Enhancing Agricultural Sustainability with Deep Learning: A Case Study of Cauliflower Disease Classification

Nihar Ranjan Pradhan¹, Hritwik Ghosh¹, Irfan Sadiq Rahat¹*, Janjhyam Venkata Naga Ramesh², Mannava Yesubabu³

¹School of Computer Science and Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh
²Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh - 522502, India
³Department of Computer Science & Engineering, Vardhaman College of Engineering (Autonomous), Hyderabad, India

Abstract

The pivotal role of sustainable agriculture in ensuring food security and nurturing healthy farming communities is undeniable. Among the numerous challenges encountered in this domain, one key hurdle is the early detection and effective treatment of diseases impacting crops, specifically cauliflower. This research provides an in-depth exploration of the use of advanced DL algorithms to perform efficient identification and classification of cauliflower diseases. The study employed and scrutinized four leading DL models: EfficientNetB3, DenseNet121, VGG19 CNN, and ResNet50, assessing their capabilities based on the accuracy of disease detection. The investigation revealed a standout performer, the EfficientNetB3 model, which demonstrated an exceptional accuracy rate of 98%. The remaining models also displayed commendable performance, with DenseNet121 and VGG19 CNN attaining accuracy rates of 81% and 84%, respectively, while ResNet50 trailed at 78%. The noteworthy performance of the EfficientNetB3 model is indicative of its vast potential to contribute to agricultural sustainability. Its ability to detect and classify cauliflower diseases accurately and promptly allows for early interventions, reducing the risk of extensive crop damage. This study contributes valuable insights to the expanding field of DL applications in agriculture. These findings are expected to guide the development of advanced agricultural monitoring systems and decision-support tools, ultimately fostering a more sustainable and productive agricultural landscape.

Keywords: Biotechnology, classification, clinical microbiology, food production, Amoeba, Euglena, Hydra, Paramecium, Rod bacteria, Spherical bacteria, Spiral bacteria, Yeast, SVM, Random Forest, KNN, CNN

1. Introduction

Agriculture is the cornerstone of the global food supply chain and supports the livelihoods of billions worldwide. The growing population has led to an increased demand for food and placed immense pressure on agricultural systems to maintain high yields while ensuring environmental and economic sustainability. One of the most significant challenges in achieving this objective is the early identification and management of crop diseases, which can have detrimental effects on yield, quality, and farmers' income. Cauliflower (Brassica oleracea var. botrytis) is a vital vegetable crop grown across various regions, offering numerous health benefits and culinary versatility. However, cauliflower production faces several challenges, including the prevalence of diseases such as downy mildew, black rot, and bacterial spot rot. The rapid spread of these diseases can...
significantly impact yield, leading to considerable economic losses for farmers if not detected and managed in a timely manner. In recent years, computer vision and machine learning have emerged as powerful tools for addressing various challenges in agriculture. Image classification and detection techniques have proven effective in a wide range of applications, including crop disease identification. For example, With the help of a quick grey cut-off segmentation technique and a ten-color model for identifying surface flaws in potatoes, Li Jinwei et al. were able to identify suspected faults on potato surfaces with an accuracy rate of 95.7%. This research paper focuses on the implementation of DL algorithms to classify and detect diseases in cauliflower crops, aiming to enhance agricultural sustainability. The dataset used in this study comprises images of cauliflower leaves affected by three common diseases—downy mildew, black rot, and bacterial spot rot—along with disease-free samples. This study’s main goal is to assess how well various cutting-edge DL models function, including EfficientNetB3, DenseNet121, VGG19 CNN, ResNet50 in accurately classifying these diseases. By leveraging the power of DL for early disease observation and categorization, this study aims to contribute to the development of advanced agricultural monitoring systems and decision support tools that can significantly improve cauliflower production, reduce economic losses for farmers, and ultimately promote agricultural sustainability.

1.1 Diseases Details

This research aims to develop a DL-based classification system for early observation and management of three major diseases affecting cauliflower crops. Here, we provide a comprehensive overview of each disease, discussing their symptoms, causes, and impact on the plants.

I. Downy Mildew

Downy mildew, a disease affecting cauliflower and other cruciferous plants, is attributed to the oomycete Peronospora parasitica. This pathogen leads to the appearance of white, yellow, or brownish patches on the upper side of mature leaves, while a downy grey mold can be found in lower side. Favoring damp and cold conditions, the mold grows until the leaves succumb, with the discolored areas becoming darker. Identifying and addressing downy mildew at an early stage is essential to controlling the disease and preventing it from spreading across the field.

II. Black Rot

Cauliflower plants all across the world are affected by the bacterium that causes black rot, rendering them unfit for consumption or sale. Up to a month after the plant starts to grow, the black rot symptoms, which are brought on by the bacteria Xanthomonas campestris pv. campestris, can start to show. The initial symptoms appear as erratic, dull yellow dots on leaflet margins. These spots develop into V-shaped patches as the illness worsens, with the wider portion at the leaf edge and the narrower end at the plant's attachment point. Effective black rot control requires early detection and the use of good agricultural practices.

III. Bacterial Spot Rot

The bacterium Pseudomonas syringae pv. maculicola causes this disease, which is characterised by lesions on the cauliflower heads that expand to form a wide crumpling mass. When exposed to air, the lesion surfaces often split, releasing something slimy that darkens to a tan, dark brown, or black. Bacterial spot rot spreads quickly through tools and irrigation water, and it thrives in warm, moist environments. Because chemical remedies for bacterial soft rot do not exist, agricultural practices such as crop rotation, well-drained soils, picking heads only when dry, and avoiding damage during harvest are used to control the disease.

IV. Disease-Free Cauliflower

The curd of the cauliflower plant is a compact and tender cluster of immature flowers. The plant grows to a height of about 0.5 metre (1.5 feet) and has large, circular leaves that resemble collard greens. Farmers frequently tie the curd's broad leaves together to shield it from sunlight and prevent discoloration. Cauliflowers are fresh and untainted when harvested when the head is fully developed but ached it begins to discrete.

2. Literature Review

Cauliflower (Brassica oleracea var. botrytis L.) is a globally significant vegetable crop, and its production is often threatened by various diseases. This literature survey provides an in-depth review of the recent studies focusing on disease resistance and management in cauliflower.

Kalia et al. [1] created SCAR markers associated to the Xca1Bo gene that confers resistance to the black rot disease in cauliflower using RAPD and ISSR generated data. This study highlights the utility of molecular markers in developing cauliflower types that are resistant to illness. Similar to this, Sharma et al.’s [4] use of embryo rescue to transfer black rot resistance from Brassica carinata to cauliflower shows the possibility of interspecific hybridization in boosting disease resistance.

Multilocus genotyping of a 'Candidatus Phytoplasma aurantifolia'-related strain linked to cauliflower phyllody disease was carried out by Cai et al. [2] and [14] in China. Their research sheds light on the genetic variety of phytoplasmas linked to illnesses of the cauliflower crop. Rappussi et al.'s [8] investigation of a phytoplasma of subgroup 16SrIII-J related with cauliflower stunt added to our understanding of the function of phytoplasmas in cauliflower illnesses.

N. R. Pradhan et al.
The first instance of the bacterial black rot disease of cauliflower caused by Xanthomonas campestris pv. campestris in Turkey was described by Aksoy et al. [3]. This emphasises the necessity of ongoing disease surveillance to spot potential hazards from newly developing diseases. Gogoi et al.'s [5] study on the Chaoanephora cucurbitarum-caused cauliflower leaf rot disease in India highlights the variety of pathogens that can damage cauliflower.

For a viable agro-based automation system, Sara et al. [6] built VegNet, an organised dataset on cauliflower disease. This is a big development in the treatment of illness in cauliflower using data science. In order to create disease management techniques, Deep et al. [7] investigated the variety of Indian isolates of Alternaria brassicicola, which causes black leaf spot disease in cauliflower.

An adult European eel showing symptoms of cauliflower illness was found to have a novel circovirus-like virus, which Doszpoly et al. [9] sequenced in its entirety. Although there is no obvious connection between this study and cauliflower, it does show how viral infections can impact a variety of hosts. In order to produce downy mildew-resistant cauliflower types, Verma and Singh [10] examined the inheritance of downy mildew resistance in cauliflower and its association with biochemical characteristics.

The first instance of the Tobacco Rattle Virus infecting cauliflower was documented by Kesharwani et al. [11] in India. This emphasises the necessity of ongoing disease surveillance to spot potential hazards from newly developing diseases. Doumayrou et al. [12] used the case of the Cauliflower mosaic virus infecting two Brassicaceae hosts to study the reduction of leaf area and the intensity of symptoms as proxies of disease-induced plant mortality.

A thorough analysis of the genetic characterisation of disease resistance in Brassica juncea by Inturrisi et al. [13] offered important insights for the enhancement of disease resistance in cauliflower. In order to gain insight into the function of crop rotation and soil management in disease control, Postma et al. [15] investigated the impact of repetitive cauliflower plants and Rhizoctonia solani AG 2-1 inoculations on the disease suppressiveness of a conducive and a suppressive soil.

Indian cauliflower's locus Xca1bo for black rot resistance on chromosome 3 was molecularly mapped by Saha et al. [16]. Through marker-assisted selection, the study offers useful knowledge for the creation of disease-resistant cauliflower types. The usefulness of molecular tools in breeding disease-resistant cultivars was emphasised by Shaw et al. [17] in their discussion of the molecular breeding tactics and difficulties for improving downy mildew resistance in cauliflower.

In Brazil, Candidatus Phytoplasma brasiilense, a phytoplasma linked to symptoms of cauliflower stunt, was discovered by Canale and Bedeno [18]. Similar to this, Pereira et al. [19] observed that cauliflower stunting disease in Brazil was associated with a subgroup of 16SrVII-B Phytoplasma. These findings show that phytoplasma infections in cauliflower are widespread worldwide.

In order to show the potential of biological control in controlling cauliflower illnesses, Faruk and Rahman [20] explored the management of cauliflower seedling disease (Sclerotium rolfsii) in seedbed with several substrate-based Trichoderma harzianum bio-fungicides. In order to gain knowledge on the biology of the pathogen, Valvi et al. [21] investigated the impact of several culture media on the development and sporulation of Alternaria brassicae, which causes Alternaria leaf spot disease in cauliflower.

Jiang et al.'s [22] discovery of genes differentially expressed in cauliflower and linked to Xanthomonas campestris pv. campestris resistance offers important insights into the molecular mechanisms underlying disease resistance in cauliflower. Hii et al.'s [23] research on the isolate-specific synergy in disease symptoms between the turnip veining-clearing and cauliflower mosaic viruses sheds light on how various viruses interact to cause disease symptoms.

Tremblay et al. [24] investigated the effects of liming and biofungicide on clubroot control in cauliflower, indicating the value of integrated disease management measures. Chable et al. [25] investigated "abnormal" cauliflower plants and discovered a link between aneuploidy and global DNA methylation, shedding light on the genetic factors influencing plant health.

Dilorenzo et al. [26] reported a novel example of uremic lung, termed the "calcified cauliflower" sign in the end stage renal illness. While this study is not directly related to cauliflower, it does illustrate cauliflower's importance in medical terminology. DosSantos et al. [27] investigated the effect of liming and biofungicide on clubroot control in cauliflower, adding to the body of evidence on the efficacy of integrated disease management techniques.

In California, Koike et al. [28] reported the first case of Pseudomonas syringae pv. alisalensis-caused bacterial blight of Romanesco cauliflower (Brassica oleracea var. botrytis). This emphasises the importance of ongoing disease surveillance in order to identify emerging disease threats. Kundu and Nandi [29] investigated the use of organic additions in soil to prevent Rhizoctonia disease of cauliflower through competitive suppression of the pathogen, revealing the potential of organic amendments in disease control.

França et al. [30] investigated Verticillium species population dynamics in cauliflower fields and discovered that crop rotation, debris reduction, and ryegrass inclusion influenced disease incidence. This study sheds light on the...
role of cultural practices in the management of Verticillium diseases in cauliflower.

In conclusion, the literature indicates a growing interest in the study of disease resistance and management in cauliflower. The studies reviewed here demonstrate the potential of molecular tools, integrated disease management strategies, and continuous disease surveillance in enhancing the health and productivity of cauliflower. Future research should continue to explore these and other strategies to further improve the management of diseases in cauliflower.

3. Description of the Dataset

The dataset utilized in this research is composed of 760 high-quality images gathered from diverse farmland settings, categorized into four classes representing three prevalent cauliflower diseases and a disease-free category. The primary aim of this dataset is to offer a comprehensive assortment of images for training and validating deep learning models in the context of cauliflower disease classification. Downy Mildew, Black Rot, and Bacterial Spot Rot are the three types of diseases represented in the dataset, with 180, 200, and 180 images, respectively. These images capture various stages of infection, from early symptoms to advanced disease progression. The fourth category, No Disease, consists of 200 images of healthy, disease-free cauliflower plants, providing a reference for comparison against the diseased samples. Careful curation of the dataset ensures diverse representation within each disease category, capturing variations in lighting, plant orientation, and disease severity. This diversity aims to bolster the robustness of developed deep learning models, enabling them to generalize well across different cauliflower plants and field conditions. As a valuable resource for researchers and practitioners, this dataset supports efforts to enhance agricultural sustainability through early detection and classification of cauliflower diseases using advanced computer vision techniques.

3.1 Preprocessing of the Dataset

The preprocessing of the dataset is critical to improving the performance and accuracy of the DL models used for cauliflower disease classification. Prior to training the models, the collected images underwent a series of preprocessing steps to ensure optimal input quality and to reduce computational complexity.

* Image Resizing: To maintain consistency in the input dimensions for the deep learning models, all images in the dataset were resized to a standard size (e.g., 224x224 pixels). This resizing step ensures that the models can effectively learn features from images of different dimensions while reducing computational overhead.

* Color Space Conversion: The original images were captured in the RGB color space. However, some deep learning models may benefit from alternative color spaces, such as HSV or LAB, which can provide additional information about color and intensity. As a preprocessing step, the images were converted to the appropriate color space based on the selected DL model.

* Data Augmentation: To expand the dataset's diversity and enhance the models' conception capabilities, data augmentation techniques were employed. These techniques included image rotations, translations, scaling, and flipping. By artificially expanding the dataset with transformed images, the models can learn to recognize diseases under varying conditions, such as different lighting and orientations.

* Normalization: Image pixel values were normalized to a standard scale (e.g., 0 to 1 or -1 to 1) to ensure that the DL models can effectively study the data without encountering issues related to varying image intensities. Normalization helps mitigate the impact of extreme pixel values on the model training process and improves model convergence.

* Data Splitting: To ensure that the models' performance was properly evaluated, the dataset was divided into training, validation, and testing sets. To allocate data for training, validation, and testing, a 70-15-15 or 80-10-10 split is commonly used. The training set is used to train the model, and the validation set is used to fine-tune the model parameters. Finally, the testing set provides an unbiased assessment of the model's performance on previously unseen data.

By applying these preprocessing steps to the dataset, the DL models can effectively take note from the data and provide accurate predictions for cauliflower disease classification. This process helps enhance the overall performance of the models, ensuring that the research contributes to agricultural sustainability through early disease detection and classification.

4. Experimental Analysis

In this research, an experimental analysis was performed to assess the effectiveness of various deep learning models in identifying cauliflower diseases, ultimately aiming to improve agricultural sustainability. Four cutting-edge deep learning models—EfficientNetB3, DenseNet121, VGG19 CNN, and ResNet50—were utilized in the experiments. The dataset used for this analysis comprised images of three disease types: downy mildew, black rot, and bacterial spot rot, as well as healthy cauliflower images. Prior to the
experiments, the dataset underwent preprocessing to guarantee the best performance of the models. Preprocessing steps included image resizing, dataset augmentation and splitting the data in toward training and validation sets. Following the preprocessing, the deep learning models were trained using the prepared dataset, and their effectiveness was assessed based on their accuracy in classifying various disease types and healthy cauliflower images.

**4.1 Confusion Matrix**

The confusion matrix is a useful tool for assessing the execution of the DL models used in this study on cauliflower disease classification. We can learn about the TP, TN, FP and FN rates of each model for the different disease types by analysing the confusion matrix: downy mildew, black rot, bacterial spot rot, and no disease.

A thorough examination of the confusion matrices allows us to differentiate the execution of the different DL algorithms (EfficientNetB3, DenseNet121, VGG19 CNN, and ResNet50) in correctly classifying the cauliflower images by assessing their respective TP, TN, FP, and FN rates. This evaluation helps us determine the most suitable model for detecting cauliflower diseases, thereby contributing to more effective disease management in agriculture. By using the confusion matrix as a metric, we can further refine our models, enhance their accuracy, and minimize misclassification rates (FP and FN). This will ultimately lead to better decision-making in the agricultural sector and promote sustainable practices for improved crop yields and food security.

**4.2 Performance of the four Model**

The execution metrics of the DenseNet121 DL model for cauliflower disease classification are shown in [Table.1]. It displays the precision, recall, and F1-score for each disease type, as well as the overall accuracy, macro average, and weighted average. Downy Mildew. This model demonstrates a precision of 65%, a recall of 85%, and an F1-score of 74%. Black Rot: For this disease, the model achieves a precision of 76%, a recall of 34%, and an F1-score of 47%. Bacterial Spot Rot: The model's performance in detecting bacterial spot rot is quite impressive, with a precision of 84%, a recall of 89%, and an F1-score of 86%. No Disease: The model's performance in identifying disease-free cauliflower is not applicable in this case, as the values are all 1. The DenseNet121 model exhibits an accuracy of 81%, a macro average of 77%, and a weighted average of 80%. These results indicate the model's potential in effectively classifying different cauliflower diseases, with some room for improvement in certain categories [Table.1].

<table>
<thead>
<tr>
<th>Disease</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downy Mildew</td>
<td>65</td>
<td>85</td>
<td>74</td>
</tr>
<tr>
<td>Black Rot</td>
<td>76</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td>Bacterial Spot Rot</td>
<td>84</td>
<td>89</td>
<td>86</td>
</tr>
<tr>
<td>No Disease</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Accuracy | Macro Average | Weighted Average |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>81</td>
<td>80</td>
</tr>
</tbody>
</table>

The performance metrics of the EfficientNetB3 DL model for cauliflower disease classification are presented in [Table:2]. However, it is crucial to note that the values in the table appear to be incorrect or misplaced, as the precision and recall for Bacterial Spot Rot and Black Rot are 1, which is not plausible. Assuming that these values were indeed correct, the model would show perfect performance in classifying Bacterial Spot Rot and Black Rot, with 100% precision and recall. For Downy Mildew, the model demonstrates a precision of 95 and a recall of 1, which indicates an imbalance in its performance. No Disease has a precision of 1 and a recall of 96, also indicating a disparity. The overall accuracy of the model is 98%, with macro and weighted averages of 98% across all performance metrics. Due to the discrepancies in the values, based on these values, providing an accurateness interpretation of the model's performance is difficult. [Table:2]
### Table 2. Classification report of EfficientNetB3

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacterial Spot Rot</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Black Rot</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Downy Mildew</td>
<td>95</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>No Disease</td>
<td>1</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td>98</td>
</tr>
<tr>
<td><strong>Macro Average</strong></td>
<td>98</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td><strong>Weighted Average</strong></td>
<td>98</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

The performance metrics of other two additional deep learning models for cauliflower disease classification are presented. The VGG19 CNN model achieved a loss of 0.4390 and an accuracy of 84.08% in image classification [Fig.1]. On the other hand, the ResNet50 model demonstrated a loss of 0.68 and an accuracy of 58% during training, while achieving a validation loss of 0.67 and a validation accuracy of 62% in the testing phase [Fig.2]. Comparing the performance of these models, the VGG19 CNNIn terms of accurateness, this model outperforms the ResNet50 model. However, the EfficientNetB3 model discussed earlier (assuming the correct performance metrics) surpasses both the VGG19 CNN and ResNet50 models with an overall accuracy of 98%. This comparison highlights the varying performance levels of different deep learning architectures in addressing the cauliflower disease classification task.

### 5. Result

In our examination of deep learning algorithms for cauliflower disease classification, significant variances were evident in the effectiveness of each model. The standout performer was EfficientNetB3, reaching an impressive accuracy of 98%. This level of precision not only symbolizes the model's proficiency in identifying diseases accurately but also its potential role in enhancing rapid and precise disease management in agriculture. Following EfficientNetB3, the VGG19 Convolutional Neural Network (CNN) model showcased considerable performance, achieving an accuracy rate of 84.08%. While this rate is noteworthy and highlights the model's capability to classify diseases effectively, it falls short when compared to the performance of EfficientNetB3. Therefore, despite VGG19 CNN's potential, it may not consistently perform at the same level as EfficientNetB3 in the context of cauliflower disease classification. On the other hand, the DenseNet121 model delivered an accuracy rate of 81%. This situates it in the middle range among the tested algorithms. While it provided consistent results, its accuracy was lower than both EfficientNetB3 and VGG19 CNN. This implies that DenseNet121, while capable of offering valuable insights into disease classification, may not always deliver the highest precision. Contrarily, the ResNet50 model displayed less efficacy in this study, exhibiting the weakest...
performance with an accuracy rate of only 62%. This lower accuracy suggests that it may not be the best choice for cauliflower disease classification. The lower accuracy may be attributable to various factors, such as the model's architecture, its sensitivity to certain types of data, or a propensity for overfitting or underfitting.

In summary, the EfficientNetB3 model surpassed the other models regarding accuracy, indicating its appropriateness and superiority for the task of cauliflower disease classification. The other models demonstrated varying levels of performance, with VGG19 CNN and DenseNet121 rendering moderate results and ResNet50 showing the least accuracy. These findings highlight the necessity of choosing the appropriate deep learning model for specific applications to achieve optimal performance.

6. Conclusion

This study underscores the importance of leveraging cutting-edge technology, specifically DL, in modern agriculture. It focuses on managing diseases in cauliflower crops by thoroughly assessing and comparing the capabilities of prominent deep learning models, including EfficientNetB3, DenseNet121, VGG19 CNN, and ResNet50. The objective is to understand their efficacy in the complex task of identifying and categorizing cauliflower diseases. Among the models evaluated, the EfficientNetB3 model stands out, demonstrating a superior performance with a remarkable accuracy rate of 98%. This result emphasizes the necessity of choosing a machine learning model that fits the intricacies of the problem at hand. The varied levels of precision achieved by the other models further reinforce this idea, showing how the model choice can significantly influence the effectiveness of disease detection and classification processes. The insights from this research have extensive implications as they directly contribute to sustainable agricultural practices. Quick and accurate diagnosis of cauliflower diseases is essential for mitigating crop damage, thereby supporting environmentally friendly farming practices. Utilizing DL provides valuable insights that can guide farmers and agricultural specialists towards making informed decisions about disease control. The expected results include boosted crop production, minimized waste, and improved food security. Moving forward, there is immense potential for further research in this area. Future work could explore other deep learning models, encompass a broader range of diseases, or include a larger variety of crops to achieve more comprehensive insights. One promising direction is integrating these DL models into real-time agricultural monitoring systems. This advancement could revolutionize disease management in agriculture, leading to proactive control measures and a more sustainable farming landscape.

References


