A survey on deep learning-based disease identification and classification on cauliflower vegetable

1.Prof. Preethi B,

Assistant Professor, Department of Computer Science and Engineering, BIET,Davangere

2.Prof. Vinutha H P,

Professor,Department of Information Science and Engineering, BIET,Davangere

3.Prof. N.S.Patil,

Associate Professor,Department of Information Science and Engineering,BIET,Davangere

Abstract

This survey paper addresses the crucial role of deep learning, in identifying and classifying diseases in cauliflower, a key crop in precision agriculture. Traditional methods based on skilled visual examination are being replaced by advanced image analysis techniques due to their limitations in efficiency and consistency. Cauliflower, being susceptible to a variety of diseases like black rot and downy mildew, poses significant challenges in agricultural production and public health. Through a comprehensive review of recent studies, we explore the application of various CNN architectures such as ResNet50, InceptionV2, and MobileNetV1, particularly focusing on their implementation in datasets and their effectiveness in achieving notable accuracy rates. This paper also discusses emerging approaches like transfer learning and domain adaptation, which address issues of dataset diversity and real-world application. The current research gaps include dataset reliance and geographic limitations, and suggest integrating handcrafted with deep features for improved disease detection in cauliflower. This study aims to offer insights into the advancements and challenges in applying deep learning for plant disease detection and classification.

Keywords: Cauliflower vegetable, Deep learning, Agriculture, CNN Architecture, Plant disease detection

1   INTRODUCTION

The utilization of image analysis for the identification of plant diseases on leaves is a crucial focus within the field of precision agriculture research. Plant tissues are subjected to visual inspection by trained specialists in order to accurately document the extent of plant diseases. [1]. The widespread adoption of digital cameras and advancements in information technology within the agricultural sector have resulted in the extensive utilization of expert systems for cultivation and management. This has had a substantial impact on enhancing plant output capacity. The extraction and description feature of expert systems for pests and diseases primarily depend on expert knowledge, resulting in elevated expenses and reduced efficiency.

According to a reliable source, cauliflower is ranked as the second most prevalent "cole" crop globally, following cabbage. Farmers are the primary cultivators of this particular seasonal crop within agricultural surroundings. The mineral content of it includes iron, magnesium, phosphorus, potassium, salt, as well as vitamins A and B1. Cauliflower is known for its high nutritional density, which can be attributed to its low-fat content and high levels of fibre, vitamin B9, L-ascorbic acid, and water [2]. Cauliflower plant diseases play a significant role in diminishing vegetable yield and negatively impacting the agro-economic sector. Cauliflower plants can be susceptible to various diseases, including clubroot, blackleg, black rot, bacterial spot rot, downy mildew, powdery mildew, black rot, sclerotinia stem rot, white rust, cauliflower mosaic, ringspot, and others. These diseases have the potential to cause harm to cauliflower plants. Furthermore, it has been observed that cauliflower and its leaves are susceptible to various diseases, such as downy mildew, black rot, and bacterial spot rot [3].

Types of diseases which affect cauliflower:

- **Brassica Alternaria Leaf Spot**:
  Brassica Alternaria Leaf Spot, caused by the fungi Alternaria brassicicola or Alternaria brassicae, is a significant disease affecting Brassica family crops like cabbage, broccoli, and cauliflower. Characterized by dark, concentrically ringed spots on leaves, stems, and occasionally flowers, it thrives in warm, moist environments. The disease can reduce photosynthesis, impair plant growth, and diminish seed quality and yield. It spreads through wind, water, and infected seeds, and management strategies include crop rotation, using disease-free seeds, good sanitation practices, and fungicidal treatments. Developing resistant cultivars also plays a crucial role in mitigating the disease's impact, making it a vital focus for farmers and gardeners to ensure healthy Brassica crops.

Figure 1 **Brassica Alternaria Leaf Spot**

- **Brassica Black Rot**
  Brassica Black Rot, caused by the bacterium Xanthomonas campestris pv. campestris, is a devastating disease impacting Brassica crops like cabbage, broccoli, and cauliflower. Characterized by yellow, V-shaped lesions on leaves that turn black and cause wilting, it leads to significant yield loss and quality reduction, particularly in warm, humid climates. The disease spreads primarily through water, infected seeds, and plant debris, making field hygiene, crop rotation, and the use of disease-free planting materials crucial for management. Chemical treatments are largely ineffective against this bacterial infection, emphasizing the importance of preventive measures and the development of resistant crop varieties. Controlling this disease is essential for maintaining the health and productivity of Brassica crops.

Figure 2 Brassica Black Rot

● **Brassica Downy Mildew**

Brassica Downy Mildew, caused by the oomycete Hyaloperonospora brassicae, is a common disease affecting Brassica crops such as cabbage, broccoli, and cauliflower, especially in cool and moist environments. It manifests as yellow or pale green patches on the upper surfaces of leaves and fluffy, downy growth on the undersides, leading to leaf decay, stunted growth, and reduced crop yields. The disease spreads through water-borne spores, often exacerbated by rain and irrigation, and can be propagated by infected seeds and plant debris. Management strategies include using resistant varieties, crop rotation, good sanitation practices, proper irrigation management, and fungicidal applications as part of an integrated approach. Regular monitoring is crucial for early detection and effective control, helping to minimize the impact of Downy Mildew on Brassica crops.

Figure 3 Brassica Downy Mildew

● **Soft rot**

Soft Rot, caused by the bacterium Pectobacterium carotovorum, is a serious disease impacting cabbages and other Brassica crops, particularly under warm, moist conditions. This disease is marked by a soft, wet decay of plant tissues, often beginning at the base of the cabbage head or in damaged areas, leading to a slimy degradation and foul Odor. The bacteria spread via water, contaminated soil, and equipment, entering the plant through wounds or natural openings. Cabbage Soft Rot poses a significant threat to crop yield and quality, as infected cabbages are typically unmarketable, resulting in substantial economic losses for farmers. The disease can also progress during post-harvest storage, further diminishing the crop's market value.

Effective management involves preventive measures such as crop rotation, proper field drainage, sanitation to avoid contamination, and careful handling to reduce plant damage and the risk of infection.

The agricultural technique of manually harvesting plants individually at various intervals is widely used in practice. The duration of the harvesting process is prolonged because workers are required to physically evaluate the size of the cauliflower head within the plant, even when it is hidden by its canopy. In the field of digital agriculture field monitoring, satellite or unmanned aerial vehicle (UAV) footage is used to monitor and track the growth and development of plants throughout the growing season [4]. Deep learning techniques are widely used in the field of data analysis. This specific analysis includes the classification of crop maturity on a large scale, along with precise predictions regarding the ripeness of the harvest, the quantity to be harvested, and the expected date of harvest readiness. Economic profitability can be attained by farmers through the application of crop trait prediction in the context of harvest. Therefore, it is crucial for the model to demonstrate reliability, thereby instilling confidence in farmers regarding its decision-making capabilities.One of the primary limitations associated with cauliflower production is its vulnerability to disease infection, despite the numerous health benefits provided by this vegetable. The cultivation of cauliflowers presents the possibility for farmers to come across bacterial and fungal infections. The presence of these infections can result in the development of different diseases, including blackleg (Leptosphaeria maculans), black rot, downy mildew (Hyaloperonospora parasitica), powdery mildew (Erysiphe cruciferarum), ring spot (Mycosphaerella brassicicola), white rust (Albugo Candida), and bacterial soft rot [5].

In the family of Cruciferous vegetables, Cauliflower, is the largest and most widely consumed group of plants all over the world. They are characterized by different levels of nutrients. However, because of their large and frequent consumption, they may become a significant source of nutrients and bioactive compounds in the daily diet. The beneficial effects of cruciferous vegetables on human health have been somewhat linked to phytochemicals. They prevent oxidative stress, induce detoxification enzymes, stimulate the immune system, decrease the risk of cancers, inhibit malignant transformation and carcinogenic mutations, as well as, reduce proliferation of cancer cells. Cruciferous vegetables contain a lot of valuable metabolites, which are effective in chemoprevention of cancer, which has been already documented by numerous studies. Due to the presence of vitamins C and E, carotenoids and antioxidant enzymes such as catalase, superoxide dismutase (SOD) and peroxidase, these vegetables are considerable source of antioxidants, and due to the presence of polyphenols and the sulfur-organic compounds exert also antimutagenic action. Moreover, these vegetables are also rich in glucosinolates, which are unstable compounds and undergo degradation into biologically active indoles and isothiocyanates under the influence of enzymes present in plant tissues- myrosynase. These substances through the induction of enzymatic systems I and II phase of xenobiotics metabolism may affect the elimination or neutralization of carcinogenic and mutagenic factors, and consequently inhibit DNA methylation and cancer development. Despite many healthy benefits upon eating of cruciferous vegetables, it has been also seen a negative impact of their certain ingredients on the human body.

Consuming cauliflower that has deteriorated or contaminated can result in significant health implications.

Allergic reactions can manifest in individuals as a range of symptoms, such as sneezing, itching, watery eyes, coughing, ear and skin infections, gastrointestinal disorders, and other related effects, following the ingestion of specific substances. The application of pesticides and insecticides on cauliflowers is intended to mitigate microbial contamination. However, it is important that these substances can have detrimental effects on human health. In addition to causing acute and chronic poisoning, the use of pesticides and insecticides also raises the risk of developing severe health conditions, including Alzheimer's disease, cancer, asthma, bronchitis, and other ailments. In addition, there has been a decline in both the volume and standard of cauliflower production within the agricultural sector [6].

The detection of cauliflower disease in agriculture using conventional techniques poses several challenges and limitations. Errors and inconsistencies frequently arise due to the dependence on subjective human visual inspection. The manual examination process is characterized by its time-consuming nature and its tendency to cause delays in disease detection, consequently contributing to the accelerated spread of infections. Many farmers, especially those in remote areas, find it financially burdensome to hire and retain agricultural experts for disease diagnosis. These approaches frequently overlook early or asymptomatic illnesses, pose scalability challenges, rely on specific environmental conditions, and lack data verification. The limited applicability of these methods stems from their significant dependence on specialized expertise. Advanced deep transfer learning systems have been developed to overcome these limitations. These systems provide automated, precise, efficient, and scalable disease detection capabilities, while also enabling continuous monitoring of crops. In order to optimize profitability and productivity in cauliflower production, it is imperative to promptly detect and identify diseases [7].

The efficacy of deep learning techniques in the detection of leaf diseases in cauliflower and other crops is on the rise. By utilizing neural networks, these techniques possess the ability to effectively identify diseases by analysing and learning from vast collections of images. Convolutional neural networks (CNNs) are widely recognized as a fundamental deep learning approach for the detection and classification of cauliflower leaf disease. Convolutional Neural Networks (CNNs) are commonly recognized as the predominant deep learning models utilized for image recognition applications. The automatic and adaptive learning of spatial hierarchies of features is achieved from the input photos. Convolutional Neural Networks (CNNs) are equipped with the ability to recognize and detect patterns and anomalies that exist on leaf surfaces. The presence of patterns and irregularities can function as reliable indicators for identifying potential diseases, particularly when it comes to diagnosing cauliflower disease. Transfer Learning: This involves using a pre-trained model (often trained on a large dataset like ImageNet) and fine-tuning it for the specific task of cauliflower disease detection. Transfer learning is beneficial when the available dataset for the specific task is relatively small, as it leverages learned features from a larger dataset. Image Augmentation: To increase the diversity of the training dataset and prevent overfitting, image augmentation techniques like rotation, flipping, scaling, and cropping are used. This helps the model learn to recognize diseases under various conditions and orientations. Recurrent Neural Networks (RNNs): Though less common than CNNs for image analysis, RNNs, especially Long Short-Term Memory (LSTM) networks, can be used for sequential data processing, which might be relevant in tracking disease progression over time. Semantic Segmentation:

This approach involves deep learning models that can classify each pixel in an image into various categories (e.g., healthy tissue, diseased tissue, background). This is useful for understanding the extent and severity of disease spread on a leaf. Generative Adversarial Networks (GANs): While more experimental, GANs can be used for generating synthetic images of diseased leaves, which can augment training datasets, particularly when certain types of disease images are scarce. Object Detection Models: Models like YOLO (You Only Look Once) or SSD (Single Shot Multibox Detector) can be adapted to not only detect the presence of disease but also localize it on the cauliflower leaves [8].

2   RELATED WORK

Convolutional neural networks (CNNs), which are a type of deep learning technique, have become increasingly popular in various agricultural applications. The utilization of CNN-based techniques has gathered significant attention and demonstrated potential in various agricultural applications. These applications encompass the identification of plant diseases, weeds, and crop pests [9]. Convolutional Neural Networks (CNNs) have become the preferred architecture for various image and video research tasks, particularly in the field of computer vision. These tasks include important applications such as image recognition, object detection, and segmentation.

In recent times, deep learning architectures have demonstrated potential in the areas of object segmentation, classification, and identification. Convolutional neural network (CNN) techniques have emerged as the predominant methods for addressing deep learning challenges [10]. In their study, in [11] focused on a collection of representative banana plant models sourced from various regions in Southern India and Africa. This collection included a single class of healthy plants and seventeen distinct classes of plants affected by different conditions. A 90% accuracy rate was attained through the utilization of various CNN architectures, including ResNet50, InceptionV2, and MobileNetV1. In [12] they have utilized images from the Plant Village collection to obtain samples of peach plant leaves exhibiting both healthy and diseased conditions for their research study. A 99% accuracy rate was achieved using the AlexNet concept through a process of trial and error. The utilization of images featuring three labels for diseased plant leaves and one label for healthy plant leaves has been observed in various studies. These studies have employed photos sourced from the Plant village dataset to identify disease classes in maize, grape, and soybean   plants. The reported accuracy rates in these studies ranged from 97% to 99% [13].

The VGG CNN architecture was utilized by [14] to analyse a dataset consisting of nine distinct categories of tomato diseases. The study reported a 90% accuracy rate. Furthermore, in their experiments, [15] were able to achieve a high accuracy rate of 97%. This was accomplished by utilizing a dataset that consisted of six classes of ill wheat plants and one class of healthy wheat plants. The modified deep convolutional neural network-based technique for tea leaf sickness detection. Upon completion of the studies, the average accuracy achieved was 92%. In their study, they employed a set of sample images comprising fourteen distinct plant species, which collectively encompassed a total of seventy-nine unique diseases. Multiple accuracy scores were obtained through the utilization of the GoogLeNet architecture, all of which exceeded 75%. In their study,

[16] conducted research on various plant species and utilized the VGG architecture to achieve an accuracy rate of 81%. It is worth noting that alternative designs were able to achieve even higher accuracy rates, reaching up to 99%.

Furthermore, a dataset consisting of 38 distinct classes that encompassed 14 diverse plant species was employed, despite their broad focus on various plant types. The accuracy rate achieved was 96%. In a study conducted [17], a VGG architecture was employed along with a dataset comprising 25 distinct plant species. The results demonstrated an impressive accuracy rate of 98%, they utilized image pre-processing techniques in conjunction with their CNN model to achieve an accuracy rate of 97.62% when combining the current research on apple leaf diseases. On the other hand, certain studies have utilized conventional machine learning techniques, which have a collective substantial amount of knowledge regarding the identification of plant diseases.

However, the procedures that they are limited to include image segmentation, feature extraction, and pattern recognition. The fundamental CNN architectures, such as AlexNet, VGGNet, GoogLeNet, DenseNet , and ResNet , have been extensively utilized for plant disease classification. However, these architectures suffer from certain limitations, such as a high parameter count and slow processing speed. Deep learning methods excel at capturing and representing high-level and low-level features. However, their ability to accurately identify local spatial attributes is limited [18]. It is recommended to integrate both the handcrafted and deep features to effectively capture the distinctive attributes of the plant leaf images.

To effectively implement this strategy, it is crucial to minimize the distances between samples from similar classes and maximize the distances between samples from dissimilar classes. Deep learning approaches demonstrate significant potential; however, their effectiveness relies on the assumption that the training and testing data originate from identical distributions. This assumption may not hold in practical situations [19]. Domain adaptation and transfer learning techniques are commonly employed to tackle the challenges arising from distribution shifts. Transfer learning is a technique that aims to address the differences in distribution between the training and testing data.

On the other hand, domain adaptation utilizes previously learned models to improve the performance in new domains. Domain adaptation is a technique that enhances the performance of a model on a target domain, even when the target domain exhibits different data distributions. This is achieved by leveraging knowledge obtained from a similar source domain. Domain adaptation is a technique that is employed in situations where obtaining labelled data in the target domain is either excessively costly or unfeasible [24], [25]. Transfer learning is currently being utilized with great success in various fields such as computer vision, robotics, natural language processing, precision agriculture, and computer vision [20]. Furthermore, it is garnering significant interest. Promising results have been observed when this technology is implemented to address real-world challenges across diverse sectors such as industrial, healthcare, security and surveillance, agriculture, automotive, and finance. Transfer learning is widely acknowledged as an essential strategy for addressing real-world challenges. Within the domain of transfer learning, there exist two distinct approaches: zero-shot learning and few-shot learning.

Embedding-based approaches commonly utilize common embeddings to facilitate the transfer of knowledge across tasks [21]. The model's ability to learn a shared embedding space during pre-training enables it to effectively map new inputs to that space. As a result, it can achieve high performance on tasks that it has not encountered previously. Methods that are dependent on models The generative model acquired through pre-training is capable of representing the data distribution in zero-shot learning. The generative model can be employed to generate examples for untested classes. Generative adversarial networks (GANs) have the capability to be utilized in the context of zero-shot learning, enabling the generation of samples for classes that have not been encountered previously.

The generator undergoes pre-training on seen classes and has the capability to generate samples for unseen classes during testing. The Deep Domain Adaptation technique was introduced by. This technique leverages privileged data from dual-domain pairings that are not related to the current task, and introduces a novel approach for zero-shot domain adaptation (ZSDA). The proposed approach utilizes adversarial learning to simultaneously train domain-invariant semantic features and task-invariant domain characteristics. The objective of this approach is to obtain domain characteristics that are independent of the specific task and are affected by domain shift. In the context of the day-night domain, [17] conducted an investigation on the zero shot scenario. This investigation utilized prior data obtained from a physics-based reflection model.

In certain real-world situations, it is possible to come across a task that requires minimal or no availability of labelled data. In scenarios of this nature, the construction of a deep learning model presents itself as a formidable task, if not an insurmountable one. In order to mitigate this issue, researchers have devised transfer learning strategies. Transfer learning is a technique that leverages pre-trained model architectures, particularly Convolutional Neural Network (CNN) architectures such as ImageNet, which have been extensively trained on labelled data. The efficacy of this approach is heavily contingent upon selecting a suitable Convolutional Neural Network (CNN) model.

[18-19] The determination of the appropriate course of action will be contingent upon the precise requirements of the current undertaking. In scenarios where processing power is a limiting factor for model evaluation, employing a CNN model with a reduced number of parameters, such as MobileNetV2, can offer distinct advantages. In cases where sufficient resources are accessible, opting for a higher parameter model like VGG19 can yield advantageous outcomes.

Table 1 survey table for plant disease identification

Study	Plant Type	Method Used	Dataset Used	Achieved Accuracy	Research Gap
G.Geethara mani et.al	Banana	CNN (ResNet50, InceptionV2, MobileNetV1)	Samples from Africa & Southern India	90%	Limited geographic diversity in the dataset.
G.Geethara mani et.al	Peach	CNN (AlexNet)	PlantVillage dataset	99%	Overreliance on a single dataset may reduce model robustness.
M G Selvaraj et.al	Maize	CNN	PlantVillage dataset	97-99%	Limited disease classes; potential lack of real-world variability.
R.A.Priyad harshini et.al	Grape	CNN	PlantVillage dataset	97-99%	Same as above.
M.Ji et.al	Soybean	CNN	PlantVillage dataset	97-99%	Same as above.
E.C.Too et.al	Tomato	CNN (VGG)	Nine classes of tomato diseases	90%	Limited to a single crop type, may not generalize well.
S.Singh et.al	Wheat	CNN	Six diseased and one healthy class	97%	Narrow focus on a small number of disease classes.
A.Kamilari s et.al	Tea	Modified CNN	-	92%	Absence of detailed dataset description, unclear generalizability.
G.S.Sujaw at et.al	Various	CNN (GoogLeNet)	79 diseases in 14 plant types	>75%	Inconsistent accuracy rates, need for more uniform results.

The present study employed a sample size of 100 wild accessions of Arabidopsis to investigate the range of diseases caused by Cauliflower mosaic virus (CaMV). CaMV is a type of cauli virus that belongs to the Caulimoviridae family and is characterized by its double-stranded DNA structure. Our investigation focused on the vegetative stage and rosette tissue of the plant due to the requirement of significant vernalization for the flowering process in many accessions. The technique introduced by [22] utilizes Multi-class Support Vector Machines to achieve accurate detection and classification of cucumber leaf diseases. The implementation of the segmentation and feature extraction processes involved the utilization of k-means clustering and image processing techniques. The diagnostic approach used to identify cucumber infections involves the application of a Convolutional Neural Network (CNN) model. A grand total of 48,311 photographs were employed, yielding an average accuracy rate of 95.5%.

A method for the identification of fruit illness was presented by [14] in a similar manner. The approach employs a random forest classifier. A total of seventy photographs were employed during the implementation process.   A novel technique has been proposed for the classification of vegetables and weeds. The support vector machine (SVM) classifier demonstrated a classification accuracy of 90%. In their study, Wahab et al. utilized Support Vector Machines (SVM) as a methodology to develop a systematic approach for the identification and categorization of challis leaf infections. The classification task was executed utilizing Support Vector Machines (SVM), whereas the segmentation process was conducted employing k-means clustering. The accuracy was determined to be 57.1 percent.

 The detection and classification of diseases affecting roses [15] is facilitated by the utilization of the MobileNet model in a specific methodology. The study incorporated a dataset comprising 2000 photographs, which were categorized into four distinct classes of disorders. The analysis yielded an average accuracy rate of 95.63%. have developed an automated technique for the identification of potato diseases. The researchers utilized a segmentation technique in combination with a Multiclass Support Vector Machine (MSVM) to classify potato diseases. The collection comprised a total of 300 photographs, which were systematically classified into three distinct groups according to the diseases they portrayed. The impact on these diseases varied across different aspects. The categorization accuracy was determined to be 95%.

 The classification of potato illnesses was performed using a Convolutional Neural Network (CNN) with a sliding window approach. A trained model was developed by utilizing a dataset consisting of 400 photographs, which yielded an accuracy range spanning from 80% to 90%. The identification of paddy illnesses was performed utilizing a designated technique [20]. The disorders were classified using the KNN classifier, which yielded an accuracy of 75.61%. A functional prototype has been developed by [16] that employs imaging techniques for the diagnosis of diseases impacting rice crops. The researchers utilized two techniques, specifically OSTU and Threshold, in order to achieve a precision rate of 94.7%. In addition, a prototype approach for the classification and detection of rice diseases was presented [17].

 The three rice plant diseases were addressed by the individuals. The disease was segmented using the K-means clustering algorithm and then classified using the MSVM (Multi-class Support Vector Machine) model. The training dataset achieved an accuracy rate of 93.33%, while the testing dataset achieved an accuracy rate of 73.33%.

This paper introduces a methodology that employs Convolutional Neural Networks (CNNs) to identify different diseases that impact rice crops. The trained model was constructed using a dataset comprising 500 photographs. A level of accuracy of 95.48% was achieved. In addition, researchers [22] have developed a deep learning-based algorithm specifically designed for the identification of plant diseases. A total of 87,848 photos were used to train the CNN models, which achieved an impressive accuracy rate of 99.53%. A novel technique has been developed for the identification of papaya disease. The segmentation process utilized the K-means technique, while the illness classification process employed Support Vector Machines (SVM). After conducting a comprehensive series of 500 papaya photo practice sessions, an accuracy rate of 90% was achieved.

A study was conducted by [23] to examine the efficacy of low UV-C or gamma radiation treatment on fresh-cut cauliflower samples. In addition, the samples underwent a light treatment using natural antibacterial formulations. The primary goal of this treatment was to reduce the presence of food-borne microorganisms and prolong the shelf life of the cauliflower samples. The objective of the process is to mitigate the adverse impact on the microbiological safety of vegetables by avoiding any potential interaction between the combined treatment methods. [24] The optimal method for preventing the growth of harmful bacteria, yeasts, and molds on recently-cut cauliflower is to employ a dual strategy consisting of gamma irradiation at a dosage of 1 kGy and the application of antimicrobial formulations via spraying. The formulations should be composed of citrus extract, lactic acid, and essential oils obtained from either oregano or lemongrass.

The objective of this study was to analyze the spectrum of diseases observed in 100 wild accessions of Arabidopsis thaliana following infection with Cauliflower mosaic virus (CaMV). CaMV is a member of the Caulimoviridae family, characterized by its double-stranded DNA structure. The Cauliflower Mosaic Virus (CaMV) exhibits the capability to infect indigenous populations of Arabidopsis, a plant species. However, the infection caused by this pathogen is restricted to plants that fall under the Brassicaceae family. This family includes a range of vegetables such as mustard, broccoli, and cabbage. The CaMV viral translational transactivator protein P6 facilitates the virus in exerting its influence on the host by generating extensive cytoplasmic viral replication foci and inducing a significant increase in overall translation. The distinctive characteristics of CaMV suggest the possible existence of a host variable network that could influence the development of CaMV disease. It is noteworthy to mention that infection by CaMV has been demonstrated to induce varying degrees of disease severity in wild accessions of Arabidopsis under conditions of water deficiency. The suitability of CaMV as a virus for conducting a genome-wide association investigation in Arabidopsis has been attributed to this particular characteristic [25].

Table 2 survey table for plant disease classification

Study	Plant Type	Classification Method	Dataset Size	Achieved Accuracy	Research Gap
M.G.Selvar aj et.al	Cucumber	Multi-class SVM	-	-	Limited information on dataset size and diversity.
R.A.Priyad harshini et.al	Fruit	Random Forest Classifier	70 images	-	Small dataset size may not be representative.
M.Ji et.al	Rose	MobileNet	2000 images of four disease classes	95.63%	Limited disease classes, potential bias towards certain conditions.
G.S.Sujawa t et.al	Potato	MSVM	300 images of 3 disease types	95%	Small dataset, lack of variation in disease types.
M.Ji et.al	Paddy	KNN	-	75.61%	Lower accuracy, potential issues with model complexity or dataset.
E.C.Too et.al	Paddy	Imaging Techniques (OSTU, Threshold)	-	94.7%	Limited information on methodology and dataset diversity.
S.Singh et.al	Rice	MSVM	-	93.33% training, 73.33% testing	Discrepancy between training and testing accuracy, indicating possible overfitting.

A.Kamilaris et.al	Various	CNN	87,848 images	99.53%	Large dataset size, but potential lack of focus on specific diseases.

2.1 Observations

- **Preference for CNNs**: Convolutional Neural Networks are the most commonly employed deep learning architecture for both disease identification and classification in plants, indicating their effectiveness in image-based analysis.

- **High Accuracy:** Many studies have reported high accuracy rates (often above 90%), demonstrating the potential of deep learning in accurate disease detection.

- **Dataset Reliance**: Several studies rely heavily on the Plant Village dataset, which could limit the diversity and variability of the training data, potentially affecting the models' ability to generalize to real-world conditions.

- **Geographical Limitations**: Some studies have datasets concentrated in specific geographic areas, which may not represent global plant disease variations adequately.

- **Transfer Learning Utilization**: Transfer learning is increasingly being adopted, as it allows leveraging pre-trained models to improve performance, especially when dataset sizes are limited or specific to a certain crop.

- **Gap in Disease Classes**: Some studies focus on a limited number of disease classes, which may not capture the full spectrum of potential plant diseases, indicating a need for more comprehensive disease coverage in future research.

- **Varied Architectural Choices**: Different CNN architectures like ResNet, InceptionV2, MobileNet, and VGG are being explored, with each offering unique advantages, suggesting no single "one-size-fits-all" model.

- **Emergence of Advanced Techniques**: Newer approaches like zero-shot learning and domain adaptation are being explored to address challenges in data availability and diversity, indicating a trend towards more adaptive and robust deep learning models in agriculture.

CONCLUSION:

The survey on deep learning applications for cauliflower leaf disease identification highlights a pivotal shift from traditional expert-based diagnostics to more advanced, automated methods using Convolutional Neural Networks (CNNs). This approach has proven effective, as evidenced by high accuracy rates achieved in studies utilizing datasets like PlantVillage. However, challenges such as reliance on specific datasets and limited geographical representation remain, suggesting a need for more diverse and comprehensive disease coverage.

Emerging techniques like transfer learning are being adopted to address these issues, enhancing model adaptability to real-world conditions. The integration of deep learning with traditional feature extraction methods offers a promising direction for more accurate and efficient disease identification in cauliflower crops. This progression in agricultural technology is crucial for improving disease management, crop yield, and overall food security, underscoring the importance of ongoing research in this domain to cater specifically to the complexities of cauliflower leaf disease detection.

References:

[1] Dan, A., Jain, R., Dwivedi, R. K. & Kumar, A. Evaluation of socio-economic conditions of cauliflower (Brassica oleracea) growers in Chaka block of Allahabad district Uttar Pradesh. J. Pharmacogn. Phytochem. *9*(5), 148–151 (2020).

[2] Kumar, Y., Singh, R., Moudgil, M. R. & Kamini, F. A systematic review of different categories of plant disease detection using deep learning-based approaches. Arch. Comput. Methods Eng. *30*, 1–23 (2023).

[3] Maria, S. K. et al. Cauliflower disease recognition using machine learning and transfer learning. In Smart Systems: Innovations in Computing 359–375 (Springer, 2022).

[4] Abdul Malek, M., Reya, S. S., Zahan, N., Hasan, Z. & Uddin, M. S. Deep learning-based cauliflower disease classification. In Computer Vision and Machine Learning in Agriculture, Volume 2 171–186 (Springer, 2022).

[5] Saad, I. H., Islam, M. M., Himel, I. K. & Mia, M. J. An automated approach for eggplant disease recognition using transfer learning. Bull. Electr. Eng. Inf. *11*(5), 2789–2798 (2022).

[6] Chowdhury, M. E. et al. Automatic and reliable leaf disease detection using deep learning techniques. AgriEngineering *3*(2), 294–312 (2021).

[7] Sivakumar, P., Mohan, N. S. R., Kavya, P. & Teja, P. V. S. Leaf disease identification: Enhanced cotton leaf disease identification using deep CNN Models. In 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT) 22–26 (IEEE, 2021).

[8] Chohan, M., Khan, A., Chohan, R., Katpar, S. H. & Mahar, M. S. Plant disease detection using deep learning. Int. J. Recent Technol. Eng. *9*(1), 909–914 (2020).

[9] S. K. Maria, S. S. Taki., M. J. Mia, A. A. Biswas., A. Majumder, F. Hasan, "Cauliflower Disease Recognition Using Machine Learning and Transfer Learning," In: Somani A.K., Mundra A., Doss R., Bhattacharya S. (eds) Smart Systems: Innovations in Computing. Smart Innovation, Systems and Technologies, vol 235, 2021.

[10] Verhoeven, A., Kloth, K.J., Kupczok, A., Oymans, G.H., Damen, J., Rijnsburger, K., Jiang, Z., Deelen, C., Sasidharan, R. & van Zanten, M. (2022). Arabidopsis latent virus 1, a comovirus widely spread in Arabidopsis thaliana collections. New Phytologist.

[11] G. Geetharamani and G. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," Comput. Electr. Eng., vol. 76, pp. 323–338, Jun. 2019. [4] E. M. F. El Houby, "A survey on applying machine learning techniques for management of diseases," J. Appl. Biomed., vol. 16, no. 3, pp. 165–174, Aug. 2018.

[12] G. Geetharamani and G. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," Comput. Electr. Eng., vol. 76, pp. 323–338, Jun. 2019.

[4] E. M. F. El Houby, ''A survey on applying machine learning techniques for management of diseases,'' J. Appl. Biomed., vol. 16, no. 3, pp. 165–174, Aug. 2018.

[13] M. G. Selvaraj, A. Vergara, H. Ruiz, N. Safari, S. Elayabalan, W. Ocimati, and G. Blomme, ''AI-powered banana diseases and pest detection,'' Plant Methods, vol. 15, no. 1, p. 92, Dec. 2019.

[14] R. A. Priyadharshini, S. Arivazhagan, M. Arun, and A. Mirnalini, ''Maize leaf disease classification using deep convolutional neural networks,'' Neural Comput. Appl., vol. 31, no. 12, pp. 8887–8895, Dec. 2019.

[15] M. Ji, L. Zhang, and Q. Wu, ''Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks,'' Inf. Process. Agricult., vol. 7, no. 3, pp. 418–426, Sep. 2020.

[16] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, ''A comparative study of fine-tuning deep learning models for plant disease identification,'' Comput. Electron. Agricult., vol. 161, pp. 272–279, Jun. 2019.

[17] S. Singh, S. Gupta, A. Tanta, and R. Gupta, ''Extraction of multiple diseases in apple leaf using machine learning,'' Int. J. Image Graph., vol. 22, no. 3, May 2022, Art. no. 2140009.

[18] A. Kamilaris and F. X. Prenafeta-Boldú, ''Deep learning in agriculture: A survey,'' Comput. Electron. Agricult., vol. 147, pp. 70–90, Apr. 2018.

[19] G. S. Sujawat, ''Application of artificial intelligence in detection of diseases in plants: A survey,'' Turkish J. Comput. Math. Educ. (TURCOMAT), vol. 12, no. 3, pp. 3301–3305, Apr. 2021, doi: 10.17762/turcomat.v12i3.1581.

[20] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, ''Using deep transfer learning for image-based plant disease identification,'' Comput. Electron. Agricult., vol. 173, Jun. 2020, Art. no. 105393, doi: 10.1016/j.compag.2020.105393.

[21] A. Malik, G. Vaidya, V. Jagota, S. Eswaran, A. Sirohi, I. Batra, M. Rakhra, and E. Asenso, ''Design and evaluation of a hybrid technique for detecting sunflower leaf disease using deep learning approach,'' J. Food Qual., vol. 2022, pp. 1–12, Apr. 2022, doi: 10.1155/2022/9211700.

[22] U. Sara, A. Rajbongshi, R. Shakil, B. Akter, S. Sazzad, and M. S. Uddin, ''An extensive sunflower dataset representation for successful identification and classification of sunflower diseases,'' Data Brief, vol. 42, Jun. 2022, Art. no. 108043, doi: 10.1016/j.dib.2022.108043.

[23] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, ''Deep learning for image-based cassava disease detection,'' Frontiers Plant Sci., vol. 8, p. 1852, Oct. 2017, doi: 10.3389/fpls.2017.01852.

[24] J. Xue, L. Huang, B. Mu, K. Wang, Z. Li, H. Sun, H. Zhao, and Z. Li, ''Detection of rotten fresh-cut cauliflowers based on machine vision technology and watershed segmentation method,'' Amer. J. Biochem. Biotechnol., vol. 18, no. 2, pp. 155–167, Feb. 2022.

[25] T. Y. Orin, M. U. Mojumdar, S. M. T. Siddiquee and N. R. Chakraborty, "Cauliflower Leaf Disease Detection Using Computerized Techniques," 2021 IEEE 6th International Conference on Computing, Communication and Automation (ICCCA), Arad, Romania, 2021, pp. 730-733, doi: 10.1109/ICCCA52192.2021.9666437.

[26] Joanna Kapusta-Duch, Aneta Kopec, Ewa Piątkowska, Barbara Borczak and Teresa Leszczynska, "The beneficial effects of Brassica vegetables on human health, " Rocz Panstw Zakl Hig 2012, 63, Nr 4, 389 - 395.