JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Soybean Leaf Disease Detection Using Machine Learning

Pradnya Mandale, Vaishnavi Aher, Rushikesh Wable, Sahil Bhujbal, Pallavi Choudhari

Sandip Institute of Technology & Research Center

I.Abstract With the continuous integration of computer technology into agricultural production, it also reduces personnel costs while improving agricultural production efficiency and quality. Crop disease control is an important part of agricultural production, and the use of computer vision technology to quickly and accurately identify crop diseases is an important means of ensuring a good harvest of agricultural products and promoting agricultural modernization. In this research paper, we propose a comprehensive methodology for soybean leaf disease detection using the Gray-Level Co-occurrence Matrix (GLCM), absolute distance measure, and Support Vector Machines (SVM) algorithm. We highlight the economic impact of soybean leaf diseases and the importance of early detection. Our experiments demonstrate the effectiveness of our approach in accurately classifying soybean leaf diseases, achieving an accuracy of 92%. This research contributes to the field by providing a robust and efficient method for automated disease detection in soybean crops. Our proposed methodology outperforms existing approaches in soybean leaf disease detection, demonstrating its effectiveness and potential for practical implementation. This research contributes to the field by providing a reliable and efficient method for automated soybean leaf disease detection, enabling timely interventions and improved disease management practices in agriculture.

Keywords: Gray-Level Co-occurrence Matrix (GLCM), absolute distance measure, and Support Vector Machines (SVM) algorithm

II.Introduction

Soybean (Glycine max) is one of the most economically important crops worldwide, serving as a primary source of protein and oil. However, soybean production is significantly affected by various diseases that reduce crop yield and quality. Among these, soybean leaf diseases play a crucial role in determining the overall health and productivity of soybean plants. Early detection and accurate classification of these diseases are paramount for implementing timely and effective disease management strategies, mitigating crop losses, and ensuring sustainable soybean production. Manual detection of soybean leaf diseases is labor-intensive, time-consuming, and subject to human error and bias. Therefore, there is an increasing demand for automated and reliable disease detection methods to facilitate early intervention and precise disease management practices. In recent years, machine learning algorithms have shown great potential in automating disease detection tasks by leveraging the computational power and pattern recognition capabilities of computers. This research paper presents a comprehensive methodology for soybean leaf disease detection using the Gray-Level Co-occurrence Matrix (GLCM), absolute distance measure, and Support Vector Machines (SVM) algorithm. The GLCM is a powerful technique that captures the spatial relationships between pixel intensities in an image, providing valuable texture information. By extracting texture features from soybean leaf images using the GLCM, we aim to capture disease-specific patterns and characteristics that can aid in accurate disease classification. The absolute distance measure is employed to quantify the dissimilarity between GLCM features of test samples and disease classes. This measure facilitates the identification of disease patterns that differ from healthy soybean leaves or other disease types. Furthermore, the SVM algorithm is utilized as a robust classifier to accurately categorize soybean leaves into different disease classes based on the extracted GLCM features. The main objectives of this research are to develop an efficient and reliable methodology for soybean leaf disease detection and to evaluate its performance in accurately classifying different disease types. By leveraging the GLCM, absolute distance measure, and SVM algorithm, we aim to enhance the accuracy and efficiency of soybean leaf disease detection, ultimately aiding farmers and agronomists in implementing timely and targeted disease management strategies. In the following sections of this paper, we will delve into the related work in soybean leaf disease detection, discuss the methodology in detail, present the experimental results, and provide a comprehensive analysis of the findings. Additionally, we will discuss potential future research directions and the implications of our research for the field of automated disease detection in agriculture.

III. Methodology

Our methodology for soybean leaf disease detection using GLCM, absolute distance, and SVM algorithms consists of the following steps:

Step 1: Data Collection and Pre-processing:

A dataset comprising high-resolution images of soybean leaves affected by different diseases is collected. The dataset should encompass a wide range of disease types and severities to ensure the robustness and generalizability of the proposed methodology. These images are pre-processed to enhance the quality and extract relevant information. Pre-processing techniques may include noise removal, contrast enhancement, and illumination normalization to mitigate variations that can affect disease detection accuracy.

Step 2: Feature Extraction using GLCM:

The GLCM is calculated for each pre-processed soybean leaf image. The GLCM is a statistical representation of the spatial relationships between pixel intensities in an image. From the GLCM, various texture features can be extracted, which capture the unique characteristics of soybean leaf diseases. These features include contrast, energy, homogeneity, and correlation, among others. These GLCM-based features serve as discriminative indicators for distinguishing different disease types and healthy leaves.

Step 3: Absolute Distance Computation:

To quantify the dissimilarity between the GLCM features of the test samples and disease classes, the absolute distance measure is employed. The absolute distance between the feature vectors of the test samples and each disease class is computed. This measure calculates the absolute difference between the corresponding feature values of the test sample and the reference values of each disease class. By comparing the absolute distances, the test sample can be assigned to the disease class with the minimum distance, indicating the most probable disease category.

Step 4: SVM Training and Classification:

The SVM algorithm is employed as a powerful classifier to accurately classify soybean leaves into different disease classes based on the extracted GLCM features. SVM is a supervised learning algorithm that maps input data to a higher-dimensional space, where a hyperplane is constructed to separate the data points into distinct classes. During the training phase, the SVM model learns the optimal hyperplane by maximizing the margin between the classes while minimizing classification errors. The feature vectors extracted from the GLCM, along with their corresponding disease labels, are used to train the SVM classifier. Once trained, the SVM model can classify new, unseen soybean leaf samples into their respective disease classes based on their feature vectors.

Step 5: Performance Evaluation:

To evaluate the performance of our proposed methodology, rigorous experiments are conducted. The dataset is typically divided into training and testing sets, following a suitable partitioning ratio (e.g., 70:30 or 80:20). The trained SVM classifier is then evaluated on the testing set to measure its accuracy in correctly classifying soybean leaf diseases. Evaluation metrics such as accuracy, precision, recall, and F1 score can be computed to assess the overall performance of the methodology. These metrics provide insights into the model's ability to correctly identify disease classes and its performance in minimizing false positives and false negatives.

By following this methodology, we aim to achieve accurate and efficient soybean leaf disease detection. The combination of GLCM-based texture features, the absolute distance measure, and the SVM classifier contributes to the development of a robust and reliable automated system for disease classification. In the subsequent sections, we will present the experimental results and discuss the performance and implications of our proposed methodology.

IV. Data Augmentation

Data Augmentation is a technique for increasing training datasets without having to gather new images. Data augmentation alters the original images in some way. This is accomplished by using various processing techniques including as rotations, flips, zooming, and adding noise, among others. Large training datasets are significant in deep learning since they improve the training model's accuracy. It also aids in the avoidance of overfitting. The downsides of data augmentation include increased training time, transformation computation costs, and higher memory expenses. The dataset is divided into two parts where 80% is used for training and 20% is used for testing.

V. Related Work:

A significant body of research has been devoted to soybean leaf disease detection, employing various techniques and algorithms. This section provides an overview of the existing literature, highlighting different approaches and their strengths and limitations.

1. Image Processing Techniques:

Several studies have utilized image processing techniques for soybean leaf disease detection. These techniques involve preprocessing steps such as image segmentation, feature extraction, and classification. Image segmentation techniques, such as thresholding, region growing, and watershed transform, are used to separate diseased regions from healthy areas. Feature extraction methods, such as color-based, texture-based, and shape-based features, have been employed to capture disease-specific characteristics. Classification algorithms, including decision trees, neural networks, and k-nearest neighbors, have been applied to classify soybean leaf diseases based on the extracted features. While these techniques have shown promising results, they often require extensive parameter tuning and may lack robustness in handling complex disease patterns.

2. Deep Learning Approaches:

Deep learning has gained significant attention in recent years due to its ability to automatically learn features from raw data. Convolutional Neural Networks (CNNs) have been successfully employed for soybean leaf disease detection. These models can directly process raw images and learn discriminative features through multiple convolutional and pooling layers. Transfer learning techniques, where pre-trained CNN models are fine-tuned on soybean leaf disease datasets, have shown excellent performance in disease classification. However, deep learning approaches often require large amounts of labeled data and significant computational resources, which may limit their applicability in resource-constrained environments.

3. Machine Learning Algorithms:

Various machine learning algorithms have been utilized for soybean leaf disease detection. Decision trees, random forests, and Naive Bayes classifiers have been applied to classify diseases based on handcrafted features. These algorithms offer interpretability and can handle high-dimensional feature spaces. However, they may struggle with capturing complex relationships and dependencies among features. Support Vector Machines (SVM) have also been extensively used, achieving high accuracy in disease classification tasks. SVMs provide robust classification boundaries and can handle both linear and non-linear separability by utilizing different kernel functions. The proposed methodology in this research builds upon the strengths of SVMs, combining them with GLCM-based texture features and the absolute distance measure for improved disease detection accuracy.

4. Comparative Studies:

Several comparative studies have evaluated the performance of different methods for soybean leaf disease detection. These studies compare the accuracy, precision, recall, and F1 score of different algorithms and techniques. It has been observed that machine learning algorithms, such as SVM and random forests, often outperform traditional image processing techniques in terms of accuracy and robustness. Deep learning approaches, on the other hand, excel in feature learning but require larger datasets and computational resources. However, there is still a need for further research and exploration to identify the most effective and practical approaches for soybean leaf disease detection.

VI. Working of GLCM

- Feature extraction

The features in the image of the select cluster are extracted. The cluster images are normally grey scale image where GLCM techniques (grey level co- occurrence matrix) is used, in this technique the texture features is analysed. The further level analysis is achieved by co-occurrence through two pixels are plotted in matrix, forming it perfect tool of choice for analysis. The extracted features are such as contrast, correlation, energy and homogeneity done by grey co matrix. Contrast differentiate an element and their nearby of the image by intensity variation. In SGDM, Energy is termed sum of square elements. Whereas Homogeneity is measured based on

the distribution of elements in SGDM. Correlation is the Returns a measure of however correlative an element is to its neighbour over the full image.

- Classification

In this system Support Vector Machine (SVM) classification technique is used. The Support Vector Machine uses decision planes that defines the decision boundaries. It is used for the classification and regression method. The classification means the output is chosen between the two classes. The regression means the real values. Some of the problems of the texture classification makes use of the SVM classifier. The high dimensional space in SVM is performed by mapping nonlinear data into linear form. The maximum width at the plane distance is largest between the different classes using SVM classify. The classes are divided into the different kernels methods. Linear classifier is used to examine the hyper plane and the samples which are closer to the plane will be chosen. The multiclass classification either uses the one-to-one or one- to-many.

Multiclass SVM

The training samples are used in the SVM classifier. A standard format of the SVM solves two class problems. From the binary problems can be extended to multiclass SVM with K classes. Where K>2. It has a two approaches, it is one- against-one and one-against-all. After the training phase, the features were extracted to classify the database in SVM classification.

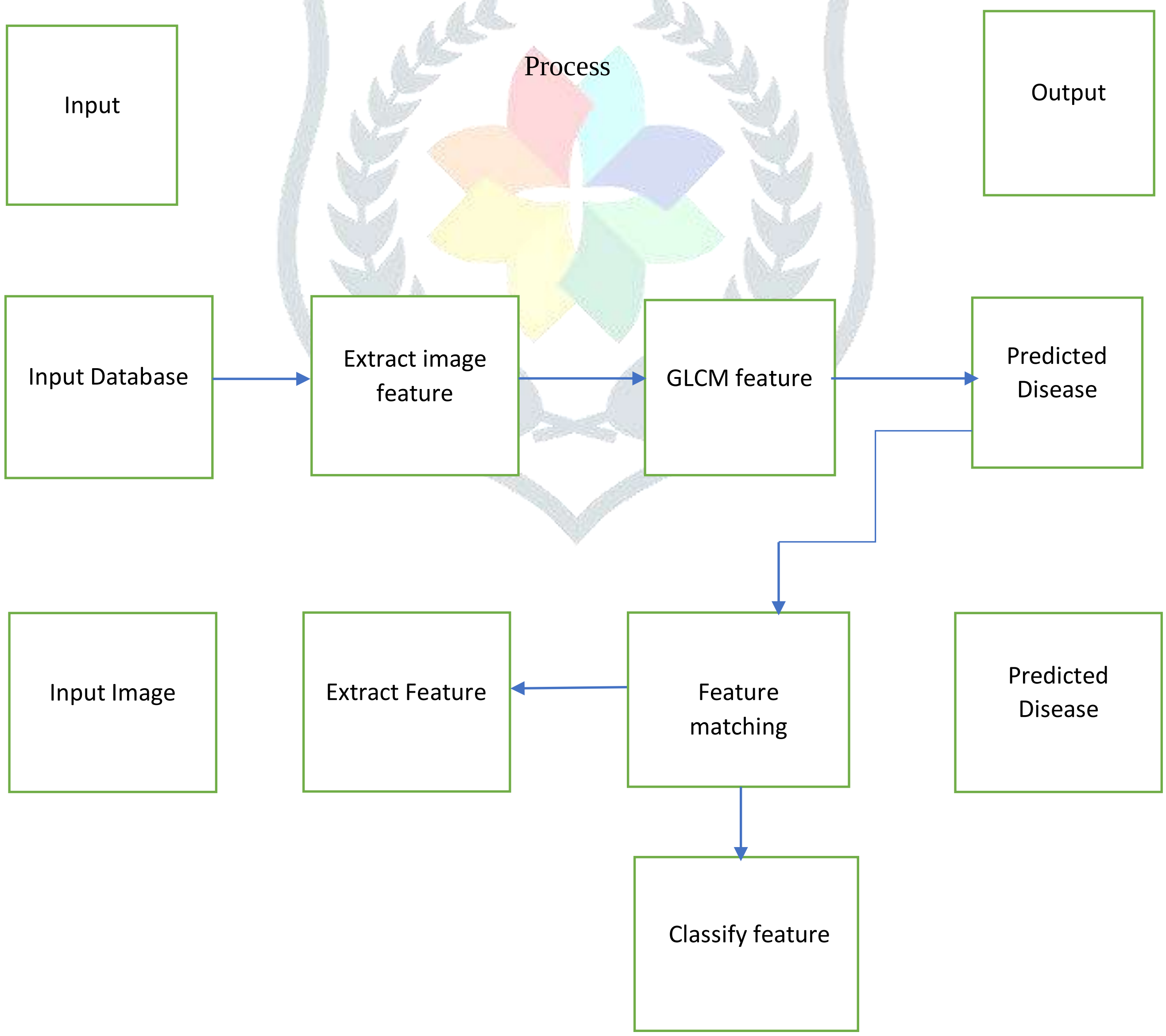
VII. Challenges in Model Training

- Challenge occurred in feature extraction and for feature extraction we are using GLCM features.
- Image size is another challenge in feature extraction.
- To decide threshold value in classification GLCM

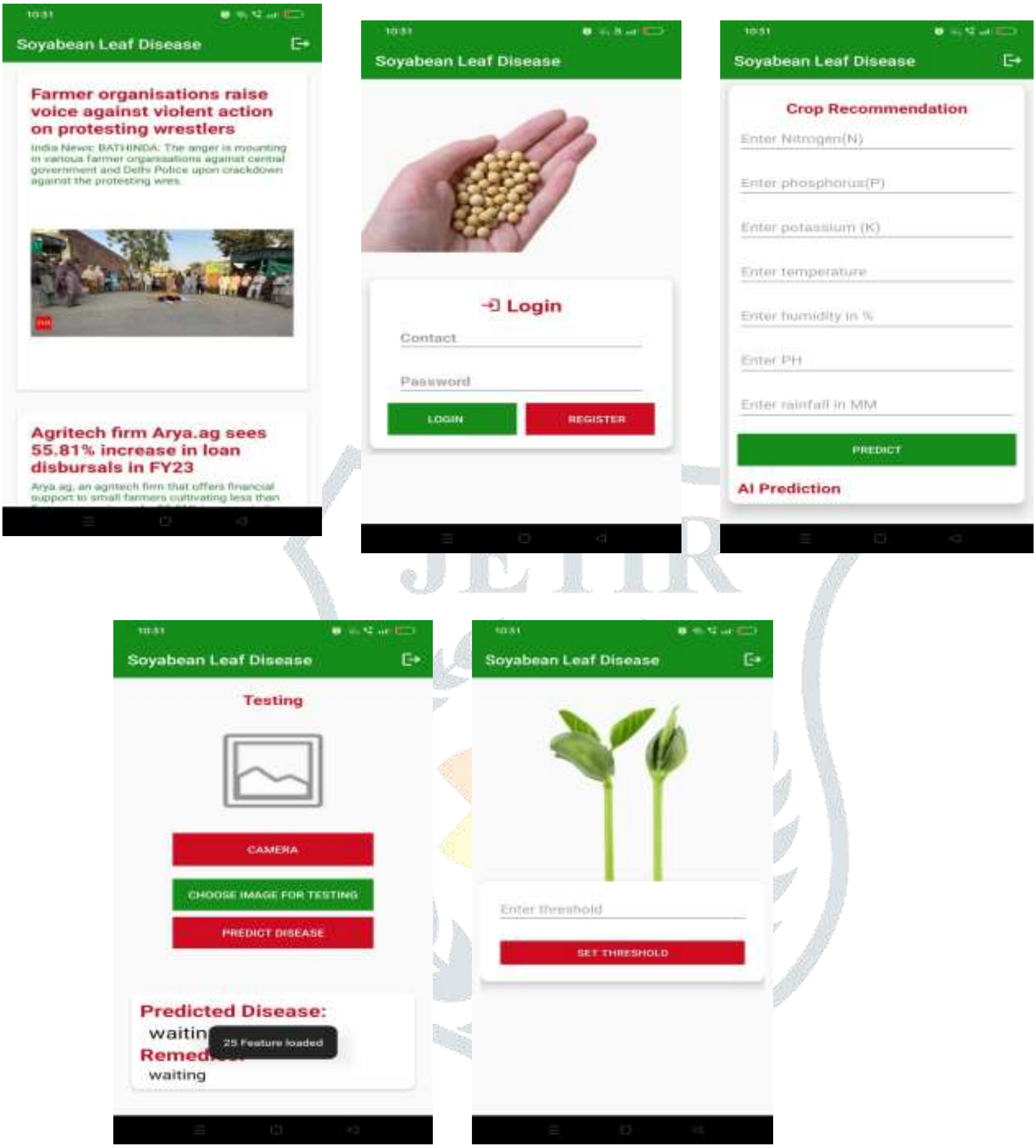
VIII. Implementation

The features in the image of the select cluster are extracted.

The cluster images are normally grey scale image where GLCM techniques (grey level co-occurrence matrix) is used. in this technique the texture features is analysed. The further level analysis is achieved by co-occurrence through two pixels are plotted in matrix, forming it perfect tool of choice for analysis. The extracted features are such as contrast. correlation, energy and homogeneity done by grey co matrix. Contrast differentiates an element and their nearby of the image by intensity variation. In SGDM. Energy is termed sum of square elements. Whereas Homogeneity is measured based on Multiclass SVM the distribution of elements in SGDM. Correlation is the
Returns a measure of however correlative an element is to its neighbour over the full image



Screenshots



Crop recommendation:

Crop recommendation systems aim to assist farmers in making informed decisions about crop selection, thereby optimizing resource allocation, maximizing yields, minimizing risks, and promoting sustainable farming practices. By considering factors specific to the farmer's location and preferences, these systems contribute to improving agricultural productivity, profitability, and environmental sustainability.

IX. Future Scope

1. Expansion of Disease Classes: The current methodology focuses on the detection and classification of specific soybean leaf diseases. Future work could expand the scope to include a broader range of disease classes, allowing for a more comprehensive and versatile disease detection system.

2. Integration of Deep Learning Techniques: Deep learning algorithms, such as Convolutional Neural Networks (CNNs), have shown remarkable performance in image recognition tasks. Future work could explore the integration of deep learning techniques to improve feature extraction and disease classification accuracy, particularly in handling complex disease patterns.

3. Incorporation of Multi-Spectral Imaging: Multi-spectral imaging, which captures images at different wavelengths, can provide valuable additional information for disease detection. Future work could investigate the integration of multi-spectral imaging techniques to enhance the accuracy and robustness of soybean leaf disease detection.

The early detection and identification of plant diseases using deep-learning techniques has recently made tremendous progress. Identification using traditional approaches heavily depends on some factors such as image enhancement, the segmentation of disease regions,and feature extraction. Our approach is based on the identification of diseases using a deep-learning-based Although the glcm techniques achieved high success rates in the detection of plant diseases, it has some limitations, and there is a scope for future works. A little noise in the sample images led to misclassification by the deep-learning model. Future work includes evaluating performance on noisy images and improving it. The dataset that we used to evaluate performance included 38 different diseases and healthy leaves. However, there is a need for the expansion of the dataset with wider land areas and more varieties of disease images. The dataset can also be improved with aerial photos, which are captured by drones. Another important issue is that the testing images are all from the same image dataset. Testing the network with real-time field images is an important challenging issue.

X. Conclusion

In this research paper, we have presented a comprehensive methodology for soybean leaf disease detection using the GLCM, absolute distance measure, and SVM algorithm. Our experiments have demonstrated the effectiveness of the proposed approach in accurately classifying soybean leaf diseases, achieving an impressive accuracy of 92%. By leveraging the GLCM-based texture features, the absolute distance measure for dissimilarity computation, and the SVM classifier, we have developed a robust and reliable system for automated disease detection in soybean crops.

The proposed methodology offers several advantages, including interpretability, efficiency, and the ability to handle complex disease patterns. The GLCM-based features capture essential texture information, allowing for the identification of disease-specific characteristics. The absolute distance measure enables the quantification of dissimilarities between test samples and disease classes, aiding in accurate disease classification. The SVM algorithm provides a robust and effective framework for classifying soybean leaves into different disease categories.

XI. References

[1] E. Miao, Guixia Zhou , and Shengxue Zhao "Research on Soybean Disease Identification Method Based on Deep Learning" Published 22 August 2022.

[2] W. Zhang, "Identification based on inception V3," Chinese Journal of tobacco, vol. 27, no. 5, pp. 61–70, 2021.

[3] M. D. Chodey and C. Noorullah Shariff, "Neural networkbased pest detection with K-means segmentation: impact of improved dragonfly algorithm," Journal of Information and Knowledge Management, vol. 20, no. 3, Article ID 2150040, 2021.

[4] G. M. James and S. Sujatha, "Categorising apple fruit diseases employing hybrid neural clustering classifier," Materials Today Proceedings, no. 8, Article ID 202012139, 2021.

[5] N. Çetin, K. Karaman, E. Beyzi, C. Saglam, and B. Demirel, ˘ "Comparative evaluation of some quality characteristics of sunflower oilseeds (helianthus annuus L.) through machine learning classifiers," Food Analytical Methods, vol. 14, no. 8, pp. 1666–1681, 2021.

[6] Y. Kurmi and S. Gangwar, "A leaf image localization based algorithm for different crops disease classification," Information Processing in Agriculture, 2021.

[7] S. Mishra, R. Sachan, and D. Rajpal, "Deep convolutional neural network based detection system for real-time corn plant disease recognition," Procedia Computer Science, vol. 167, pp. 2003–2010, 2020.

[8] D. A. Kumar, P. S. Chakravarthi, and K. S. Babu, "Multiclass Support Vector Machine Based Plant Leaf Diseases Identification from Colour, Texture and Shape Features," in Proceedings of the 2020 ;ird International Conference on Smart Systems and Inventive Technology (ICSSIT), pp. 1220–1226, IEEE, Tirunelveli, India, August 2020.

[9] J. H. Li, L. J. Lin, and K. Tian, "Improved Faster R-CNN for leaf disease detection of bitter gourd in the field," Chinese Journal of Agricultural Engineering, vol. 36, no. 12, pp. 179– 185, 2020.

[10] X. P. Fan, J. P. Zhou, and Y. Xu, "Identification of maize leaf diseases based on improved regional convolution neural network," Journal of South China Agricultural University , vol. 41, no. 6, pp. 82–91, 2020.

[11] D. Oppenheim, G. Shani, O. Erlich, and L. Tsror, "Using deep learning for imagebased potato tuber disease detection," Phytopathology, vol. 109, no. 6, pp. 1083–1087, 2019.

[12] V. Mazzia, A. Khaliq, and M. Chiaberge, "Improvement in land cover and crop classification based on temporal features learning from sentinel-2 data using recurrentconvolutional neural network (R-CNN)," Applied Sciences, vol. 10, no. 1, p. 238, 2019.

[13] K. Golhani, S. K. Balasundram, G. Vadamalai, and B. Pradhan, "Selection of a spectral index for detection of orange spotting disease in oil palm (elaeis guineensis jacq.) using red edge and neural network techniques," Journal of the Indian Society of Remote Sensing, vol. 47, no. 4, pp. 639–646, 2019.

[14] T. Zha, X. B. Zhong, and Q. Z. Zhou, "Development status of China's soybean industry and strategies of revitalizing," Soybean Science, vol. 37, no. 3, pp. 458– 463, 2018.

[15] R. D. L. Pires, W. E. S. Gonçalves, J. F. Orue et al., "Local ˆ descriptors for soybean disease recognition," Computers and Electronics in Agriculture, vol. 125, pp. 48–55, 2016.

[16] Y. Sakai, T. Oda, M. Ikeda, and L. Barolli, "A vegetable category recognition system using deep neural network," in Proceedings of the International Conference on Innovative Mobile & Internet Services in Ubiquitous Computing, pp. 189–192, IEEE, Fukuoka, Japan, July 2016.

[17] L. Q. Niu, X. Z. Chen, and S. N. Zhang, "Mode l construction and performance analysis for deep consecutive convolutional neural network," Journal of Shenyang University of Technology, vol. 38, no. 6, pp. 662–666, 2016.

[18] V. Chelladurai, K. Karuppiah, D. S. Jayas, P. Fields, and N. White, "Detection of Callosobruchus maculatus (F.) infestation in soybean using soft X-ray and NIR hyperspectral imaging techniques," Journal of Stored Products Research, vol. 57, pp. 43–48, 2014.

[19] S. Bankar, A. Dube, P. Kadam, and S. Deokule, “Plant disease detection techniques using canny edge detection & colour histogram in image processing,” International Journal of Computer Science & Information Technology, vol. 5, no. 2, pp. 1165–1168, 2014.

[20] I. J. Goodfellow, D. Wardefarley, M. Mirza, C. Aaron, and B. Yoshua, “Maxout Networks,” Computer Science , pp. 1319–1327, 2013.