Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-7, Issue-12, December 2025
DOI: 10.53469/jrse.2025.07(12).06

Enhanced Identification of Apple Leaf Diseases Through Optimized Machine Learning Algorithms

April Manallo[1], Vivien Chua[2]

[1]Research Scholar, Department of Research Computer Science, Tirupur Kumaran College for Women, Tirupur, Tamilnadu, India

[2]Associate Professor, Department of Research Computer Science, Tirupur Kumaran College for Women, Tirupur, Tamilnadu, India

Abstract: ***Apple orchards play a vital role in global food production, but they face constant threats from diseases like scab, rust, and rot. Accurate and timely identification of these diseases is paramount to prevent crop losses and ensure food security. In this study, we present a novel approach that integrates the Firefly Optimization (FFO) algorithm with Rough Set Theory (RST) to optimize feature selection for the identification of apple leaf diseases. We employ the K - Nearest Neighbors (k - NN) machine learning algorithm to classify various types of apple leaves, including healthy, rot, rust, and scab leaves. The results reveal the remarkable effectiveness of the optimized k - NN algorithm in comparison to the non - optimized version. The optimized approach significantly improves the True Positive Rate (TPR) and reduces the False Negative Rate (FNR) for each type of apple leaf disease, indicating enhanced accuracy in disease identification. These findings have substantial practical implications for apple growers, enabling informed management decisions to reduce crop losses. Furthermore, this study underscores the potential of machine learning, image processing, and feature selection techniques in advancing disease identification accuracy, with broader applications in agriculture. Ultimately, this research contributes to sustainable agricultural practices and global food security by harnessing technology to combat crop diseases effectively.***

Keywords: Firefly Optimization, Roughset Theory, Apple orchards, Disease severity measurement, Image processing techniques, Orchard management, Precision agriculture

1. Introduction

In the ever - evolving landscape of agriculture, the accurate identification and management of diseases affecting apple orchards stand as a critical challenge. Apple trees, esteemed for their bountiful fruit, confront relentless threats from diseases such as scab, rust, and rot. Swift and precise disease detection is essential for safeguarding orchard health, ensuring food security, and optimizing crop yields. This pursuit of accuracy and efficiency in disease identification finds its solution at the intersection of cutting - edge technology and innovative methodologies, as exemplified in this study.

Apple leaf diseases not only jeopardize crop productivity but also carry significant economic implications. Timely and accurate detection of these diseases is instrumental in mitigating crop losses and facilitating informed management decisions. The traditional methods of disease identification often relied on labor - intensive, time - consuming processes that were susceptible to human error. However, the integration of advanced technology, particularly machine learning and image processing, has redefined the landscape of disease detection in apple orchards(Zhang Y.).

The concept of feature extraction an indispensable process that bridges the gap between raw data and actionable insights. Feature extraction involves the distillation of pertinent information from intricate datasets, with a specific focus on those attributes that contribute most substantially to the classification and identification of apple leaf diseases(Jyotismita Chaki). It is within this context that our study introduces a novel approach: the integration of the Firefly Optimization (FFO) algorithm with Rough Set Theory (RST) to optimize feature selection. This approach aims to significantly enhance the accuracy and efficiency of apple leaf disease identification.

This introduction sets the stage for the comprehensive exploration of the research presented in this study. It highlights the importance of disease identification in apple orchards, outlines the challenges faced, and introduces the innovative approach employed to address these challenges. The subsequent sections will delve into the methodology, experimental results, discussions, and implications, ultimately showcasing the significant potential of this research in advancing disease management and agricultural sustainability. As we navigate the intersection of technology and agriculture, we embark on a journey to revolutionize disease detection in apple orchards, empowering growers and researchers to make informed decisions that will shape the future of apple cultivation.

2. Proposed Methodology

Traditionally, disease identification in apple orchards relied heavily on visual inspection by trained experts. However, this manual approach is time - consuming, labor - intensive, and inherently subjective. Human judgment, while valuable, is susceptible to inconsistencies and errors, especially in the face of subtle disease symptoms (Rome.).

The integration of technology, particularly machine learning and image processing, has emerged as a game - changer in the realm of disease detection in agriculture. These advanced techniques offer the promise of objective, rapid, and accurate identification of diseases based on visual cues.

The technological advancements lies the concept of feature extraction. This process involves distilling key information from complex datasets, with a focus on attributes that are most informative for disease classification. The selection of relevant features is pivotal in reducing computational complexity, enhancing classification accuracy, and enabling the development of efficient disease identification models (Wani).

Our study introduces a novel approach that marries the Firefly Optimization (FFO) algorithm with Rough Set Theory (RST) to optimize feature selection. By doing so, we aim to unlock the full potential of machine learning for apple leaf disease identification. The integration of FFO and RST offers a powerful means to identify the most informative features, thereby improving the accuracy and efficiency of disease classification.

2.1 Dataset Description

The dataset used in this study was sourced from the Plant Pathology Challenge, a part of the Fine - Grained Visual Categorization (FGVC) workshop held at the prestigious 2020 Computer Vision and Pattern Recognition conference (CVPR 2020). This dataset, meticulously compiled and categorized, forms the cornerstone of our research, enabling us to conduct a thorough analysis of apple leaf diseases and the development of an optimized identification approach(Holb).

- **Total Samples**: The dataset comprises a substantial total of 4840 samples of apple leaves, providing a diverse and comprehensive representation of apple leaf conditions.
- **Categorization by Health Condition**: The leaves within the dataset have been meticulously categorized into four distinct health conditions, namely:
- **Healthy Leaves**: Serving as the baseline reference for disease detection, this category encompasses 1809 samples of healthy apple leaves.
- **Scab - Affected Leaves**: Among the diseased leaves, 1460 samples belong to the scab category.
- **Rust - Affected Leaves**: Rust, another common disease, is represented by 750 samples.
- **Rot - Affected Leaves**: The dataset also includes 821 samples of leaves affected by rot, completing the quartet of health condition categories.

2.2 Preprocessing Methods

- **Noise Reduction**: To enhance the quality of the images and reduce data noise, noise reduction techniques such as Gaussian smoothing and median filtering were applied.
- **Contrast Enhancement**: Contrast adjustment techniques, like histogram equalization, were utilized to improve image clarity and accentuate relevant features.
- **Image Resizing**: Images were resized to a standardized resolution to ensure uniformity across the dataset and to reduce computational complexity during subsequent processing.
- **Color Normalization**: For datasets with color images, color normalization techniques were applied to standardize color values and reduce the impact of color variations on feature extraction (Konstantinos N. Plataniotis).

2.3 Feature Extraction

- **Shape - Based Features**: Essential characteristics of the physical structure of apple leaves, such as area, perimeter, form factor, w - max, and h - max, were extracted. These features serve as indicators of the unique shapes of both healthy and diseased leaves.
- **Texture - Based Features**: Variations in leaf textures are crucial cues for disease identification. Metrics like standard deviation, smoothness index, skewness, and kurtosis were computed to distinguish textures across leaves.
- **GLCM - Based Features**: Leveraging the Gray - Level Co - occurrence Matrix (GLCM), features like entropy and homogeneity were extracted to capture spatial relationships between adjacent pixels. These features unveil vital patterns within leaf images.
- **IDM (Inverse Difference Moment) Feature**: IDM, measuring the inverse difference between pixel intensities, was utilized to characterize the texture of the images, contributing to the overall feature set (Zhang Y.).

2.4 Feature Selection

While extracting a variety of features from apple leaf images provides a wealth of information, it's important to recognize that not all extracted features are equally relevant or useful for disease detection. The sheer number of features can lead to dimensionality challenges, increased computational complexity, and even the risk of overfitting in machine learning models. This is where feature selection, a critical step in the FFO - RST algorithm, comes into play(Singh) (Russ).

The Firefly Optimization with Rough Set Theory (FFO - RST) algorithm serves as an efficient and powerful tool for feature selection. It helps identify the most relevant features in a dataset while eliminating those that may not significantly contribute to the accuracy of disease detection. Here's a closer look at how FFO - RST accomplishes this:

1) **Firefly Optimization (FFO):** FFO, the initial step of the algorithm, ranks features based on their importance. By assessing the "brightness" or significance of each feature in terms of classification accuracy, FFO effectively identifies the most crucial features. This process effectively reduces the dataset's size, leading to improved computational efficiency.
2) **Rough Set Theory (RST):** After FFO, the optimized features are subjected to Rough Set Theory. RST provides a further refinement of the feature selection process. It accomplishes this by assessing the "dependency" of each feature within the dataset. Features with higher dependencies are considered more important, while those with lower dependencies may be removed. This step not only enhances the accuracy of feature selection but also minimizes computational time.

The integration of FFO and RST within the FFO - RST algorithm offers a balanced approach to feature selection. It leverages the strengths of both techniques to ensure that only the most relevant features are retained, effectively reducing noise in the dataset. By focusing on the most informative features, FFO - RST not only enhances accuracy but also enables earlier disease detection. This, in turn, empowers growers and researchers to implement timely control measures, ultimately leading to improved crop productivity and enhanced food security.

2.5 K - Nearest Neighbors (k - NN) Algorithm and Optimization with FFO - RST

The K - Nearest Neighbors (k - NN) algorithm is a versatile and intuitive machine learning method that falls under the category of supervised learning. It is particularly well - suited for classification tasks, such as the identification of apple leaf diseases in our study. Here, we'll explain the k - NN algorithm and then delve into how it was optimized using the Firefly Optimization (FFO) algorithm with Rough Set Theory (RST).

K - Nearest Neighbors (k - NN) Algorithm:
The k - NN algorithm operates on a simple yet effective principle: objects or data points that are similar tend to belong to the same class. In the context of our study, this means that similar patterns in apple leaf images are likely to belong to the same disease category. Here's how k - NN works:

1) **Training Phase**:

During the training phase, the algorithm stores the entire dataset along with their corresponding labels (in our case, the health condition of apple leaves). This training data serves as the knowledge base for the algorithm.

2) **Prediction Phase**:

- When presented with a new, unlabeled data point (an apple leaf image in our case), the k - NN algorithm makes predictions based on the similarity of that data point to its k - nearest neighbors in the training data.
- The "k" in k - NN represents the number of nearest neighbors to consider. For example, if k is set to 5, the algorithm looks at the five nearest neighbors to the new data point.

3) **Classification**:

The k - NN algorithm counts how many of the k - nearest neighbors belong to each class (e. g., healthy, scab, rust, or rot). The class with the majority of neighbors is assigned as the predicted class for the new data point.

4) **Decision Rule**:

Typically, a simple majority vote is used as the decision rule. In case of a tie, some variants of k - NN may use weighted voting, considering the distance to neighbors as weights.

Optimization with FFO - RST:
Now, let's explore how the k - NN algorithm was optimized using the Firefly Optimization (FFO) algorithm in conjunction with Rough Set Theory (RST):

1) **Feature Selection with FFO - RST**:

- Before feeding the data into the k - NN algorithm, feature selection is crucial. Not all extracted features may be equally relevant for classification.
- Firefly Optimization (FFO) was applied to rank the features based on their importance in terms of classification accuracy. FFO helps identify the most critical features while reducing the dataset's dimensionality, which is especially important for efficient computation.

2) **Further Refinement with Rough Set Theory (RST):**

- After FFO ranking, Rough Set Theory (RST) was employed for additional feature refinement. RST utilizes an attribute dependency measure to identify the most informative features while eliminating less important ones. This step helps improve the accuracy of the feature selection process and reduce computational complexity.

3) **Optimized k - NN Classification**:

- The selected features optimized by FFO - RST were then used as input for the k - NN algorithm. This optimization process ensured that the k - NN algorithm was working with the most relevant and informative features, enhancing its classification accuracy.

3. Experimental Results

The study's primary goal was to enhance the accuracy of identifying various types of apple leaves, including healthy, rot, rust, and scab leaves, by employing the K - Nearest Neighbors (k - NN) machine learning algorithm. The findings from the experiments have affirmed the effectiveness of the optimized k - NN algorithm when compared to its non - optimized counterpart.

One of the key performance metrics evaluated in the study was the True Positive Rate (TPR), which measures the proportion of actual positive cases correctly identified by the algorithm, and the False Negative Rate (FNR), which measures the rate of false negatives or instances where the algorithm failed to detect a positive case.

- For healthy leaves, the TPR improved from 90.8% to an impressive 95.8%, while the FNR decreased from 9.2% to a significantly lower 4.2%. This indicates a substantial enhancement in the algorithm's ability to correctly identify healthy leaves.
- In the case of Rot leaves, although the TPR decreased slightly from 87.5% to 92.5%, it is noteworthy that the FNR increased only modestly from 12.5% to 7.5%, demonstrating a more balanced and accurate classification of Rot leaves.
- Rust leaves exhibited an improvement in TPR from 89.3% to 90.3%, accompanied by a decrease in FNR from 10.7% to 9.7%, indicating a positive impact on the algorithm's ability to detect Rust leaves.
- For Scab leaves, the optimized approach achieved an enhanced TPR, increasing from 97.7% to 98.3%, while the FNR decreased from 2.3% to a mere 1.7%. This showcases the remarkable accuracy achieved in identifying Scab leaves.

The presented comparison of results from the non - optimized and optimized K - Nearest Neighbors (k - NN) algorithms serves as a compelling testament to the effectiveness of the proposed approach in enhancing the accuracy of apple leaf disease identification. The true positive rate (TPR) and false negative rate (FNR) metrics provide a clear and quantitative evaluation of the algorithm's performance for each type of apple leaf disease.

In the non - optimized k - NN algorithm, the initial TPR and FNR values were indicative of its capabilities: for healthy leaves, a TPR of 90.8% and an FNR of 9.2%; for Rot leaves, a TPR of 87.5% and an FNR of 12.5%; for Rust leaves, a TPR of 89.3% and an FNR of 10.7%; and for Scab leaves, a TPR of 97.7% and an FNR of 2.3%.

The real transformation emerges after the optimization process with the Firefly Optimization with Rough Set Theory (FFO - RST) algorithm. The substantial improvements in TPR and reductions in FNR are indeed noteworthy: for healthy leaves, a TPR of 95.8% and an FNR of 4.2%; for Rot leaves, a TPR of 92.5% and an FNR of 7.5%; for Rust leaves, a TPR of 90.3% and an FNR of 9.7%; and for Scab leaves, a remarkable TPR of 98.3% and an FNR of 1.7%.

These results unequivocally demonstrate that the optimized k - NN algorithm excels in its ability to accurately identify apple leaf diseases across all types. By significantly reducing the number of false negatives and increasing the true positive rate, the optimized approach equips growers with a robust tool for making informed management decisions, ultimately leading to the reduction of crop losses.

In conclusion, this study showcases the potential of machine learning algorithms and image processing techniques in advancing the accuracy of apple leaf disease identification. The innovative feature selection approach using the FFO - RST algorithm successfully optimized the K - Nearest Neighbors algorithm, resulting in superior accuracy in disease identification. The practical implications of these findings for apple growers are substantial, as precise disease identification enables timely and targeted interventions, such as pesticide applications, which can mitigate disease spread and boost crop yields.

4. Conclusion

Our study presents a pioneering approach to apple leaf disease identification, integrating advanced technology, feature optimization through Firefly Optimization with Rough Set Theory (FFO - RST), and the versatile K - Nearest Neighbors (k - NN) algorithm. The optimized k - NN model exhibited remarkable improvements in accuracy, reducing false negatives and empowering growers with a powerful tool for informed disease management. This research not only contributes to the precision of agricultural practices but also underscores the potential of technology - driven solutions to enhance food security and sustainability in agriculture.

References

[1] Abbas, A., Khan, M. M., & Khan, S. U. " Apple Diseases Detection and Classification Using Convolutional Neural Network. " *In Intelligent Systems Technologies and Applications.* Singapore: Springer, 2020.109 - 121.

[2] Abdul Saboor, Muhammad Usman, Sikandar Ali, Ali Samad, Muhmmad Faisal Abrar and Najeeb Ullah. "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms. " *Mobile Information Systems* (2022).

[3] Holb, I. J., & Schnabel, G. "Apple scab: biology, epidemiology, and management. " *Springer Science & Business Media* (2014).

[4] Jones, A. L., & Schilder, A. M. *Plant diseases caused by Pseudomonas syringae in the Pacific Northwest: identification and management. Pacific Northwest Plant Disease Management Handbook*.2015.

[5] Jyotismita Chaki, Nilanjan Dey. *Texture Feature Extraction Techniques for Image Recognition*. Springer Nature Singapore, 2019.

[6] Khan, S. M., Maqbool, M., & Bhat, M. M. "Bacterial canker of apple and its management: a review. " *International Journal of Agriculture and Biology* 17.6 (2015): 1091 - 1096.

[7] Konstantinos N. Plataniotis, Anastasios N. Venetsanopoulos. *Color Image Processing and Applications*. Springer Berlin Heidelberg, 2013.

[8] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. "SSD: Single shot multibox detector. . " *In European conference on computer vision.* Springer, Cham, 2016.21 - 37.

[9] Rome., FAO.2021. World Food and Agriculture - Statistical Yearbook 2021. n. d.

[10] Russ, John C. *The Image Processing Handbook*. CRC Press, 2006.

[11] Singh, P., & Choudhary, P. "PotatoNet: A Deep Learning - Based Approach for Potato Disease Detection. " *In International Conference on Intelligent Computing and Communication*. Springer, Cham, 2021.455 - 463.

[12] Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. *Advances in deep learning*. Springer, 2020.