LEVERAGING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES FOR ACCURATE CLASSIFICATION OF STORED GRAIN PESTS

M S Santhanambika[1], G Maheswari[1]*, N Valliammal[1] and G Sudhamathy[1]

[1]Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore 641043, Tamil Nadu, India
*Email: maheswari_zoo@avinuty.ac.in (corresponding author): ORCID ID 0000-0002-9764-4819

ABSTRACT

The present study proposes a methodology utilizing machine learning and deep learning techniques for stored grain insect pest classification. Relevant morphological features extracted from captured pest images were fed to K-nearest neighbors (KNN), Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Naïve Bayes (NB) algorithms. The effectiveness of the proposed approach was evaluated using a comprehensive dataset compiled with images of various stored grain insect pests. The order of classification accuracy was NB < KNN < SVM < CNN where KNN achieved 76% accuracy, SVM exhibiting 81% accuracy, CNN achieving 98% accuracy, and NB achieving 33% accuracy. Though CNN required more computation time for classification, better accuracy was achieved and this could be utilized to identify the insects infesting stored food grains. The intelligent classification provides a valuable tool for identifying and differentiating stored grain insect pests, the primary step in IPM.

Key words: Image dataset, classification models, machine learning K-nearest neighbor (KNN), support vector machine (SVM), convolutional neural network (CNN), Naïve Bayes (NB)

Globally 70% of the human population depends on agriculture for their income (Devipriya et al., 2022). Agricultural productivity is affected by various factors, including the grain loss which is estimated to be 25% whereas 6% of it happens during storage (Shankar and Adrol, 2012). Though India is the world's largest producer of pulses, its productivity is low because of biotic and abiotic stresses (Kumari et al., 2022). Technical care in grain storage is essential to gain maximum profit, as there may be invasion of insect pests, pathogens and rodents. Stored grain pests are a diverse group of insects posing significant threats to the quality and quantity of stored products, resulting in economic losses and food security concerns (Wilberforce and Kalita*,* 2023).The identification of stored grain pests plays a crucial role in effective pest management strategies. The major pests of the stored grains come under the order Coleoptera, Lepidoptera, Psocoptera, and the class Acarina. These are categorized into Primary and secondary storage pests based on their type of infestation (Deshwal, 2020). The primary storage pests are the insects that damage whole or healthy grains. These are further classified into internal and external feeders depending on their feeding habits (Singh and Saini, 2023). The internal feeders are the insects or larvae that feed within the kernels of whole grains or seeds, leaving behind hollow shells when they emerge as adults whereas the external feeders feed on germ and endosperm from outside. The secondary storage pests are the insects that damage broken or already damaged grains (Ahmad et al., 2021). These pests can cause significant economic losses by reducing the quality and quantity of stored grains, leading to spoilage and contamination of stored goods. The pertinent pest management tool requires effective identification based on their characteristics and behaviour to control these pests (Daglish et al., 2018).

Small-scale farmers traditionally store their grains in small quantities at home, while larger quantities are stored in warehouses for future use. In addition to relying on artificial pesticides and insecticides to minimize storage losses and maximize yields, there is a growing push to incorporate modern technological solutions utilizing machine learning and deep learning techniques. These interventions aid in detecting the presence of insect pests, identifying them, and preventing damage to stored grains. Accurate identification and classification of these pests are essential for implementing suitable management techniques to minimize the infestation risks (Hagstrum and Finn, 2012). In recent years, machine learning algorithms have emerged as powerful tools for automated and efficient pest classification. Saikumar et al., 2023 validated the machine learning models for the classification and detection of the major insect pests in brinjal. Mendoza et al., 2023 built a simple insect detection system for stored grain using a camera, a low-power computer, and a trained deep learning model. This system identify insect pests in real-time under different lighting conditions, potentially aiding in faster and more effective insect control. Shi et al., 2020 proposed a method for automatically detecting and identifying eight common stored grain insects. The key innovation is a multiscale training strategy that extracts detailed features from images and locates potential insects.

Integrating machine learning algorithm enhances the efficiency and scalability of pest classification, accommodating a wide range of stored grain pest species. The present study advances stored grain pest management by leveraging machine learning algorithms for accurate classification. The necessity of using CNN, KNN (Kasinathan et al., 2021), NB and SVM (Quan et al., 2018) for the classification of stored grain pests is driven by several factors, including the complexity of pest identification, the need for accuracy and efficiency in pest management, and the availability of diverse data types. The raw images of *Sitophilus oryzae* Linn. (rice weevil), *Sitophilus granarius* Linn. (Wheat weevil), *Rhizopertha dominica* Fab. (lesser grain borer), *Callosobruchus maculatus* Linn. (cowpea weevil), *Callosobruchus chinensis* Linn. (adzuki bean weevil), *Tribolium castaneum* Herbst. (red flour beetle) and *Oryzaephilus surinamensis* Linn. (saw-toothed grain beetle) were employed in this study.

MATERIALS AND METHODS

The mass trapping of insect pests from stored cereals and pulses were performed with probe traps and an insect removal bin developed by TNAU without any attractants or chemicals (Mohan and Rajesh, 2016). The traps and bins with 2 mm perforation for cereals and 3mm for pulses were procured from Vridha Traders, Coimbatore. The ensnared insects' images were captured with a high-resolution camera to create a dataset. The stored grains infested with insects were collected from the local traders and households in and around Coimbatore city using traps and insect removal bins. The collected insects were identified using the standard identification keys of Southgate et al. (1957), Howe and Curie (1964) and Halstead (1993). The pests from stored cereals, *S. oryzae, S. granarius, O. surinamensis*, *R. dominica* and *T. castaneum* were collected using 2mm diameter perforated probe trap and insect removal bin (Mohan and Rajesh, 2016). While the pests of stored pulses, *C. maculatus*, and *C. chinensis* were collected using a 3mm probe trap and insect removal bin. The collected stored grain insect pests were focused under Stereo Binocular Microscope (Karl Zeiss Stemi DV4) and their images were captured using Samsung galaxy note 10 plus mobile phone to create a real time dataset. The dataset has a total of 789 images belonging to 7 classes, *C. chinensis* (64 images), *C. maculatus* (28 images), *O. surinamensis* (162 images), *R. dominica* (260 images), *S. granarius* (43 images), *S. oryzae* (100 images) and *T. castaneum* (127 images). The entire dataset was divided into a train-test ratio of 70:30. The image representation of the datasets were given in Fig. 1. All the images was employed with enhancement techniques (Krizhevsky et al., 2017) to reduce the noise and sharpen the images for better accuracy. This improves the quality of the image for better detection and classification of insects. The laptop used for this investigation has 8 GB RAM, an Intel® core ™ i5-8250U CPU, 64-bit Windows running at 1.60GHZ and the python programming language with Tensorflow and Keras packages.

Data augmentation was employed during the training phase to increase the diversity of the training dataset and improve the model's generalization ability (Fig. 2). Augmentation techniques such as shear transformations, zooming, and horizontal flipping were applied to the training images using the ImageDataGenerator class. Therefore, data augmentation helps to recognize complex objects (Krizhevsky et al., 2017). Leveraging more data for training deep learning neural networks and augmentation techniques to generate several variants of the images can improve the models' performance and help them adapt to unseen images.

C. maculatus *C. chinensis* *T. castaneum* *O. surinamensis* *S. oryzae* *S. granarius* *R. dominica*

Fig. 1. Image representation of the dataset

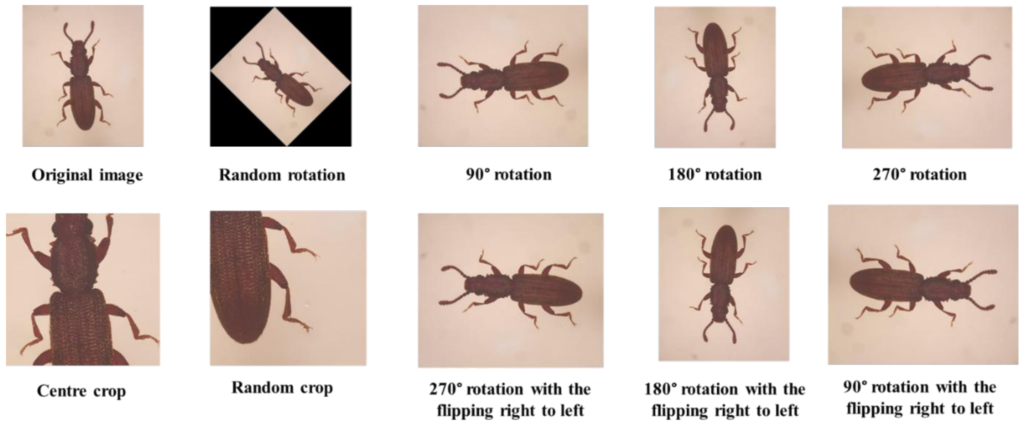

Fig. 2. Image augmentation of *Oryzaephilus surinamensis*

Following the augmentation of the image dataset, shape features were extracted from the images. Subsequently, a variety of machine learning algorithms, including Support Vector Machines (SVM), k-nearest Neighbors (KNN), and Naïve Bayes (NB), were employed to classify the different insect classes. A Convolutional Neural Network (CNN)-based insect classification approach was also employed to establish a benchmark performance. This experiment aimed to assess and compare the classification performance of these machine learning algorithms. The insects were categorized using SVM, KNN, NB, and CNN models.

RESULTS AND DISCUSSION

Investigation into the performance of insect classification was conducted across seven distinct insect species, namely *C. chinensis, C. maculatus, O. surinamensis, R. dominica, S. granarius, S. oryzae,* and *T. castaneum.* This study employed machine learning algorithms including support vector machines (SVM), k-nearest neighbors (KNN), and naïve bayes (NB), while also leveraging convolutional neural networks (CNN) to establish a performance benchmark against traditional machine learning techniques. The dataset was enriched with 789 insect images, which were augmented to facilitate robust model training, testing and validation for accurate classification outcomes. The SVM classifier was strategically selected due to its aptness for linearly separable data in feature space, incorporating a linear kernel. This classifier effectively segregated data into distinct classes by delineating hyperplanes. Its computational efficiency and proficiency in handling more superficial separation boundaries were vital. Employing a KNN classifier with a parameter setting of two neighbours indicated that, for each test instance, labels from its two nearest training data neighbours were considered. The choice of neighbour count influenced the balance between model bias and variance. The euclidean distance metric was the default measure for evaluating distances between points in multidimensional space.

The naïve bayes classifier, adapted to gaussian distribution, demonstrated swift classification results, albeit with lower accuracy. The deficiency in accuracy could be attributed to the assumption of feature independence and uniform feature weighting, which might not hold true in the intricate dataset. The CNN model was trained with specific hyperparameters, including a batch size of 16, 25 epochs, and a learning rate of 0.0001. It harnessed its convolutional and max-pooling layers to automatically extract intricate, high-level features from the insect dataset. This led to a remarkable accuracy of 98%, primarily attributed to CNN's capacity to grasp hierarchical features and intricate patterns, underlining its superiority in image classification tasks. Figure 3, representing the trend of accuracy and loss over epoch exhibits the training dynamics and performance of CNN model. The trend of accuracy over epochs ensures that the model is learning from the training dataset with few plateaus which may be due to noise or outliers, whereas the loss curve indicates that the model is improving.

The classifier model performance was evaluated through a classification report detailing metrics like precision, recall, F1-score, and overall accuracy (Table 1) for each class. SVM achieved 81% accuracy as depicted in Table 1, across the seven insect classes, demonstrating its proficiency in classification. As a comparison with Xie et al. (2015), the authors applied moment-invariant features with SVM classifier and obtained an accuracy of 70.5% on 24 different insect classes of xie dataset. Further, Cheng et al., 2017 used ten classes of insects from Xie dataset and obtained 25.3%

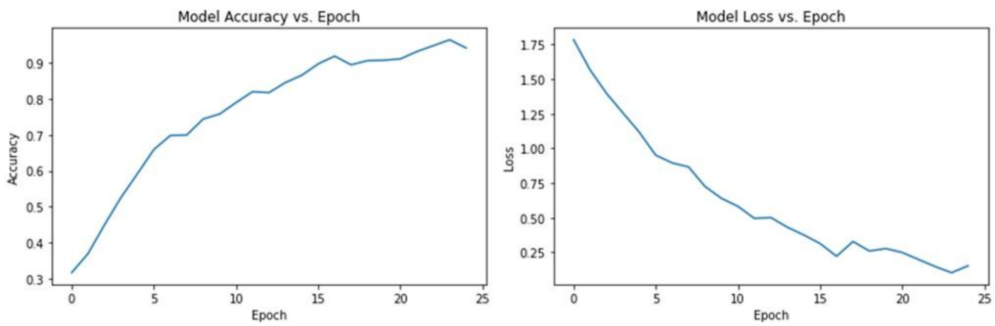

Fig. 3. CNN Model learning curve for accuracy and loss

Table 1: Classification results of insect dataset

Classifier	Classification results			
	Precision (%)	Recall (%)	F1 Scores (%)	Accuracy (%)
SVM	81	81	81	81
NB	80	76	76	33
KNN	80	76	76	76
CNN	98	98	98	98

classification accuracy by applying SVM classifier. Conversely, Naïve Bayes yielded lower accuracy 33% due to feature independence assumption violations and uniform feature weighting. KNN outperformed Naïve Bayes with 76% accuracy, effectively leveraging neighbours' labels to distinguish between insect species. With an accuracy of 81%, SVM showcased improved performance by aptly learning decision boundaries and achieving accurate predictions for most instances. However, CNN stood as the pinnacle performer, boasting a 98% accuracy. This outcome underscored its prowess in capturing intricate patterns, reaffirming its efficacy in classifying various insect species.

The computational time for classifier processing differed based on image type, size, algorithms, and processing units. Notably, SVM (3 min 18.33 s), KNN (3 min 58.36s), and NB (2 min 43.47s) exhibited quicker processing times. In contrast, CNN demanded increased processing time of about 28 min 33.11s due to its heightened accuracy, which justified its extended processing duration for accurate insect recognition and classification. The present investigation outlined the superior accuracy of the CNN model in insect classification, corroborated by its ability to extract high-level features effectively. This information could guide further research into optimizing insect recognition and classification systems. The convolutional neural network (CNN), emerging as the standout performer, owed its supremacy to its innate capacity for assimilating hierarchical features from the dataset. This capability enabled the extraction of pivotal attributes essential for delineating nuances between various insect species.

ACKNOWLEDGEMENTS

The authors acknowledge the Department of Zoology and Centre for Machine Learning and Intelligence, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu for providing the infrastructure to carry out the research work.

FINANCIAL SUPPORT

The authors acknowledge the financial support provided by the DST-CURIE-AI, Centre for Machine Learning and Intelligence, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore within the Phase II Project No.39 framework.

AUTHOR CONTRIBUTION STATEMENT

MSS carried out the experiment, acquisition of images, analysed the data and wrote the manuscript. GM supervised the experiments and critically revised the manuscript. NV and GS substantively revised it. All authors read and approved the final manuscript.

CONFLICT OF INTEREST

No conflict of interest.

REFERENCES

Ahmad R, Hassan S, Ahmad S, Nighat S, Devi Y K, Javeed K, Hussain B. 2021. Stored grain pests and current advances for their management. Postharvest Technology-Recent Advances, New Perspectives and Applications. pp. 45-82.

Cheng X, Zhang Y H, Wu Y Z, Yue Y 2017. Agricultural pests tracking and identification in video surveillance based on deep learning. Intelligent Computing Methodologies: 13th International Conference, Liverpool, UK. 2017. pp. 58-70.

Daglish G J, Nayak M K, Arthur F H, Athanassiou C G. 2018. Insect pest management in stored grain. Recent advances in stored product protection. pp. 45-63.

Deshwal R, Vaibhav V, Kumar N, Kumar A, Singh R. 2020. Stored grain insect pests and their management: An overview. Journal of Entomology and Zoology Studies 8(5): 969-974.

Devi Priya R, Anitha N, Devisurya V, Vidhyaa V P, Shobiya K, Suguna C. 2022. Insect and pest detection in stored grains: analysis of environmental factors and comparison of deep learning methods, WSEAS Transactions on Environment and Development, 18: pp. 759-768.

Hagstrum D W, Finn P W. 2012. Insect pest management for raw commodities during storage, fundamentals of stored-product entomology. Kansas State University. pp. 213-218.

Halstead D G. 1993. Keys for the identification of beetles associated with stored products. Laemophloeidae, Passandridae and Silvanidae. Journal of Stored Products Research 29(2): 99-197.

Howe R E, Currie J E. 1964. Some laboratory observations on the rates of development, mortality and oviposition of several species of Bruchidae breeding in stored pulses. Bulletin of Entomological Research 55(3): 437-477.

Kasinathan T, Singaraju D, Uyyala S R. 2021. Insect classification and detection in field crops using modern machine learning techniques. Information Processing in Agriculture 8(3): 446-457.

Krizhevsky A, Sutskever I, Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM 60(6): 84-90.

Kumari S, Yadav S S, Rolania K, Dhanda, S. 2022. Effect of seed Protectants against pulse beetle *Callosobruchus chinensis* infesting mungbean. Indian Journal of Entomology 84(1): 176–177

Mendoza Q A, Pordesimo L, Neilsen M, Armstrong P, Campbell J, Mendoza P T. 2023. Application of Machine Learning for Insect Monitoring in Grain Facilities. AI 4(1): 348-360.

Mohan S, Rajesh A. 2016. Tools for stored product insects management and technology transfer. Indian Journal of Entomology 78(special): 59-63.

Saikumar N, Emmanuel N, Sri Phani Krishna K, Chinnabbai C, Uma Krishna K. 2023. Artificial Intelligence for Classification and Detection of Major Insect Pests of Brinjal. Indian Journal of Entomology 85(3): 563-566.

Shankar U, Abrol D P. 2012. Integrated pest management in stored grains. Integrated pest management: principles and practice. pp. 386-407.

Shi Z, Dang H, Liu Z, Zhou X. 2020. Detection and identification of stored-grain insects using deep learning: A more effective neural network. IEEE Access 8: 163703-163714.

Singh S, Saini M K. 2023. Broad Spectrum Activity of Essential Oils in Managing Stored Grain Pests. Indian Journal of Entomology 85(3): 854-864.

Southgate B J, Howe R W, Brett G A. 1957. The specific status of Callosobruchus maculatus (F.) and *Callosobruchus analis* (F.). Bulletin of Entomological Research, 48(1): 79-89.

Quan W, Wang J, Lei L, Gao M. 2018. Target recognition method based on multi-class SVM and evidence theory. In Advances in Internetworking, Data & Web Technologies: The 5th International Conference on Emerging Internetworking, Data & Web Technologies, 2017. pp. 262-272.

Wilberforce S, Kalita S. 2023. Efficacy of Piper nigrum and Cuminum cyminum Seed Powders against *Callosobruchus chinensis* L. Indian Journal of Entomology 85(4): 973-976.

Xie C, Zhang J, Li R, Li J, Hong P, Xia J, Chen P. 2015. Automatic classification for field crop insects via multiple-task sparse representation and multiple-kernel learning. Computers and Electronics in Agriculture 119: 123-132.

(Manuscript Received: February, 2024; Revised: May, 2024;
Accepted: July, 2024; Online Published: August, 2024)
Online First in www.entosocindia.org and indianentomology.org Ref. No. e24973