## Insect Identification and Related Recommendations Using No-code ML-based ORANGE Data Mining and Literature Survey.

**BY**

**Harshilkumar Ponkiya (202319002)**

**Sanskar Shete (202319004)**

**Satyam Kumbhar (202319005)**

**Hemanth B S (202319009)**

**Abhay Kumar (202319012)**

MASTER OF SCIENCE IN AGRICULTURE ANALYTICS

**Submitted To:**

**Mr.Kapil Oberoi**

Scientist - Geoinformatics Department

IIRS-ISRO, Dehradun - 248001 India.



INTRODUCTION

With the continuous advancement of agricultural reform, modern technology is closely related to the development of agriculture. The sustainable development of modern agriculture is no longer limited to the use of natural resources but also includes the understanding and control of information resources. In recent years, the continuous deterioration of the ecological environment makes the ecological structure more fragile; crop diseases and insect pests often have large-scale outbreaks. Frequent outbreaks of crop diseases and insect pests directly affect the quantity and quality of agricultural products, resulting in economic losses. Therefore, it is necessary to study the control of crop diseases and insect pests to avoid unnecessary losses.

The first task of crop diseases and insect pests control is to quickly and accurately identify crop diseases and insect pests, so as to carry out risk assessment and control treatment in the future. Among them, statistical analysis and prediction of the same type of diseases and insect pests through a large amount of case data are also very important. In the past, pest control was usually carried out by manual statistics and analysis, and the relevant technical personnel or agricultural experts relied on experience to determine the type of pest through tedious and repetitive inspection, measurement, and statistical calculation. However, due to the differences in artificial experience and technology, the identification of diseases and pests is not accurate, and there will be some deviations and omissions in the way of data processing, so modern information technology is urgently needed to provide support for it.

In recent years, advanced models in machine learning were successfully achieved the best performance in pest classification and detection. Among these works, the various models were trained by using extracted features from the insects and different categories of insect images were classified. It is very difficult to classify and detect insects with similar feature types and different positions in the natural environment. Wang [11] and Xie [12] used ANN and SVM model for the classification of insects in the crops.

The classification accuracy compared between various machine learning techniques include ANN, SVM, KNN, Naive Bayes, and the CNN model. The results proved that the CNN model provides the highest classification accuracy of 91.5% and 90% for 9 and 24 classes of insects

Tools like Orange and KNIME provide functionality that enables users to start creating applications by drag-and-drop (visual programming). In contrast to coding approaches, no-code/low-code technologies offer an opportunity for users to concentrate on the analysis and much less on the lines of code.

Orange has a more intuitive and user-friendly visual interface, making it easier for non-technical users, such as researchers and students, to quickly grasp and use for data analysis and machine learning tasks.

KNIME also offers a visual interface, some users may find it slightly more complex than Orange, especially beginners who are not familiar with data mining concepts. The extensive node library and flexibility can sometimes lead to a steeper learning curve.

**Objectives :**

1. **Identification of Insects using AI-ML Techniques.**
2. **Recommending the Solution based on the identification of insects.**
3. **Comparison between Modern Data Mining Techniques and Traditional Coding in Python Programming.**

* **About Orange :**

Orange is an open-source data visualization, machine learning and data mining toolkit. It features a visual programming front-end for exploratory qualitative data analysis and interactive data visualization.

Orange is a component-based visual programming software package for data visualization, machine learning, data mining, and data analysis.

Orange components are called widgets. They range from simple data visualization, subset selection, and prepossessing to empirical evaluation of learning algorithms and predictive modeling.

Visual programming is implemented through an interface in which workflows are created by linking predefined or user-designed widgets, while advanced users can use Orange as a Python library for data manipulation and widget alteration.

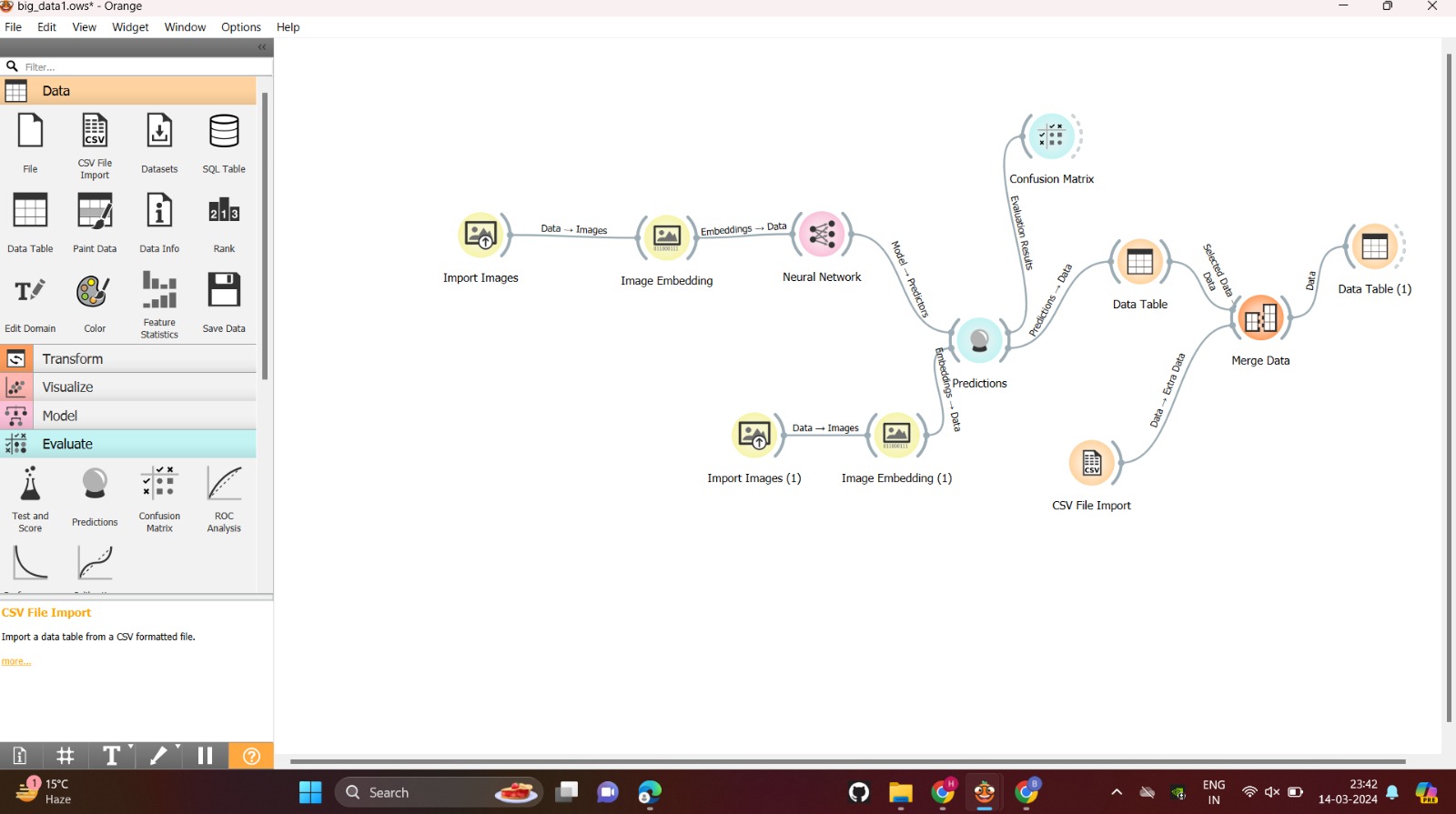
Dataset and Method :

The Dangerous Farm Insects Image Dataset is a curated collection of images featuring 15 different types of insects commonly found in agricultural settings. This dataset provides valuable visual resources for studying, identifying, and understanding the characteristics of these potentially harmful insects. Each insect is represented by multiple high-quality images, showcasing their distinct features, colors, and patterns.

Dataset Link: <https://www.kaggle.com/datasets/tarundalal/dangerous-insects-dataset>

Workflow :

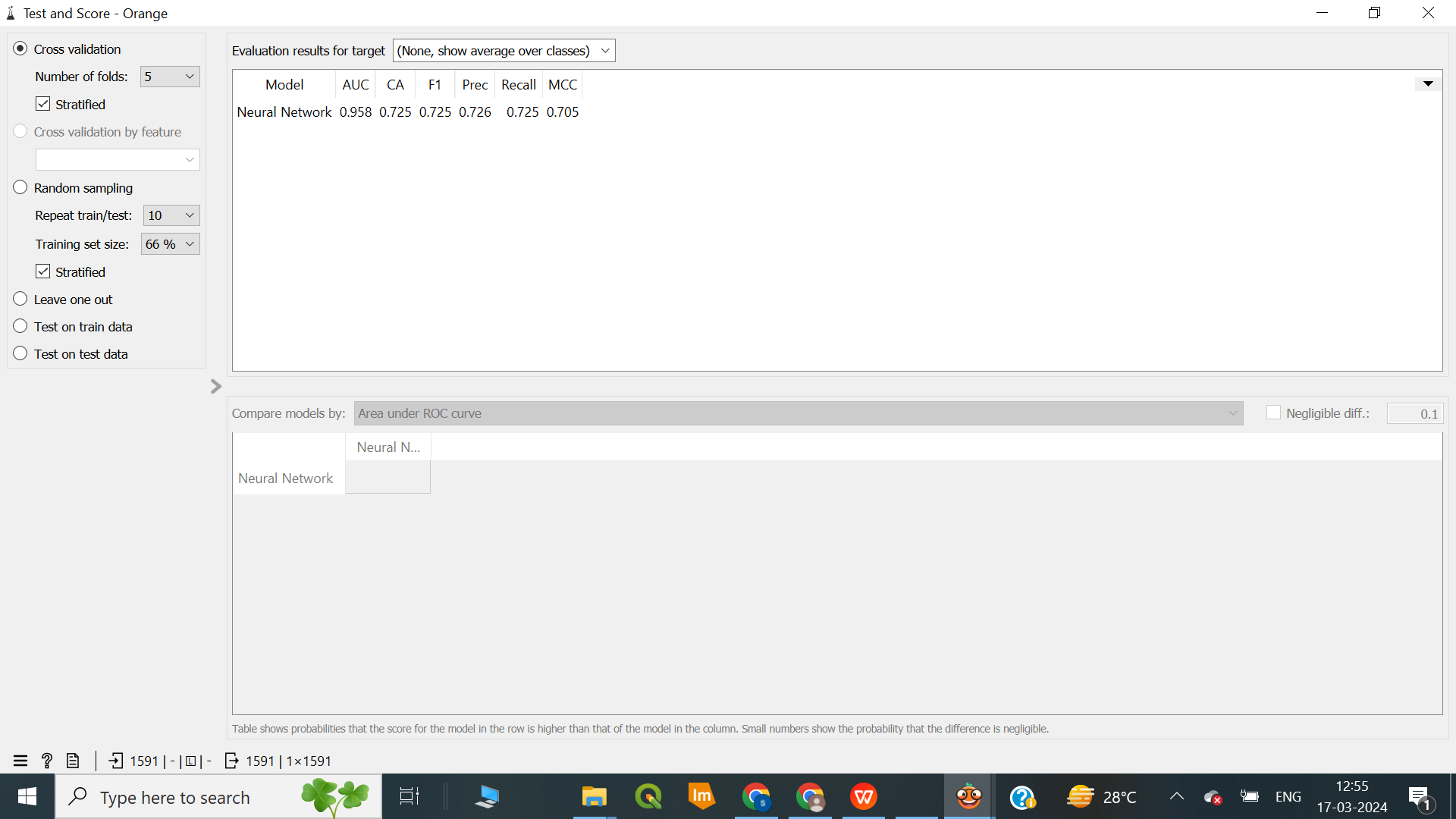
The process of identifying crop diseases and insect pests based on convolutional neural networks mainly includes the collection and preprocessing of image data sets, the construction and training of convolutional neural networks, and the verification of the accuracy of neural network. The system flow is shown in Figure.

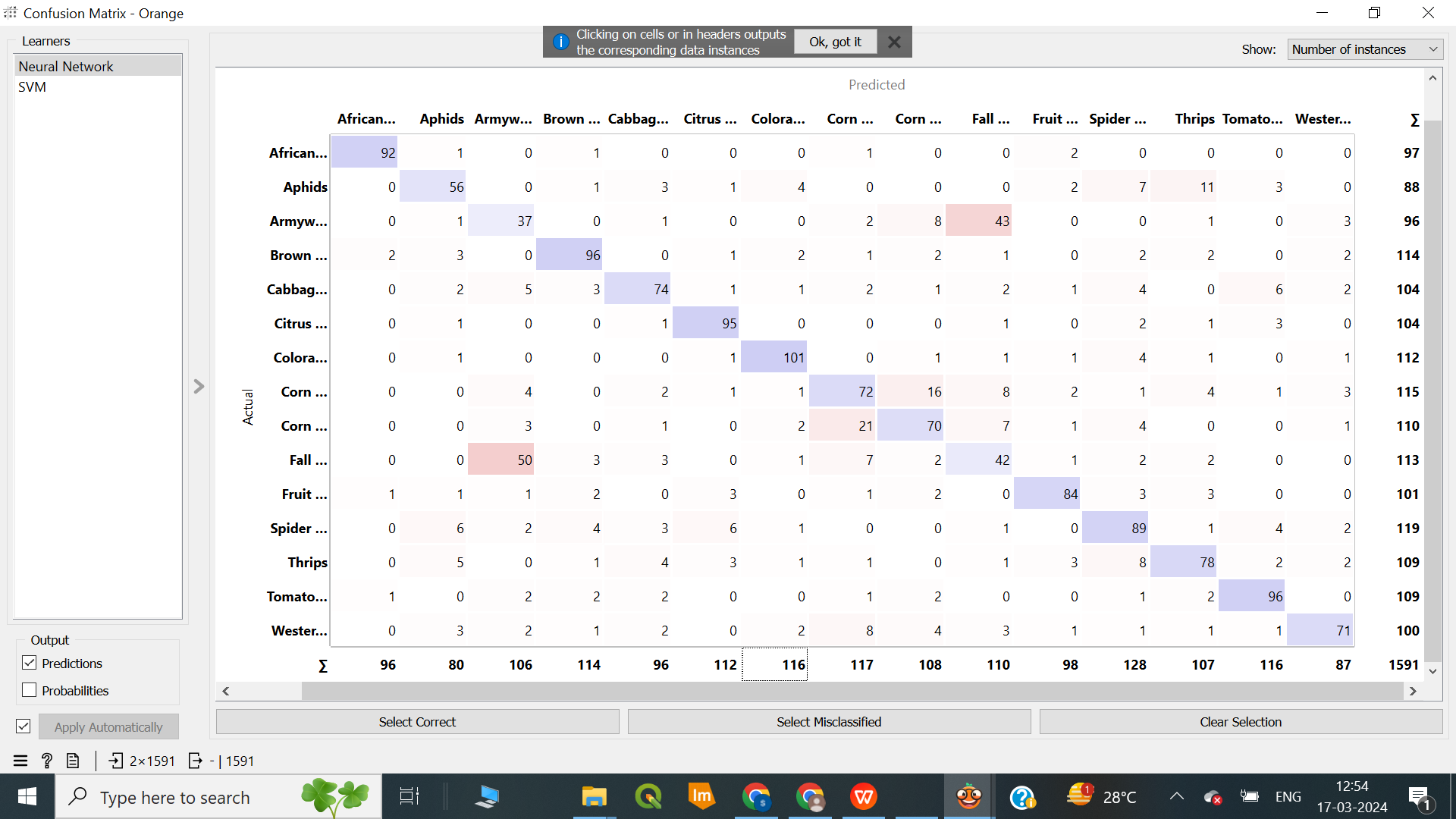


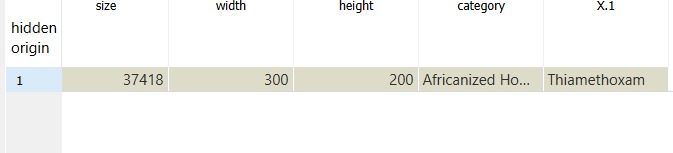
The description of each part of the system is as follows:

1. **Data set collection and preprocessing**: preprocessing mainly includes four steps: data set optimization, image transformation, image standardization, and data enhancement.
2. **Constructing and training a convolutional neural network:** this mainly includes four parts: defining the structure of the convolutional neural network, defining the loss function, iterative training, and accuracy assessment. Defining the network structure means defining the algorithm formula of the forward calculation of the neural network; defining the loss function will affect the training speed of the convolutional neural network; the iterative operation uses the backpropagation principle to continuously calculate and update the weights of the convolutional neural network until training ends; accuracy assessment is to evaluate and improve the identification accuracy rate of the convolutional neural network.
3. **Verification of the model accuracy of neural network:** a number of comparative experiments are designed to verify the identification accuracy of various convolutional networks.

**Results :**







**Conclusions :**

* The Orange Data Mining Model outperforms the CNN model in predictive performance. It has higher AUC, F1 score, precision, recall, and MCC, indicating its superior ability to classify instances correctly.
* The model was able to identify the insect from the new image, and as a result, it provided successful recommendations.
* When it comes to achieving the best predictive performance, the Orange Data Mining Model is a better choice than the CNN model using Python.
* To train the model in orange, it takes less than a minute, whereas in Python, it takes nearly 8 minutes.

References :

1. <https://orangedatamining.com/blog/explain-scoring-sheet/>
2. <https://www.hindawi.com/journals/sp/2022/9179998/>
3. <https://www.sciencedirect.com/science/article/pii/S2214317320302067>
4. <https://ijisae.org/index.php/IJISAE/article/view/4197>

*Thank You!!*