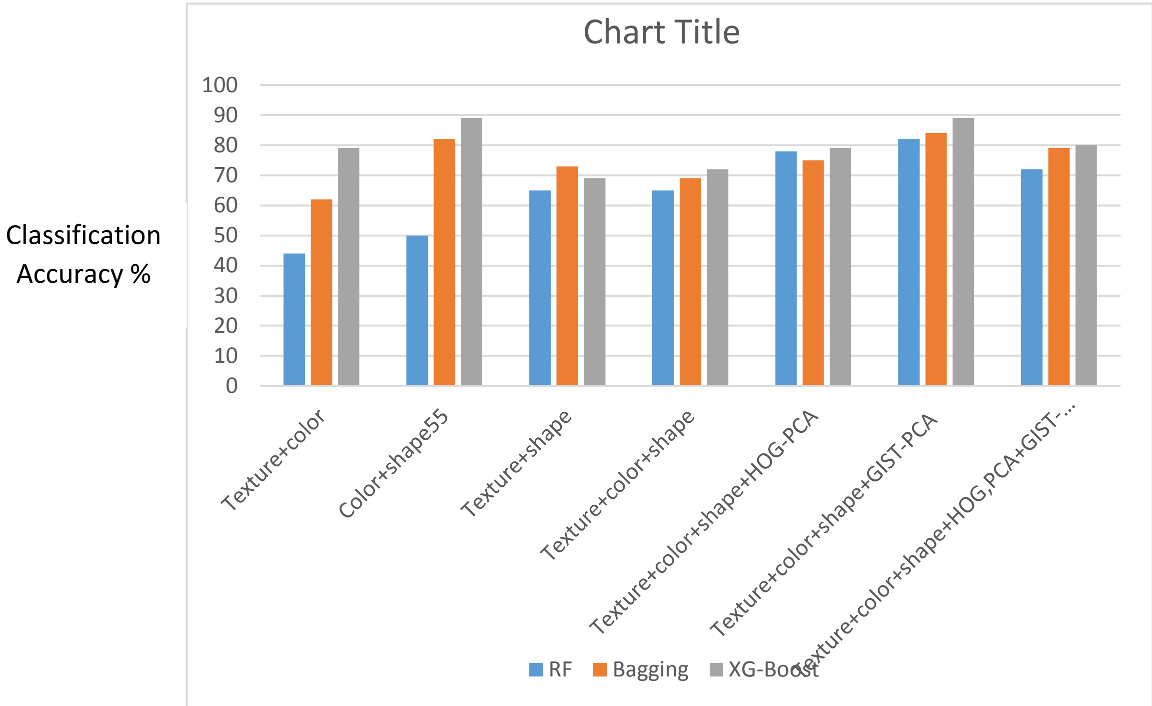
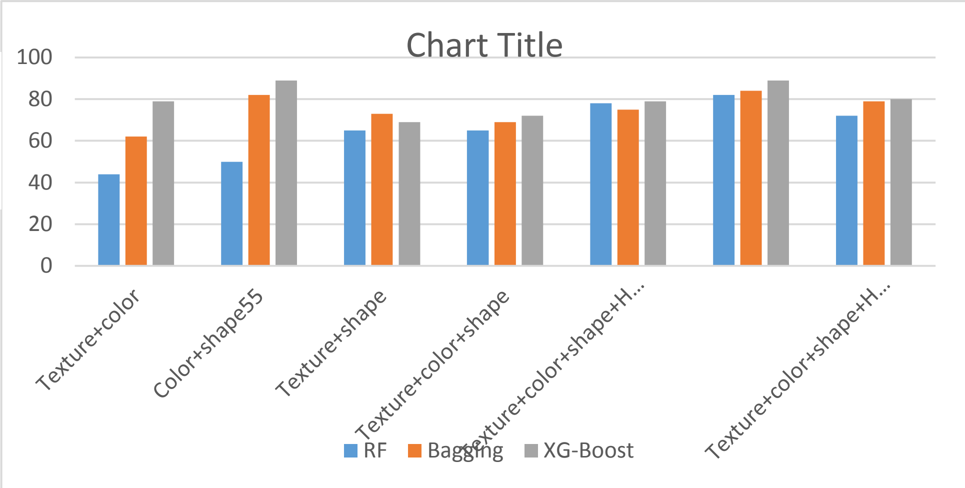


Figure. 8 Ensemble classifiers results for a Xie insect dataset, b Wang insect dataset and c butterfly image dataset



Classification Accuracy %



Classification Accuracy %

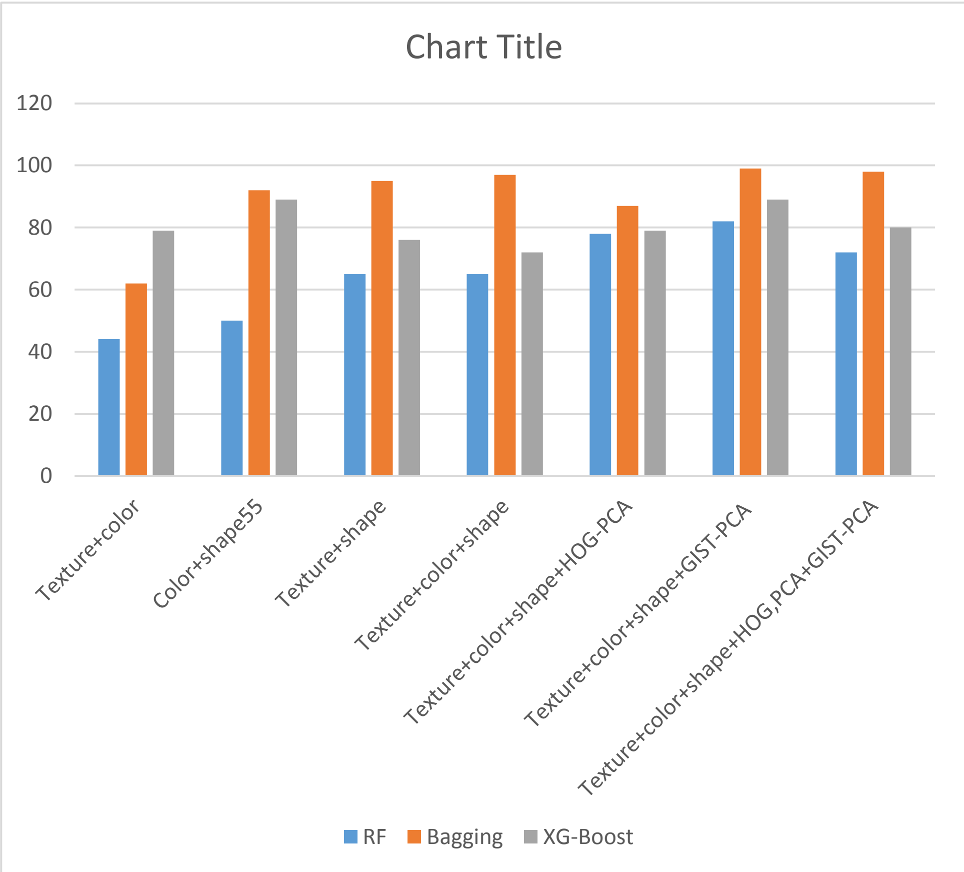


Table 2 Classification accuracy values (%) of majority voting

Feature combination	Majority voting						
	Base classifiers (fusion of NB, SVM, KNN and MLP)					Ensemble classifiers (fusion	
Data set	Xie et al.	Wan get al.	Ours	Xiao et al. (20classes)	Ours		voting
Xie insect dataset	–	–	80.9 ± 0.37	–	81.7 ± 0.69	91.2 ± 0.50	92.1 ± 1.10
Wang insect dataset	70.0 ± 1.10	92 ± 0.0	86.5 ± 0.48	–	85.6 ± 0.40	90.3 ± 1.40	96.5 ± 0.80
Butterfly image dataset	80.1 ± 1.60	–	77.2 ± 0.26	78.0 ± 0.0	78.8 ± 0.34	97.2 ± 1.00 (20 classes)	92.3 ± 1.42

of RF, bagging and XGBoost)

	Xie insect data set	Wang insect dataset	Butterfly image dataset	Xie insect data set	Wang insect dataset	Butterfly image dataset
Texture + color	74.16 ± 0.66	70.19 ± 0.87	63.11 ± 0.44	81.09 ± 0.63	72.79 ± 0.87	74.74 ± 0.22
Color +shape	79.01 ± 0.79	84.33 ± 0.91	67.49 ± 0.03	87.75 ± 0.72	87.73 ± 0.76	85.89 ± 0.04
Texture +shape	72.69 ± 0.77	83.39 ± 0.36	68.63 ± 0.20	85.05 ± 0.63	84.76 ± 0.19	79.28 ± 0.84
Texture + color+ shape	80.10 ± 0.71	86.05 ± 0.23	73.85 ± 0.35	86.87 ± 0.18	88.45 ± 0.49	80.50 ± 0.70
Texture+ color+ shape + HOG-PCA	81.88 ± 0.34	86.05 ± 0.23	76.25 ± 0.12	88.38 ± 0.47	93.39 ± 0.67	88.66 ± 0.61
Texture + color + shape+ GIST-PCA	86.06 ± 0.70	87.45 ± 0.56	77.52 ± 0.41	89.47 ± 0.06	94.67 ± 0.53	91.47 ± 0.43
Texture + color + shape+ HOG-PCA + GIST-PCA	89.76 ± 0.76	90.02 ± 0.92	84.70 ± 0.77	92.09 ± 0.55	96.48 ± 0.34	92.37 ± 0.28

Table 3 Comparison of classification accuracy (%) for texture +color + shape + HOG-PCA + GIST-PCA feature combination in three public insect datasets

SVM KNN Xie et al.

MKL our majority

3.3 Performance study of majority voting

For each of the 3 insects datasets, majority polling has been implemented to the outcomes from of the features combination of texture, colour, shape, HOG-PCA, & GIST-PCA using fusion of all base (NB, SVM, KNN, & MLP) and fusion of all ensemble (RF, bagging, & XGBoost) classification models. This was done in order to ensure that the most accurate results were obtained. Table S4 of the supplementary information presents the base and ensemble classifier setup parameters of the SKLEARN tool that were utilised in this work. Table 2 displays the outcomes of the categorization process that uses majority voting to forecast which category (class) will have the most members. It can be seen from Table 2 that applying majority polling outcomes with integrating RF, bagging, & XGBoost ensembles classifications results in higher classification outcomes across all 3 insects datasets. This is the case when the datasets are analysed using XGBoost, bagging, as well as bagging with RF. In addition, the outcomes of majority voting were more favourable for the fusion of HOG-PCA & GIST-PCA features combinations, as well as texture, colour, & shape.

The categorization outcomes achieved using SVM, KNN, & majority voting (fusion of RF, Bagging, & XGBoost ensembles classifications) for the 3 insects datasets are compared and contrasted in Table 3. A vote accuracy rating of 92.1percentage points were found for the majority vote, whether it be for texture, colour, or shape. While applied to Xie's insects dataset, the HOGPCA and GIST-PCA features combinations was examined. The merger of many ensembles classifications was necessary in order to get this higher level of accuracy. The integrated features of our system achieved the best level of accuracy (96.5%), when measured against the results of the majority voting in the Wang insect dataset. As can be seen in Table 3, the accuracy of majority voting was found to be 92.3% across the 24 different classes of butterfly images in the dataset. When compared to the accuracy of 20 different classes of butterfly images, this reliability becomes significantly more important. The findings demonstrated that our approach was capable of accurately classifying insects, despite the fact that many insects share visual, geometric, and textural characteristics.

3.3.1 Evaluation metrics for majority voting classifier

The efficiency of the classifications model is evaluated based on its metrics, which include texture, colour, and shape, as well as majority vote. HOG-PCA - GIST-PCA capability

Table 4 Majority voting classifier metrics for base and ensemble classifiers with texture +color + shape+ HOG-PCA+ GIST-PCA feature for three insect datasets

Classifier metrics	Fusion of base classifiers				Fusion of ensemble classifiers			
	Precision	Recall	*F*-measure	ROC area	Precision	Recall	*F*-measure	ROC area
Xie insect dataset	0.899 ± 0.09	0.898 ± 0.08	0.897 ± 0.05	0.947 ± 0.03	0.937 ± 0.01 0.11	0.934 ±	0.935 ± 0.08	0.965 ± 0.21
Wang insect dataset	0.953 ± 0.06	0.970 ± 0.01	0.962 ± 0.02	0.982 ± 0.05	0.994 ± 0.08 0.02	0.975 ±	0.984 ± 0.04	0.987 ± 0.05
Butterfly image	0.927 ± 0.01	0.924 ± 0.07	0.921 ± 0.01	0.959 ± 0.06	0.968 ± 0.07 0.05	0.938 ±	0.952 ± 0.01	0.968 ± 0.03

Combinations. Classifications metrics such as precision, recall, F-measure, & receiver operating characteristic-ROC area were computed as following for such purpose of studying classifications performances.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{TP + FN}$$

$$\mathrm{F - measure} = 2\mathrm{x}\frac{(\mathrm{Recall\ X\ Precision})}{(Recall + Precision)}$$

Where TP stands for "true positive-TP," FP for "false positive-FP," and FN for "false negative-FN."
To calculate that likelihood that the "positive (+ve)" class of insects would be ranked higher than the "negative (-ve)" class by the classifications used, calculate the area under the receiver operating characteristic curve. Maximum levels of precision, recall, F-measure, and ROC area indicate that the fusion of ensembles classification models outperforms the fusion of base classifications across all 3 datasets (see Table 4).
It should be pointed out that the Xie insect datasets, the Wang insect dataset, as well as the butterflies picture datasets all contain numerous classes of insects, as well as that the area under the ROC curve scores were over 0.9, representing greater classifications accuracy of insects. Thus, the popular vote strategy that incorporates textures, colour, and shape, HOG-PCA, & GIST-PCA features is appropriate for the classification & identification of insects.

4 Conclusion

Insects pose a significant threat to agricultural fields, therefore this research presents a method for identifying & categorising these pests using machine learning-ML techniques. Base (Navie Bayes, Support Vector Machine, K - nearest neighbors, & MLP) & ensembles classifications had also applied to all the possible permutations of features such as textures, colour, shape, HOG, & GIST (RF, Bagging and XGBoost). Using majority polling on both the classification algorithm as well as the ensembles classification enhanced the reliability of the classification. It has been shown experimentally that majority voting in ensemble classifiers improves performance over other methods. Incorporating the features of textures, colour, form, HOG-PCA, & GIST led to the greatest classification accuracy of 92.1%, 96.5%, & 92.3%, respectively, from majorities vote outcomes for the Xie insects datasets, the Wang insects dataset, as well as the butterflies picture datasets. To train an effective network for insects classification, a larger number of insects training examples are needed because insects share many comparable properties. Research had shown that minority voting in ensembles classifications can help entomologists identify insects in agricultural plants. Farmers & academics would benefit from this suggested research because it would aid them in spotting insects in their crops sooner rather than later. In order to enhance classifier using real-time insect information, we would integrate deep learning-DL algorithms for faster training of insect's photos.

References

[1] John V.Stafford , "Implementing Precision Agriculture in the 21st Century" Journal of Agricultural Engineering Research,Volume 76, Issue 3, pp. 267-275, July 2000.
[2] Sami Khanal et.al "An overview of current and potential applications of thermal remote sensing in precision agriculture", Journal of Computers and Electronics in Agriculture, Volume 139, Pages 22-32, 15 June 2017.
[3] iqian Zhang et.al "Precision agriculture—a worldwide overview", Journal of Computers and Electronics in Agriculture, Volume 36, Issues 2–3, Pages 113-132, November 2002.
[4] hristopher Brewster et.al "IoT in Agriculture: Designing a EuropeWide Large-Scale Pilot", IEEE Communications Magazine, Volume: 55 , Issue: 9 ,pp.26-33, Sept. 2017
[5] Nurzaman Ahmed et.al "Internet of Things (IoT) for Smart Precision Agriculture and Farming in Rural Areas" IEEE Internet of Things Journal, Vol. 5, No. 6, December 2018.
[6] M. R. M. Kassim, I. Mat, and A. N. Harun, "Wireless sensor network in precision agriculture application," in Proc. IEEE Int. Conf. Comput. Inf. Telecommun. Syst. (CITS), 2014, pp. 1–5.

[7] Y. Zhu, J. Song, and F. Dong, “Applications of wireless sensor network in the agriculture environment monitoring,” Procedia Eng., vol. 16, pp. 608–614, Nov. 2011.

[8] T. Ojha, S. Misra, and N. S. Raghuwanshi, “Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges,” Comput. Electron. Agricult., vol. 118, pp. 66–84, Oct. 2015.

[9] Juan M. Nunez V et. al, “Design and implementation of WSN for precision agriculture in white cabbage crops”, IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON) ,pp. 1-4, 2017.

[10] T. Ojha, S. Misra, and N. Singh, “Wireless sensor networks for agriculture : The state-of-the-art in practice and future challenges,” Comput. Electron. Agric., vol. 118, pp. 66–84, 2015.

[11] Kiran Kumar M., Kranthi Kumar S., Kalpana E., Srikanth D., Saikumar K. (2022) A Novel Implementation of Linux Based Android Platform for Client and Server. In: Kumar P., Obaid A.J., Cengiz K., Khanna A., Balas V.E. (eds) A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems. Intelligent Systems Reference Library, vol 210. Springer, Cham. https://doi.org/10.1007/978-3-030-76653-5_8

[12] Shravani, C., Krishna, G. R., Bollam, H. L., Vatambeti, R., & Saikumar, K. (2022, January). A Novel Approach for Implementing Conventional LBIST by High Execution Microprocessors. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 804-809). IEEE.

[13] Nagendram, S., Nag, M. S. R. K., Ahammad, S. H., Satish, K., & Saikumar, K. (2022, January). Analysis For The System Recommended Books That Are Fetched From The Available Dataset. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1801-1804). IEEE.

[14] Jothsna, V., Patel, I., Raghu, K., Jahnavi, P., Reddy, K. N., & Saikumar, K. (2021, March). A Fuzzy Expert System for The Drowsiness Detection from Blink Characteristics. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 1976-1981). IEEE.

[15] Appalaraju, V., Rajesh, V., Saikumar, K., Sabitha, P., & Kiran, K. R. (2021, December). Design and Development of Intelligent Voice Personal Assistant using Python. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 1650-1654). IEEE.

[16] Naidu, T. P., Gopal, K. A., Ahmed, S. R., Revathi, R., Ahammad, S. H., Rajesh, V., ... & Saikumar, K. (2021, December). A Hybridized Model for the Prediction of Heart Disease using ML Algorithms. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 256-261). IEEE.

[17] Teju, V., Sowmya, K. V., Yuvanika, C., Saikumar, K., & Krishna, T. B. D. S. (2021, December). Detection of Diabetes Melittus, Kidney Disease with ML. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 217-222). IEEE.

[18] Mannepalli, K., Raju, K. B., Sirisha, J., Saikumar, K., & Reddy, K. S. (2021, December). LOW complex OFDM channel design using underwater-acoustic-communication using machine learning techniques. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1505-1513). IEEE.

[19] Kumar, K. S., Vatambeti, R., Narender, M., & Saikumar, K. (2021, December). A real time fog computing applications their privacy issues and solutions. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 740-747). IEEE.

[20] Ajay, T., Reddy, K. N., Reddy, D. A., Kumar, P. S., & Saikumar, K. (2021, December). Analysis on SAR Signal Processing for High-Performance Flexible System Design using Signal Processing. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 30-34). IEEE.

[21] Raju, K. B., Lakineni, P. K., Indrani, K. S., Latha, G. M. S., & Saikumar, K. (2021, October). Optimized building of machine learning technique for thyroid monitoring and analysis. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1-6). IEEE.

[22] Nagavarapu, S., & Bhavani, V. K. (2020). The REDUCTION OF TRAFFIC LOAD IN CLOUD COMPUTING USING ENERGY EFFICIENT CLUSTERING TECHNIQUE. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 11(3), 1118-1124.

[23] Nagavarapu, S., Krishna, M. V., & Pavan, N. Cloud Storage: Data Reliability Solutions. Vol 11, Issue 10,OCT /2020 Journal of Engineering Sciences (JES).

[24] Detecting cyber-attacks on web applications by applying different machine learning techniques

[25] An efficient secure matching rank search mechanism over encrypted cloud data.

[26] Dr. MANAM VAMSI KRISHNA Mrs. BHAVANI KOGANTI, MACHINE LEARNING TECHNIQUES FOR DETECTING CYBER ATTACKS I N NETWORK'S, 9(36) 1-10. 2021/12, 2021/12

[27] Priyanka, B., & Krishna, M. V. Multifactor Authentication and Lightweight Cryptography are used to create a scalable and secure Big Data IoT system. 2021/11 IJARST.