Machine Learning Technology for Disease Classification in Agriculture Using Convolution Neural Networks (CNN)

Trushar B Patel [1] Dr.Hitesh Raval[2]

Research Scholar, Department of Computer Science, Sankalchand Patel University, Visnagar[1]
Associate Professor, Department of Computer Science, Sankalchand Patel University,Visnagar [2]

Abstract

Precision agriculture (PA) represents an advanced farming management concept that uses modern technology to enhance crop yield and efficiency. Machine learning (ML), a subset of artificial intelligence (AI), is pivotal in PA for analyzing data and making informed decisions. This paper explores the integration of ML in PA, focusing on techniques for tracking, evaluating, and improving agricultural precision. Therefore, the primary objective of this work is to categories the wheat leaf image into one of three groups—healthy, septoria, or stripe rust—using the power of DL algorithms. In this context, we propose an SVM DL-based feature model, a potential game-changer in the field of precision agriculture, where the ML component is responsible for classifying the wheat leaf images into the three disease categories based on the features extracted from each image using various DL models.

Keywords: ML,DL,AI,CNN,SVM,AP

Introduction

Machine learning (ML), a subset of artificial intelligence (AI), has emerged as a pivotal tool in the PA domain, revolutionizing the precision and effectiveness of agricultural practices. ML algorithms can process and analyze vast amounts of data from diverse sources, providing actionable insights that help farmers make informed decisions. The integration of ML in PA not only allows for real-time monitoring, predictive analysis, and automated interventions, but also holds the potential to revolutionize the precision and effectiveness of agricultural practices.

The role of machine learning in precision agriculture is significant.

Data-Driven Decision Making

Machine learning enhances decision-making processes by transforming raw agricultural data into meaningful insights. For example, ML models can analyse historical weather data, soil composition, and crop performance to predict future conditions and recommend optimal planting schedules and crop varieties. This predictive capability helps farmers mitigate risks associated with unpredictable environmental factors.

Monitoring and evaluation

ML techniques allow for continuous monitoring and evaluation of agricultural fields. Sensors and drones collect real-time data on soil moisture, nutrient levels, and plant health, which ML algorithms then analyze to detect anomalies or trends. Early detection of nutrient deficiencies, water stress, or pest infestations allows for timely interventions, reducing crop losses and improving overall farm management.

Resource Optimization

One of the primary goals of precision agriculture is to optimize resource use. ML algorithms can analyze data from various sources to determine the precise amount of water, fertilizer, and pesticides needed for different field parts. This targeted approach minimizes waste, reduces costs, and lessens the environmental impact of farming practices.

The paper's scope and structure

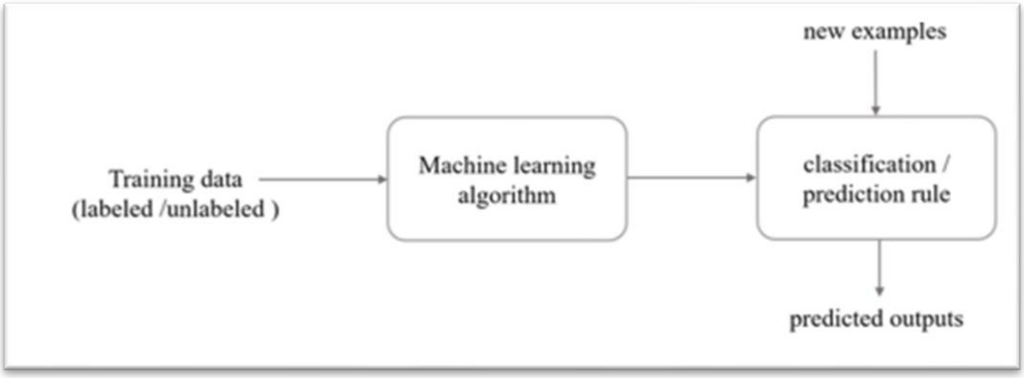
This paper explores the various applications of ML in precision agriculture, focusing on techniques for tracking, evaluating, and improving agricultural precision. It also talks about the theories underlying each leaf disease.

- ✓ Offering an SVM DL-based feature set based on various TL models to compare which is better at distinguishing between multiple illnesses and healthy leaves.
- ✓ Comparing the outcomes of segmented image feature extraction with direct image feature extraction.
- ✓ Comparing the outcomes with other models suggested in the literature using various datasets.
- ✓ Among the models in the literature, to the best of our knowledge, our model performs the best with the dataset utilized.

Machine Learning in Precision Agriculture

Overview of Machine Learning

Machine learning (ML) is a pivotal branch of artificial intelligence that focuses on developing algorithms capable of learning from and making predictions based on data. Unlike traditional programming, where explicit instructions are given to a machine, ML allows computers to identify patterns and insights from large datasets, adapting and improving over time with more data and feedback.



ML is increasingly being leveraged in agriculture to process and analyze extensive datasets sourced from various inputs, such as sensors, satellite imagery, and historical weather records. By extracting meaningful patterns from this data, ML algorithms can help farmers make informed decisions, optimize resource use, and improve crop yields.
Essential ML techniques used in precision agriculture include:

- **Supervised Learning**: Training algorithms on labelled datasets to make predictions or classify data (e.g., crop disease detection).
- **Unsupervised Learning**: Identifying hidden patterns in unlabeled data (e.g., clustering different soil types).
- **Reinforcement Learning**: Algorithms learn optimal actions through trial and error (e.g., robotic weed control).
- **Deep Learning**: Using neural networks with many layers to handle complex tasks (e.g., image recognition for pest detection).

Data Sources and Acquisition

The effectiveness of ML in precision agriculture heavily relies on the quality, quantity, and diversity of data available for training and analysis. Typical data sources in PA include:

- **Remote Sensing**: Satellite and drone imagery provides comprehensive views of agricultural fields, capturing data on crop health, soil conditions, and overall field management. These images can be analyzed to monitor crop growth, detect stress or disease, and assess yield potential.
- **Soil Sensors**: These sensors measure critical soil parameters such as moisture, temperature, pH, and nutrient levels. The data collected helps understand soil health, manage irrigation systems, and optimize fertilizer application.
- **Weather Stations**: Localized weather data from on-site weather stations include information on temperature, humidity, precipitation, and wind speed. This data is crucial for predicting weather patterns, assessing risk factors, and planning agricultural activities.
- **Yield Monitors**: Installed on harvesting equipment, yield monitors track crop yields in real-time, providing detailed spatial data on productivity across different field areas. This information is essential for evaluating the effectiveness of farming practices and planning future planting strategies.

Integrating these diverse data sources enables a holistic approach to farm management, allowing for more precise and timely interventions. For instance, real-time data from soil sensors and weather forecasts can help optimize irrigation schedules, while remote sensing imagery can guide targeted pest control measures. By harnessing the power of ML, farmers can turn this data into actionable insights, leading to more efficient and sustainable agricultural practices.

Review of Literature

Sezer et al. (2018), The massive amount of data gathered by industrial systems includes details about the actions, occurrences, and alarms throughout a manufacturing line in the industry. Additionally, when processed and analyzed, these data can yield important insights into the dynamics of the manufacturing process. It is possible to find interpretive results for strategic decision-making using analytical data-based methods. This has benefits, including decreased maintenance costs, decreased machine fault rates, decreased repair stops, decreased spare part inventories, increased spare part life, increased production, improved operator safety, repaired verified, and higher overall profits.

Poojan Panchal et al. (2019) created a dataset consisting of four classes: a healthy leaf class and Bacterial Spot, Early Blight, and Late Blight as leaf illnesses. Next, the authors divided the healthy and affected sections using the HSV-Herpes Simplex Virus segmentation technique to improve feature extraction. Then, using the Grey Level Co-occurrence Matrix (GLCM) features that were taken out of the photos, they employed a random forest classifier to detect plant leaf disease, and they were able to reach a 98% accuracy rate.

According to Shah Hosseini et al. (2019), agronomists and farmers may make better decisions if they can forecast crop production outcomes such as grain yields and nitrogen (N) losses before the growing season. The substantial amount of data required and the protracted processing times associated with simulation crop models constrain their application despite their potential usefulness in scenario planning. Thus, we can use less computationally intensive methods to make more extensive forecasts. XGBoost was the most accurate machine learning model for yield prediction, with an RRMSE of 13.5%. Conversely, Random Forests, with an RRMSE of 54.5%, was the most accurate machine learning model for forecasting N loss at the time of planting. ML meta-models adequately reconstructed anticipated maise yields using the data available at planting; however, they did not consider N loss. Their sensitivity to the quantity of the training dataset varied. As the size of the training dataset increased from 0.5 to 1.8 million data points, the yield prediction error for all ML models improved by 10%–40%;

Jiang et al. (2021), a multi-task model uses a shared layer to train a model that alternatively detects the kind of leaf disease for wheat and rice crops. They made use of two datasets that were made accessible to the public for wheat and rice, which included the following classes: wheat rust, wheat powdery mildew, brown spot, bacterial leaf blight, and smut. Subsequently, they performed pre-processing operations such as resizing and rescaling and then trained the customized VGG16 TL model. For rice leaf diseases and wheat leaf diseases, they obtained 97.22% and 98.75%, respectively.

N. Nandhini et al. (2018) proposed a plant leave disease classification technique. They performed multiple pre-processing stages on the images, such as image segmentation, to improve performance. Next, they used GLCM to extract features from the segmented images, and then they used SVM, KNN, and DTs as classifiers for the extracted features. SVM had the best performance of 98.51%.

Dataset

The Kaggle dataset utilized in this work is accessible to the general public. It has three classes: Septoria, Stripe Rust, and Healthy. Fig. 1 displays an example for each class. The collection includes 208 photos of Stripe Rust leaf disease, 97 pictures of Septoria leaf disease, and 102 photos of healthy crops. This section details the pre-processing processes required because the dataset is tiny and imbalanced.

Healthy wheat leaf Septoria leaf disease Stripe Rush leaf disease

The first kind is yellow or stripe rust, a fungal disease that can damage wheat crops at any point in their life cycle. In colder climates, it can primarily impact the weight, size, and quantity of grains per spike of the wheat crop. Under favorable weather circumstances, the second form of wheat rust fungal disease, also known as brown or leaf rust, affects the glumes and sheaths of the wheat and reduces the number of grains per plant. The last kind is "black" or "stem rust," the most harmful fungal disease for cereal crops worldwide. It causes damage to the stem, sheaths, glumes, and spikes, lowers kernels per spike, and shrinks grain size in warm weather. Fig. 2 depicts the three different forms of rust. Since stripe rust is the most harmful kind, it is the only kind of rust that is the primary subject of this essay. The other kind of leaf disease that this paper focuses on is septoria, which ranks second in importance among wheat diseases after rust and poses a severe risk to the world's food supply. It results in a 35–50% reduction in wheat yields. It is primarily active in cold climates.

Black Rust **Yellow Rust** **Brown Rust**

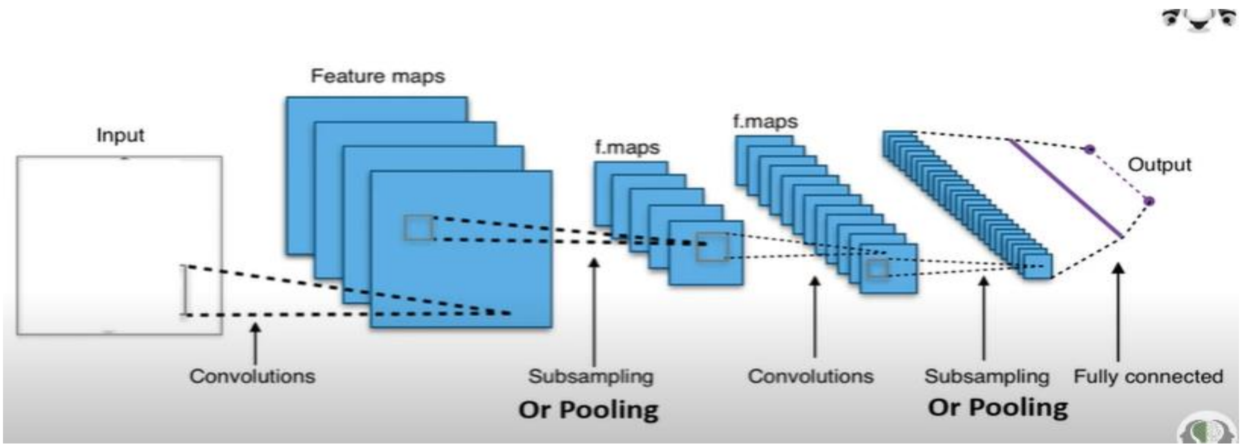
DL models that have been pre-trained on the 1000 classes of various categories that make up the ImageNet dataset are called TL models. The fundamental layers of each model are dense or wholly connected, max-pooling, and 2D convolutional. Additional layers, such as Batch Normalization and Dropout, can be added for optimization. Since the purpose is to obtain the output features before the classification phase rather than classify them, any model's final fully connected layers are eliminated when using any of the TL models as a feature extractor.

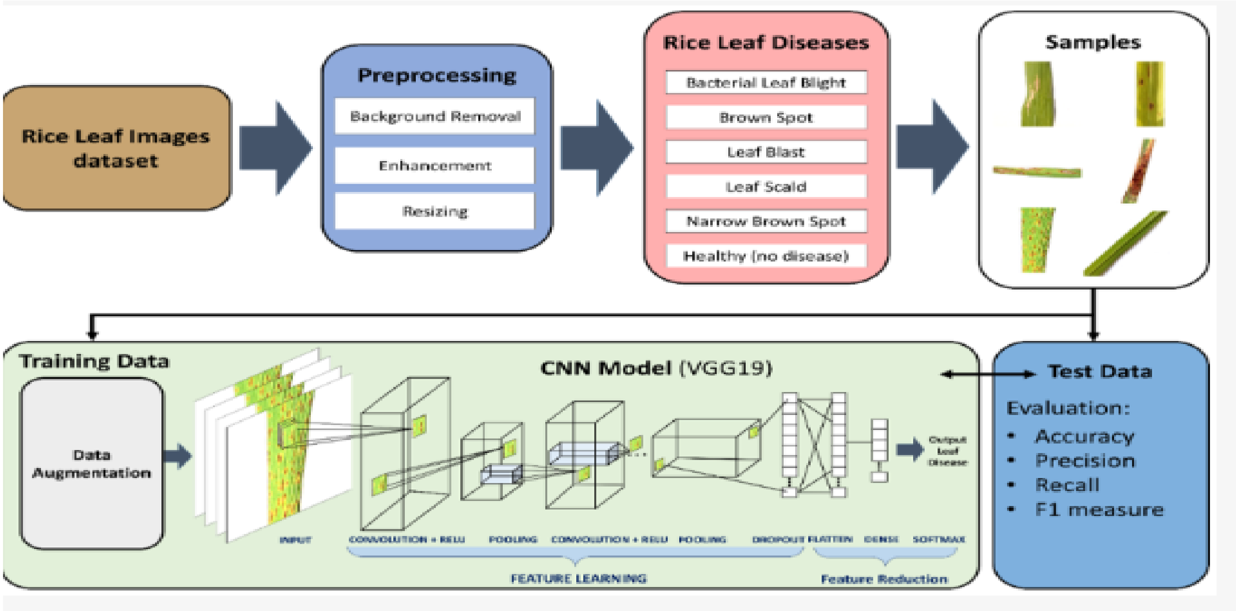
The VGG19 TL and VGG16 TL models are nearly identical. Convolutional Neural Network (CNN) architectures, which rank among the finest vision models, are what both of them represent. The input and output shapes are the same for each. There are two differences: the total number of parameters, with VGG16 having 16 convolutional layers and VGG19 having 19 convolutional layers.

Machine Learning CNN Algorithm

A Convolutional Neural Network (CNN) is a type of neural Network architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data.

A Convolutional Neural Network (CNN), also known as ConvNet, is a specialized type of deep learning algorithm mainly designed for tasks that necessitate object recognition, including image classification, detection, and segmentation.



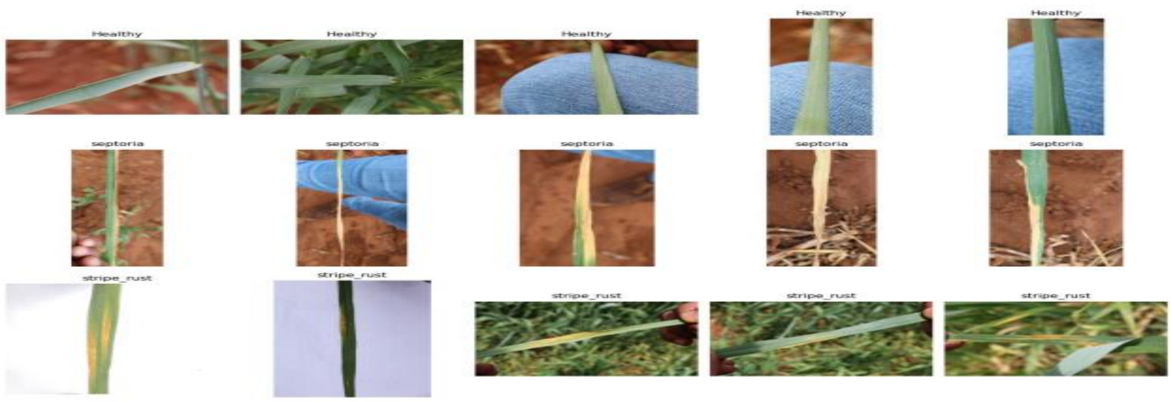


Loading the dataset and creating the image data generators.

First Load the library of ‘tensorflow’ in Python

```
# Path to the dataset
data_dir = "C:\\Users\\Bipin Bhong\\Desktop\\Python\\wheat_leaf"

# Plot some sample images from each class
classes = ['Healthy', 'septoria', 'stripe_rust']
fig, axes = plt.subplots(3, 5, figsize=(15, 10))
axes = axes.flatten()
```

Exploratory Data Analysis (EDA): Displaying some sample images from each class. Create three class of Image. In this Process collect Photos of different category in like 'Healthy', 'septoria', 'stripe rust‘.



- In CNN Model First We Need to Convolution layer of CNN.
  - Convert single image into multiple image in the form of 16,32,64 Matrix
  - In This Layer Algorithm First Filter the Images in the form of Pixel. So that Images Are Filter and Filter image generated in the Form of length and Width of images.

Convolution layer is applied in the form of 16,32,64,128

- IN Second Layer of CNN is **Pooling Layer**.
- When we got number of image from convolution image then we apply pooling.
- Means when we got multiple image so we need to compress image.
- This layer is periodically inserted in the converts and its main function is to reduce the size of volume which makes the computation fast reduces memory and also prevents overfitting.

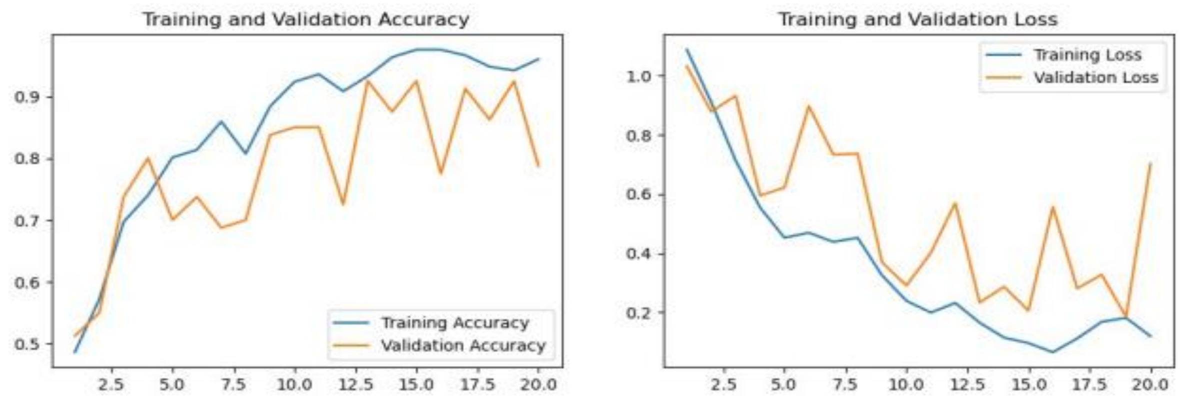
```
train_generator = train_datagen.flow_from_directory(
    data_dir,
    target_size=(128, 128),
    batch_size=32,
    class_mode='categorical',
    subset='training'
)

validation_generator = train_datagen.flow_from_directory(
    data_dir,
    target_size=(128, 128),
    batch_size=32,
    class_mode='categorical',
    subset='validation'
)

Found 327 images belonging to 3 classes.
Found 80 images belonging to 3 classes.
```

After the Pooling layer the model will Print summary in the form Layer,Conv2d,Max_Pooling.

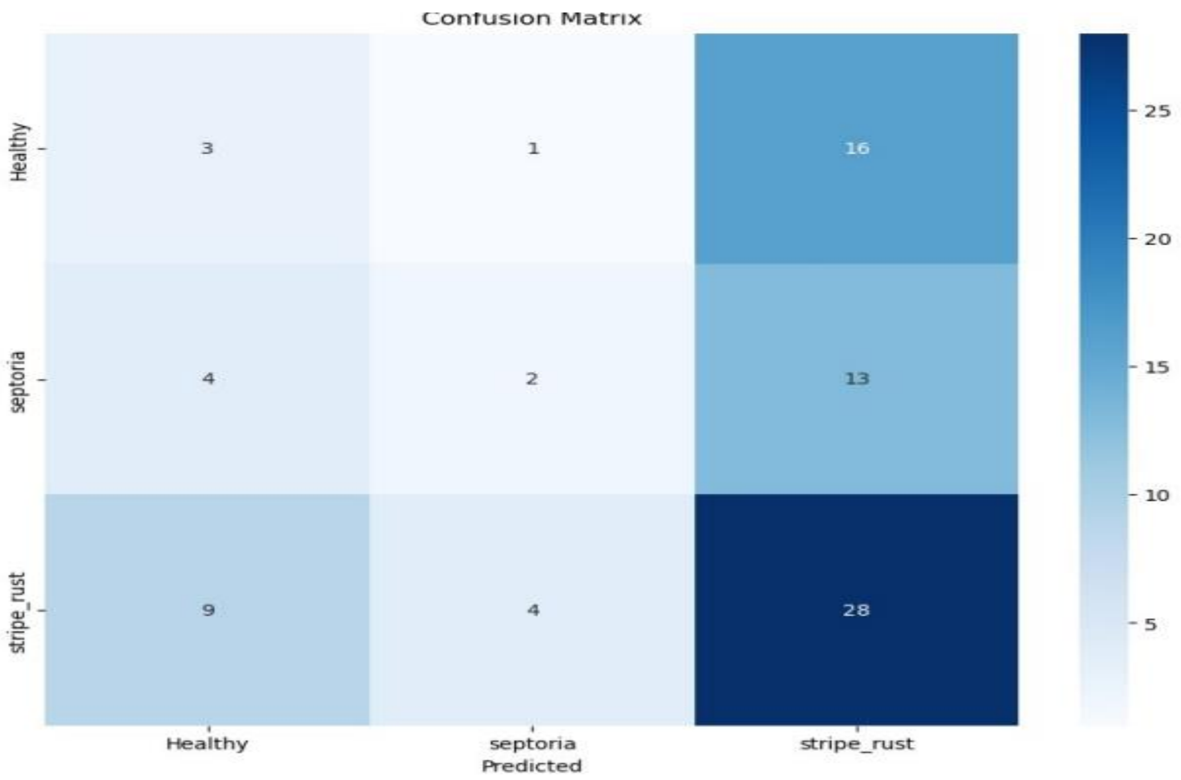
```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_1 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_1 (MaxPooling2D)	(None, 30, 30, 64)	0

Learning Rate Annealing (LRA) & Model Training:



Use Confusion Matrix with CNN:



```
Classification Report
              precision    recall  f1-score   support

     Healthy       0.19      0.15      0.17        20
    septoria       0.29      0.11      0.15        19
 stripe_rust       0.49      0.68      0.57        41

    accuracy                           0.41        80
   macro avg       0.32      0.31      0.30        80
weighted avg       0.37      0.41      0.37        80

# Confusion Matrix
print('Confusion Matrix')
cm = confusion_matrix(validation_generator.classes, y_pred)
plt.figure(figsize=(10, 8))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues",
xticklabels=classes, yticklabels=classes)
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()

Confusion Matrix
```

Finding

the confusion matrix is an essential instrument for evaluating the effectiveness of classification models. Insights into a model's accuracy, precision, recall, and general efficacy in classifying instances are provided by the thorough analysis of true positive, true negative, false positive, and false negative predictions it offers. The article provided examples to illustrate each metric's computation and discussed its importance. It also demonstrated how confusion matrices can be implemented in Python for binary and multi-class classification scenarios. Practitioners can make well-informed decisions regarding model performance—particularly when dealing with imbalanced class distributions—by comprehending and applying these metrics. Still Confusion matrix are more effective for identifying crop data set.

This section reviews the findings of existing models in the literature and presents and contrasts the outcomes of the suggested model. Every experiment is run on the lab notebook's GPU. The dataset is split into three sets: fifteen per cent test, fifteen per cent validation, and seventy per cent train. Each set's features are extracted sequentially, with a batch size of one and a target image size of 224 by 224 pixels.

Three experiments are carried out for our proposed model. In each experiment, the features from the HSI-segmented images are extracted using a single TL model. Afterwards, the characteristics are sent to the SVM classifier. Feature extraction using one of the following models is part of the experiments: VGG16, VGG19, or InceptionResNetV2.

The characteristics derived from VGG19 are the most accurate, with a 98% accuracy rate, as can be seen. The VGG16 model's features are pretty decent, too. The VGG16's features could be improved with additional tuning or by further processing the input photos. Since the InceptionResNetV2 model is too sophisticated for this tiny dataset, its accuracy could be higher.

Conclusion

Precision agriculture (PA) heralds a transformative shift in farming operations, enhancing efficiency, productivity, and sustainability through machine learning (ML) integration. This paper has highlighted ML's significant role in enhancing precision agriculture, from tracking and evaluating field conditions to improving resource use and crop management practices.

ML algorithms enable data-driven decision-making by analyzing vast and diverse datasets, including satellite imagery, soil sensor readings, weather data, and yield monitors. These insights help farmers make informed decisions about planting schedules, irrigation, fertilization, and pest control. Specifically, ML techniques like supervised, unsupervised, reinforcement, and deep learning are pivotal in addressing various agricultural challenges, such as soil health monitoring, crop management, pest and weed detection, and yield prediction.

Recent advances in the theory of human development are crucial for the growing focus on food security. Government policies and research can help achieve more sustainable food systems. Food quality schemes are complex ideas based on the interaction of unique elements created by the European Union. A food product's overall quality is the result of multiple elements interacting. More than ever, scientists analyze food quality features and variations using sophisticated, multivariate technology. Utilizing suitable data analysis techniques is crucial for extracting valuable and definitive information from these extensive data sets throughout the execution of these methodologies.

This study presents several trial results on the suggested wheat leaf crop disease classification model. The input images provided to the feature extractors are HSI-Segmented images, and the model is built on deep features extracted from various TL models that are passed to the SVM classifier. Three classes from a publicly accessible wheat leaf disease dataset are used to train the model. With an accuracy of 98%, the VGG19 feature extractor surpasses other models in the literature, as evidenced by the findings

References

- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., & Linus, O. U. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), e00938.
- Aravind, K. R., Raja, P., & Devadhas, D. P. (2020). Applications of machine learning in agriculture: A survey. *Indian Journal of Science and Technology*, 13(12), 1324- 1331.
- Bechar, A., & Vigneault, C. (2016). Agricultural robots for field operations: Concepts and components. *Biosystems Engineering*, 149, 94-111.
- Bendre, M. R., Thool, R. C., & Thool, V. R. (2015). Big data in precision agriculture: Weather forecasting for future farming. *Proceedings of the IEEE International Conference on Big Data*, 2015, 1894-1901.
- Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147, 70-90.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674.
- Reyes, M. D., Chiou, M., & Khanna, M. (2020). Adoption of precision agriculture: An application of duration analysis. *Technology in Society*, 63, 101353.
- Sharma, A., Kundu, S., Sarkar, P., & Munshi, S. (2020). Machine learning applications

for precision agriculture: A comprehensive review. *IEEE Access*, 8,

- Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote Sensing*, 12(18), 3136.
- Sun, S., Sun, C., & Du, W. (2021). Internet of Things (IoT) based smart agriculture: Toward making the fields talk. *IEEE Access*, 9, 73617-73630.
- Wang, G., Liu, Z., Lin, H., & Zhao, H. (2021). A review of deep learning in multi- modal remote sensing data for precision agriculture. *Remote Sensing*, 13(9), 1756.
- Sezer, E., Genaidy, A., Lawley, M., Garcia, H., Gabbouj, M., & Bayraktar, M. (2018). A comprehensive review of industry 4.0 applications and data-driven analytics for industrial systems. *IEEE Access*, *6*, 36556-36569
- Sajja, G. S., Jha, S. S., Mhamdi, H., Naved, M., Ray, S., & Phasinam, K. (2021, September). An investigation on crop yield prediction using machine learning. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 916-921). IEEE
- Shahhosseini, M., Martinez-Feria, R. A., Hu, G., & Archontoulis, S. V. (2019). Maise yield and nitrate loss prediction with machine learning algorithms. Environmental Research Letters, 14(12), 124026.
- Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. *Science*, *327*(5967), 828-831. https://doi.org/10.1126/science.1183899.
- Panchal, P., Raman, V.C., Mantri, S.: Plant diseases detection and classification using machine learning models. In: 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), vol. 4, pp. 1–6 (2019). IEEE
- Jiang, Z., Dong, Z., Jiang, W., Yang, Y.: Recognition of rice and wheat leaf diseases based on multi-task deep transfer learning. Computers and Electronics in Agriculture 186, 106184 (2021)
- Aasha Nandhini, S., Hemalatha, R., Radha, S., Indumathi, K.: Web-enabled plant disease detection system for agricultural applications using wmsn. Wireless Personal Communications 102(2), 725–740 (2018)