Heliyon 10 (2024) e29583

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Review article

Deep learning and content-based filtering techniques for improving plant disease identification and treatment recommendations: A comprehensive review

Folasade Olubusola Isinkaye [a,*], Michael Olusoji Olusanya [a,**], Pramod Kumar Singh [b]

[a] Department of Computer Science and Information Technology, Sol Plaatje University Kimberley, 8301, South Africa
[b] Department of Computer Science and Engineering, ABV-Indian Institute of Information Technology and Management Gwalior, Gwalior, 474015, MP, India

ARTICLE INFO

Keywords:
Convolutional neural networks
Food security
Plant disease
Recommender system
Sustainable agriculture

ABSTRACT

The importance of identifying plant diseases has risen recently due to the adverse effect they have on agricultutal production. Plant diseases have been a big concern in agriculture, as they affect crop production, and constitute a major threat to global food security. In the domain of modern agriculture, effective plant disease management is vital to ensure healthy crop yields and sustainable practices. Traditional means of identifying plant disease are faced with lots of challenges and the need for better and efficient detection methods cannot be overemphazised. The emergence of advanced technologies, particularly deep learning and content-based filtering techniques, if integrated together can changed the way plant diseases are identified and treated. Such as speedy and correct identification of plant diseases and efficient treatment recommendations which are keys for sustainable food production. In this work, We try to investigate the current state of research, identified gaps and limitations in knowledge, and suggests future directions for researchers, experts and farmers that could help to provide better ways of mitigating plant disease problems.

1. Introduction

Plant diseases have adverse effects on crop yields, quality, and economic stability in agricultural systems worldwide [1]. They cause crop production losses that vary from 20 % to 40 % annually [2]. Approximately 83 % of plant diseases are caused by fungus, 9 % by viruses and phytoplasmas, and more than 7 % by bacteria according to Ref. [3]. Also, about 13 % of the inhabitants of developing nations are affected by malnutrition [4] due to inadequate availability of food production caused by the activities of these plant diseases on farm yields. Hlophe-Ginindza and Mpandeli [5] suggested that the existing food production level must increase by at least 70 % in order to guarantee sustainable food production. Therefore, sustainable food production demands immediate proactive management strategies to identify diseases accurately and provide appropriate treatment recommendations. Recently, the advent of deep learning and content-based filtering techniques have revolutionized the field of plant disease management [6]. Deep learning

* Corresponding author.
** Corresponding author.
E-mail addresses: folasade.isinkaye@spu.ac.za (F.O. Isinkaye), michael.olusanya@spu.ac.za (M.O. Olusanya), pksingh@iiitm.ac.in (P.K. Singh).

https://doi.org/10.1016/j.heliyon.2024.e29583
Received 14 September 2023; Received in revised form 30 March 2024; Accepted 10 April 2024
Available online 16 April 2024
2405-8440/

techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated outstanding performance in several image analysis tasks [7]. By training on large-scale datasets of plant images, these models can learn to accurately identify different types of diseases based on visual patterns and characteristics. On the other hand, content-based filtering techniques has the potential to leverage disease characteristics, treatment strategies, and expert knowledge to provide personalized treatment recommendations. These techniques analyze the specific symptoms, environmental conditions, and disease phases to tailor the treatment approach to individual plants. By considering a wide range of factors, content-based filtering methods can provide more precise and effective treatment recommendations [8]. Also, the integration of deep learning and content-based filtering techniques can provide a comprehensive approach to plant disease identification and treatment recommendation. By combining the strengths of both techniques, it becomes possible to not only accurately detect diseases but also provide personalized treatment support based on the unique attributes of each plant.

Traditional approaches used to identify diseases in plants depend on visual assessment and professional knowledge. Nevertheless, these approaches can waste time, biased and be impreciesed. Also, general treatment recommendations may not consider the features of plants and diseases leading to less optimal outputs. This review explores the use of deep learning and content-based filtering techniques for plant disease identification and treatment recommendations. By examining the advantages, limitations and future research directions in this field we, aim to contribute towards the development of efficient and sustainable practices for managing plant diseases.

Specifically, the study has been organized to answer the following important research questions for plant disease identification and treatment recommendation.

1. Which traditional methodologies and current deep learning techniques are exploited to identify plant diseases in the literature?
2. What are the possible content-based filtering techniques that can be used to provide treatment recommendations for plant diseases?
3. How can deep learning techniques be combined with content-based filtering techniques to provide tailored and efficient treatment recommendations for plant diseases?
4. What datasets are available for deep learning model training and validation in the context of identifying plant diseases?
5. What performance measures are applied to evaluate the accuracy and efficiency of the deep learning and content-based filtering techniques identified?
6. What are the gaps and areas of improvement in current research, and what are the likely future directions for advancing this field?

2. Traditional methodologies and current deep learning techniques exploited to identify plant diseases in literature

2.1. Traditional methodologies for plant disease identification

Traditional methodologies of plant disease identification describes the conventional methods used to diagnose and identify plant diseases based on visual symptoms, signs, and field observations [9]. These methods have been practiced for many years and are often employed by farmers, gardeners, and field workers. This section discusses traditional methods used for plant disease identification, such as visual symptom observation, laboratory-based test, and expert knowledge-driven approaches (Fig. 1). It discusses their strengths, limitations, and applications.

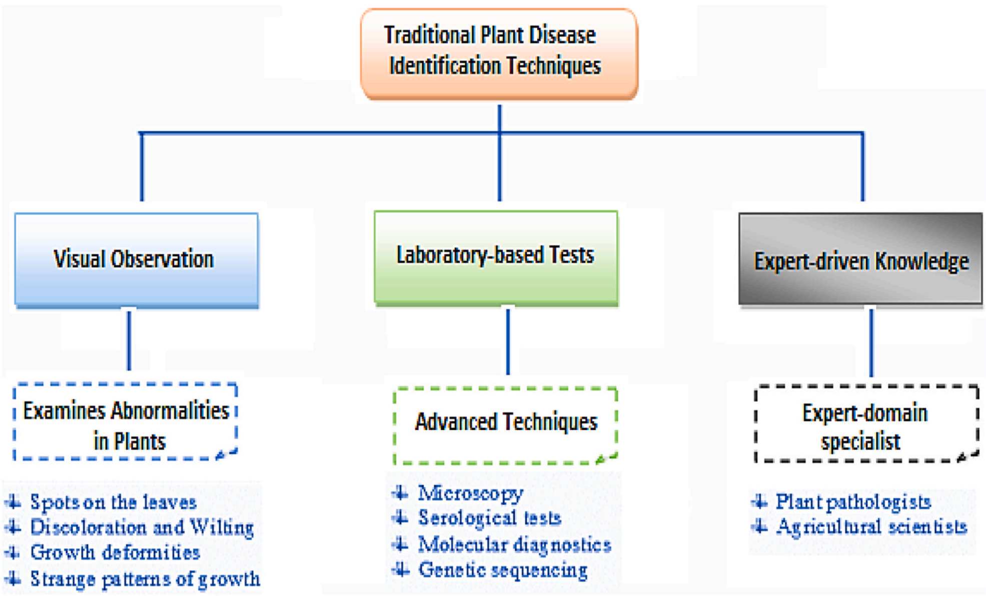

Fig. 1. Traditional techniques of plant disease identification.

Visual observation: One way to detect diseases in plants is through visual observation of disease symptoms. This method involves examining the plants for any changes or abnormalities that may indicate the presence of a disease [10]. These symptoms can include spots on the leaves, discoloration, wilting, growth, deformities or strange patterns of development. When these observed symptoms are contrasted with reference information or personal knowledge one can easily identified the actual disease. Visual symptom assessment is a straightforward method that can be easily used by farmers and field workers without requiring equipment or extensive training [11]. It also allows for feedback if disease symptoms are detected. Likewise, visual symptom observation can be done right in the field enabling on site diagnosis of diseases [12].

Visual observation of symptoms can have problems such as, subjectivity, which depends on human interpretation and judgment that can introduce bias and variability into the diagnosis [13]. It also lacks the precision and accuracy provided by more advanced diagnostic techniques [14]. Some diseases may not show visible symptoms until later stages of infection, leading to delayed or missed detection. Limited information about the presence of symptoms may sometimes not offer comprehensive details about the underlying cause, such as the specific diseases or environmental factors contributing to the disease. Also, while visual symptom observation can provide valuable initial indications, it often requires additional verification through laboratory assessment or qualified experts.

Laboratory-based tests: These involves the use of advanced techniques such as microscopy, serological tests, molecular diagnostics, and genetic sequencing to detect plant diseases [15]. These techniques provide highly accurate identification of plant diseases with the ability to differentiate between closely related species and strains [16]. Laboratory techniques can detect plant pathogens at early stages of infection, even before visible symptoms appear. Also, laboratory methods often provide quantitative data on the severity of infection, pathogen load, or disease progression [17].

Despite the successes recorded by this approach, they require specialized equipment, reagents, and trained personnel, making them more expensive than traditional field-based techniques. These methods usually involve sample preparation, testing, and analysis, which can be time-consuming [18]. Delay may limit the immediate decision-making required for disease management. Proper interpretation of laboratory results requires trained personnel with expertise in plant pathology and diagnostic techniques. Also, laboratory-based methods rely on collecting and properly handling plant samples. Improper sampling or sample deterioration during transit can compromise the reliability of the results.

Expert knowledge-driven approaches: They exploit the expertise of domain specialists, such as plant pathologists or agricultural scientists, to develop robust and accurate identification systems [19,20]. Expert knowledge-driven approaches frequently depend on explicit rules or decision trees, making the reasoning behind the diagnosis transparent and interpretable [21]. This is particularly important in agricultural settings, where farmers and agricultural experts need to understand the reason behind disease identification. Expert systems can be used to support either a sequence of tactical decisions or single decisions regarding design, selection, interpretation, prediction, and diagnosis applied to agricultural problems [22].

The disadvantages of expert knowledge-driven approaches to plant disease identification include, limited scalability, as the approaches rely on the expertise of specialists [22], which may not be readily available in all locations or for all plant diseases. Scaling up

Table 1
Summary of traditional techniques of plant disease identification.

S/N	Plant disease identification techniques	Style of inspection	Strength of disease identification technique	Limitation of the disease identification technique	Improvement strategy
1	Visual symptom observation [10]	They look for abnormalities that indicate the presence of disease such as leaf spots, discoloration, wilting, stunting, deformities, or unusual growth patterns.	Simple & straightforward, no special equipment needed, no special skills required, instant feedback, cost-effective and early detection.	Bias and variability in diagnosis, overlapping symptoms issues, limited accuracy, delay in symptom manifestation can lead to missed detection of disease.	Diagnosis methods such as pathogen isolation and molecular techniques can be used to validate initial observations.
2	Laboratory-based test [15]	They use advanced techniques such as microscopy, serological tests, molecular diagnostics, and genetic sequencing to detect plant diseases.	Highly accurate, can detect plant pathogens at early stages of infection, provision of quantitative data on the severity of infection, detailed pathogen characterization.	They are expensive as they require specialized equipment, reagents, and trained personnel, time-consuming, not good for immediate decision-making, cannot be used in remote or economically disadvantaged areas where there are no well-equip laboratories.	Establish standardization of procedures for consistency and comparability of results, implement automation technologies, such as robotics, provide training programs for laboratory personnel, implement robust laboratory information management systems
3	Expert knowledge-driven approaches [19]	They combine expert knowledge with computational techniques to analyze plant disease symptoms for accurate diagnosis.	Accurate diagnosis, transparent and interpretable diagnosis, they can be adapted and updated easily, limited data requirements.	Limited scalability, can be influenced by the subjective judgments and biases of the specialists, difficulty in capturing complex relationships in formal knowledge representation, inability to capture knowledge gaps for emerging or less-studied diseases.	A hybrid approach that combines expert knowledge with data-driven techniques, such as machine learning can be employed.

these approaches to cover a wide range of crops, diseases, and regions can be challenging, as it requires access to experts and the ability to transfer their knowledge effectively. These approaches are influenced by the subjective judgments and biases of the specialists involved. There may also be difficulty in capturing complex relationships as plant diseases and their interactions with environmental factors can be complex and multifaceted. To overcome some of these limitations, a hybrid approach that combines expert knowledge with data-driven techniques, such as machine learning, can be beneficial. This allows for leveraging the strengths of the approaches and mitigating their respective weaknesses. Table 1 presents the summary of the findings of the traditional plant disease identification techniques considered.

Traditional approaches are still in use today despite the development of newer technologies because of their affordability, accessibility, and simplicity. Combining traditional and modern methods can help to detect diseases earlier, manage crop diseases better, and ultimately promote sustainable agriculture.

2.2. Current deep learning technique for plant leaf disease identification

Due to their potential in a variety of domains, deep learning techniques have lately made their way into numerous agricultural applications. Deep learning focuses on training artificial neural networks to learn and make predictions directly from data [23]. Deep learning has demonstrated outstanding performance in diverse domains, such as computer vision, natural language processing, speech recognition, and recommendation systems [24]. Their ability to learn hierarchical representations from raw data, couple with advances in hardware and large-scale datasets, have contributed to their significant progress in the domain of artificial intelligence. Generally, current deep learning techniques, such as Convolutional neural networks, Generative adversarial networks, Recurrent neural networks and Transformers (Fig. 2), offer powerful tools for plant disease identification. These techniques are discuss in the following sections.

2.2.1. Convolutional neural network in plant disease identification

Convolutional Neural Networks (CNNs) have emerged as a powerful tool for plant disease identification and classification [25]. By exploiting their ability to automatically learn and extract features from images, CNNs have shown an outstanding performance in diagnosing various plant diseases. The following section describes how CNN has contributed to plant diseases identification.

CNNs has powerful feature extraction potentials as they are particularly built to extract relevant features from images [26]. Using multiple layers of convolutional and pooling operations, CNNs can automatically learn and extract significant features such as color, texture, shape, and patterns that are essential for disease identification [27]. This ability to extract relevant features makes CNNs very effective for distinguishing between healthy and diseased plants [28]. CNNs have demonstrated to be very precise when it comes to detecting plant diseases [29]. They can explore large datasets of plant images and learn complex patterns and features associated with different diseases [30]. By training on a various range of healthy and diseased plant images, CNNs can classify new images with a high level of accuracy [31], enabling farmers and researchers to identify diseases early and take appropriate actions.

CNN has three categories of layers which are, convolutional layers, pooling layers, and fully connected layers. For the convolutional layer, the input plant image passes through filters to produce some feature maps. In the pooling layer, the dimension of each of the feature maps is reduced to keep the number of weights small. The fully connected layer helps to transform a two-dimensional feature map into a one-dimensional vector for final classification [32]. For plant leave disease identification using CNN, activation map analysis is utilized in determining the detection quality. The maps highlight the areas of a plant image that are crucial to the model's ability to make decisions. An analysis of the detection can be made based on the network's learnt features and patterns, high activation in particular regions indicate the possible presence of disease. CNNs have shown incredible performance in plant disease identification due to their ability to capture spatial patterns in images effectively. The technique of CNN for plant disease identification is shown in Fig. 3.

CNN models contribute to the development of accurate, efficient, and deployable systems for identifying plant diseases. Several

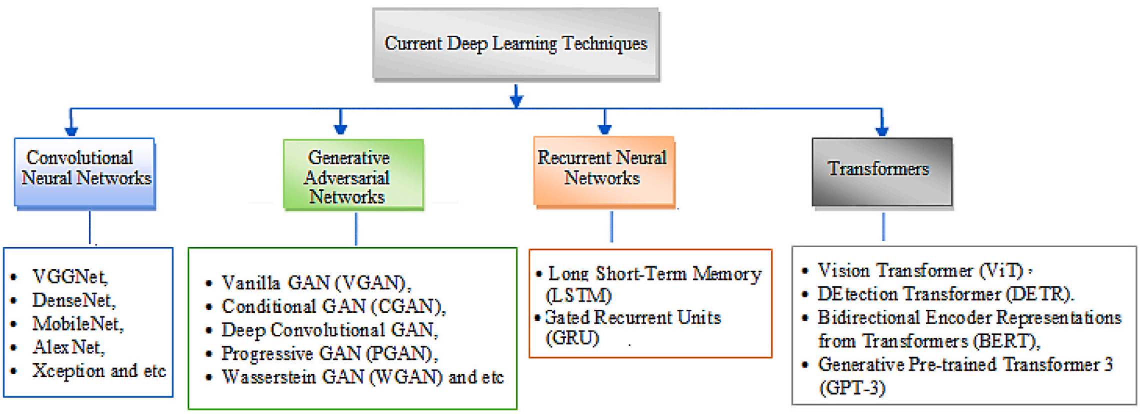

Fig. 2. Current deep learning techniques for plant disease identification.

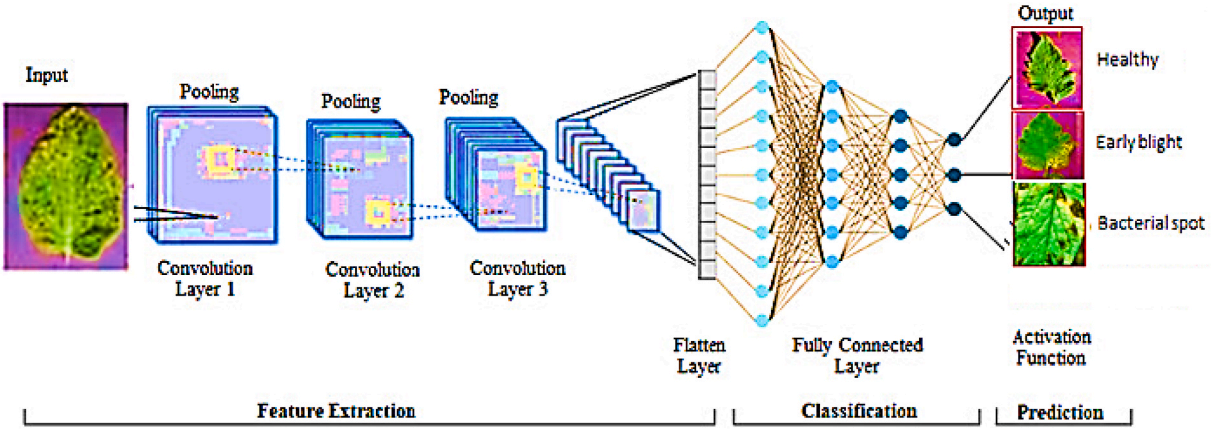

Fig. 3. Convolutional neural networks technique for plant disease identification.

CNN models such as AlexNet, MobileNet, ResNet, VGGNet, Inception, DenseNet, and EfficientNet use transfer learning approach which have had a significant impact on plant disease identification. They facilitate the use of pre-trained models on new tasks with limited training data [33]. This is specifically useful in plant disease identification, where gathering and labeling of huge datasets can be difficult. CNNs can rapidly adjust to new disease identification tasks by using these pre-trained models which decreases the need for extensive training and efficiency improvement. That is why several studies have applied transfer learning in their disease-discovery techniques [34,35].

Scalability of CNN is also vital as it gives room for efficiently and quickly performs lots of computation when managing diseases across different crops, regions, and seasons. In the work of [36], a Bi-linear Convolution Neural Network (Bi-CNNs) was utilized for plant leaf disease identification and classification where Visual Geometry Group (VGG) was fine-tuned and ResNet was pruned and used for feature extraction to improve scalability in the disease detection task. Another significant advantage of CNNs is their ability to perform real-time disease diagnosis. With the advancements in edge computing and the integration of CNN models into portable devices, such as smartphones or drones, farmers can quickly capture plant images and receive instant disease diagnosis. For instance, Lanjewar and Panchbhai [37] developed and implemented a real-time disease prediction system using a convolutional neural network (CNN) on the Platform-*as*-a-Service (PaaS) cloud. This timely feedback enables prompt action and effective disease management.

Finally, ensemble methods, such as bagging and boosting, have been successfully applied to CNNs for plant disease identification [38]. When multiple CNN models are combined, either through averaging or weighted voting, ensemble methods can improve the classification accuracy and enhance the robustness of the system. They help mitigate overfitting and capture different aspects of the disease patterns, leading to more reliable predictions. This was demonstrated in the work of Al-Gaashani et al. [39] where three lightweight CNNs models (MobileNetv2, NasNetMobile and a simple CNN model from scratch) were used to address the high bias of the single CNN model in order to boost the efficiency of CNN in plant disease identification.

Issues with the implementation of CNNs for plant disease identification.

There are some issues that could come with the implementation of CNN for plant disease identification.

- The performance of CNNs heavily relies on the quality and diversity of the training dataset [40]. Limited or biased datasets can lead to inaccurate disease identification. In order to ensure that datasets are characterized of diverse disease classes, different stages of infection, and environmental conditions is crucial for robust and reliable models.
- Overfitting is a critical issue that can result in a model that struggles to identify plant diseases when given new or diverse plant images yet performs incredibly well on the images it was trained on. This constraint makes it more difficult for the model to generalize to a wide range of datasets, which reduces its usefulness in detecting new plant diseases or variations in plant conditions [41]. Regularization techniques, data augmentation, and accurate model selection can help mitigate overfitting issues and improve generalization abilities [42].
- CNNs are often considered as black box models, since they lack interpretability [43]. Understanding the rationale behind a model's decision can be difficult, especially when it comes to complex image data. Providing insights into the features of plant images that influence the decision of the model is difficult due to lack of interpretability of CNN. Therefore, if the logic underlying the disease identification is unclear, farmers, researchers, or end users can find it difficult to apply the suggestions of the model. The decision-making process can be made more transparent and clearer by exploiting techniques like attention mechanisms, visualization tools, or explainable AI approaches to provide insights into the plant image regions that contribute most to the disease identification of the model.
- Other issues with CNN and other deep learning models for plant disease identification include, hyperparameter tuning which requires experimenting with different settings for hyperparameters such as batch sizes, regularization terms and learning rates to determine the setup that optimized model performance. Techniques such as grid search, randomized search, and automated hyperparameter optimization are often used to combat this issue. Execution time is equally another concern because of lengthy

training and inference process; solutions to this hurdle include using hardware accelerators, optimization techniques, and investigating cloud-based services for faster computation. Also, significant computational power demands require specific hardware like GPUs or TPUs to guarantee compatibility with common deep learning frameworks.

Convolutional neural networks are exceptional tools for identifying plant diseases. They are important in the agricultural domain because of their correctness, scalability, and capacity to extract relevant features from images. With further advancements in their research, they are anticipated to play a crucial part in disease management to increase agricultural output and guarantee food security.

2.2.2. Generative adversarial networks in plant disease identification

Generative Adversarial Networks (GANs) is also a type of deep learning technique that can be used in plant disease identification context to augment dataset with synthetic samples in order to improve the generalization and robustness of the models [36]. Some examples of GANs models include Vanilla GAN, Conditional GAN, Deep Convolutional GAN, Progressive GAN (PGAN), and Wasserstein GAN (WGAN) [44]. GANs can be used for anomaly detection in plant disease identification by learning the distribution of healthy plant images and detecting deviations from this learned distribution [45,46].

Basically, GANs are valuable for generating synthetic data that looks like the real plant disease patterns in plant disease detection. The generator networks within the GAN framework creates synthetic samples of plant leaves affected by various diseases. The discriminator network then learns to differentiate between these generated samples and the real data, which consists of images of plant leaves with actual diseases [47] as depicted in Fig. 4.

GANs are effective when it comes to training with unlabeled data which usually assist them to quickly generate feasible and high-quality outputs. GANs are very easy to train, hence they usually converge faster than other types of generative models. For instance, Stephen et al. [48] employed optimized deep generative adversarial network (DGAN) and 3D 2D CNN to successfully classify rice leaf diseases with an improved accuracy of 98.7 %. Wang and Cao [49] proposed a generative adversarial classified network (GACN) that generated synthetic images to balance plant disease datasets and at the same time the model was able to directly identify plant disease. In the same vein, Song et al. [50] addressed the maize disease identification problem for improved accuracy. They used Attention Generative Adversarial Network (Attention-GAN) to capture relevant information on the vital parts of images and Generative Adversarial Network (GAN) for data augmentation to produce more training data. They reported high performance of their method.

Issues with the implementation of GANs for plant disease identification.

Although Generative Adversarial Networks (GANs) are celebrated for their ability to generate reliable data, they also come with certain drawbacks.

- One of the drawbacks of GANs is mode collapse. This is when GANs produce only a limited number of outputs instead of exploring the entire distribution of the training data. This makes the output produced to be repetitive and usually unconnected to the training data. In the situation of plant diseases, mode collapse can result in the generation of only a few types of diseases, which can limit the capability of the model to identify an extensive range of conditions. To overcome this challenge, model architecture can be modified for better performance.
- Training Generative Adversarial Networks (GANs) can be challenging when it comes to plant disease identification. This is based on the fact that the generator and the discriminator are always in constant competition with each other. The competition can lead to training instability and a slow convergence which can impact the development of an effective model for identifying plant diseases. Careful parameter tuning and optimization could be introduced to achieve stable and efficient training.
- Large GANs training can be computationally intensive since it requires significant computational resources, which can lead to longer training times that could be impractical for certain research or practical applications. Strategies such as model optimization, distributed training and cloud computing can be investigated to address these computational demands to speed up training process.

2.2.3. Recurrent neural networks in plant disease identification

Recurrent Neural Networks (RNNs) are a model of choice for processing data sequences [51], which makes them suitable for plant

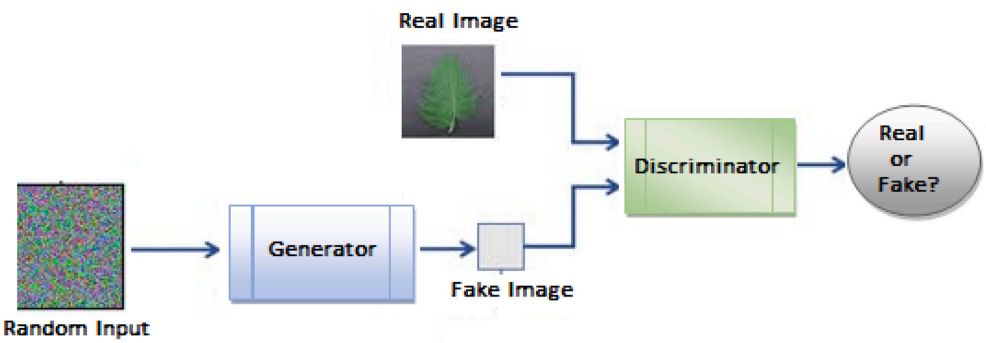

Fig. 4. Generative adversarial networks for plant disease identification.

disease identification. RNNs are particularly excellent at identifying patterns and temporal dependencies in datasets that demonstrate the progression of diseases over time. Their capability to process sequential information provides a detailed knowledge of how plant diseases progress. This makes RNNs an effective tool for the analysis and identification of temporal patterns associated with different stages of diseases in plants. RNN architecture typically allows the output from the previous step to be fed as input to the current step [52] as seen in Fig. 5. The architecture is defined as a network of interconnected units which can be connected to features of plant leaves in the context of detecting plant diseases. The input layer receive information about these features, and the connections between nodes, with their adjustable weights to allow the network to learn relationships and patterns within the data. The output layer then produces the final classification that determines if a plant leaf is healthy or unhealthy. Since RNNs are well-suited for processing sequential data, they are easily applied to tasks where the order of inputs matters, such as evaluating temporal data like time series or sequential data.

For instance, Lee et al. [53] developed a novel approach based on Recurrent Neural Network (RNN) to automatically locate infected areas in plant leaves and extract relevant features for disease identification, they showed through experiments that RNN-based technique is more robust and has a better ability to generalize unseen infected plant leaves as compared to traditional CNN approaches. Lately, RNN and CNN are jointly used to address time-series image classification problems. In such situation, convolutional layers are used for extracting features from raw image data in deeper layers and producing high-level representations. These features are used by the recurrent layers for learning the time dependencies. For example, Nandhini et al. [54] proposed a G-RecConNN which is a new sequential image classifier for plant images grounded on CNN and RNN directly accept sequences of images without recalculating the image differences in the series. The proposed model achieved high accuracy in identifying different plant diseases. In addition, Daniya and Vigneshwari [55] developed a classifier, called RSW-based Deep RNN by modifying the training algorithm of Deep RNN with RideSpider Water Wave (RSW) algorithm. They reported experimentally that the proposed RWS-based Deep RNN provides superior performance with the highest accuracy of 90.5 %. RNNs also have ability to retain information from previous time steps, enabling the incorporation of contextual information and accurate predictions based on historical data [56]. Other variations of RNN that are applied in plant disease identification include Long Short-Term Memory (LSTM) which can be applied to plant diseases identification using sequential data like temporal patterns in plant health observations [57], also, LSTM architecture permits it to recall information over long periods, making it able to learn complex relationships between past and present observations [58]. They are robust to noise and missing data, which can be common in real-world plant health datasets [59]. LSTM models can operate in real-time, enabling continuous monitoring of plant health [60]. While Recurrent Neural Networks (RNNs) have shown some potentials in plant leaf disease identification, they also come with some limitations as discusses in the following section. Also, Gated Recurrent Units (GRUs) is another variation of RNNs that have the potential to improve the identification of plant diseases through efficient temporal data analysis. They are used to analyze time-series data, capture the sequential dependencies and temporal nuances in the progression of plant diseases.

Issues with the implementation of rnn in plant disease identification.

- RNNs are susceptible to the issue of vanishing or exploding gradients, especially when dealing with long sequences [61]. This can lead to difficulties in training the model effectively to capture long-range dependencies in the spread of leaf disease. Techniques such as gradient clipping, testing with advanced optimization algorithms, such as RMSprop or Adam and applying gradient regularization methods like dropout or recurrent dropout can be implemented.

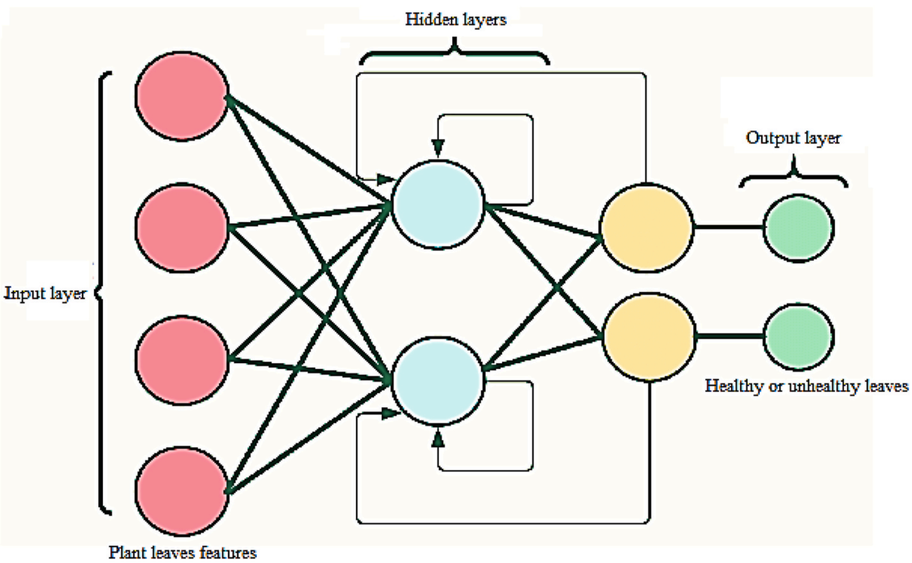

Fig. 5. Recurrent neural networks for plant disease identification.

- RNNs process data sequentially, it can restrict their parallelization abilities and make them slower when compared to other architectures like CNN [62]. This Sequential processing of RNNs might cause delays when processing a significant amount of plant image data or when quick disease identification depends on real-time processing. Using deep learning frameworks optimized for parallel computation, such as TensorFlow or PyTorch with GPU support can mitigate this problem.
- Memory constraints is another problem that comes up when managing long sequences [63] or complex disease spread, as the memory of the network capacity may become reduced. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) cells can address this issue to some extent by incorporating techniques that can selectively remember or forget information.
- RNNs can also be sensitive to noisy information in the input sequences [64] as there may be possible differences in lighting conditions, backgrounds and other factors that can introduce noise into plant images. It is therefore very important to thoroughly clean the data to eliminate such noise and improve the performance of RNNs. Training RNNs effectively requires a substantial amount of labeled data [65], collection and labelling such large-scale plant leaf datasets can be frustrating and expensive. Also, limited or imbalanced datasets can impact the generalization and accuracy of RNN models. To address some of the issues raised above, alternative architectures like Transformer-based models or hybrid strategies that incorporate RNNs with other neural

Table 2
Summary of deep learning techniques used for plant disease identification.

S/N	DL algorithm in plant disease identification	Strength of DL algorithms in plant disease identification	Linitations of the DL algorithm in plant disease identification	Improvement strategies for DL algorithm in plant disease identification	Area of application of the DL algorithm in plant disease identification
1.	CNNs	Ability to extract useful features for disease identification, use of transfer learning for pre-trained models, it can scale effectively to handle large volumes of plant images, can mitigate overfitting through ensemble methods and capture different aspects of disease patterns, has ability to perform real-time disease diagnosis	Limited or biased datasets can lead to inaccurate disease identification, they can be prone to overfitting, they lack interpretability.	Ensure that dataset is representative of various disease types, different stages of infection, and environmental conditions, use regularization techniques, data augmentation, and careful model selection to mitigate overfitting.	Leaf disease classification, early disease diagnosis, multi-class disease identification, disease severity estimation, disease progression monitoring, disease localization.
2.	GANs	They can generate synthetic images that closely resemble real plant disease samples. GANs pre-trained on several image datasets can be fine-tuned for plant disease identification. They have anomaly detection capability which can enhance the sensitivity of the model to subtle disease symptoms.	GANs are prone to mode collapse which can lead to generating only a subset of disease symptoms. Their training could be ustable which sometimes causes difficulties in converging to an optimal solution. Also, Training large GAN models can be computationally intensive and can lead to long training times.	Model architecture can be modified for better performance. Engage in careful parameter tuning and optimization to achieve stable and efficient training. Also, distributed training and cloud computing can be considered to address these computational demands to speed up training process.	Synthetic samples of plant disease generation (data augmentation), early disease detection, disease identification accuracy
3.	RNN	Sequential modeling ability of RNNs allow them to effectively capture the temporal nature of plant disease symptoms, RNNs can process data with varying lengths, which is common in plant disease identification tasks that involve time series data, RNNs can adapt to different types of input data allowing for complete disease identification.	Difficulties in training models effectively and capturing long-range dependencies in leaf disease progression, sequential data processing ability which can limit their parallelization capabilities, need to maintain memory of previous inputs to capture temporal dependencies, be sensitive to noisy or irrelevant information, challenges to interpret and understand why certain decisions or predictions are made.	Use LSTM and GRU cells to incorporate mechanisms to selectively remember or forget information, preprocess data to remove noise, introduce attention mechanisms and visualization methods to provide some level of interpretability to predictions made by RNN, use large data.	Disease severity assessment, disease management decision support, disease progression prediction, disease detection and classification, image-based disease diagnosis.
4	Tranformers	They could capture long-range dependencies in image sequences, process spatial data effectively, scale to handle big datasets, and use pre-trained models through transfer learning,	Computational power, resource requirements, and possible omission of spatial relationships are critical for disease identification. Also, the need to include substantial labeled data for training could be a major limitation, which can make optimizing models for efficiency crucial.	Use transfer learning with a variety of datasets, investigate hybrid models to handle spatial relationship, use efficiency-optimized model architectures, and experiment with methods like image cropping for computing efficiency.	To extract spatial information in plant images, capture long-range dependencies for disease identification, to enhance scalability of massive datasets, and harness the capability of pre-trained models through transfer learning.

network architectures could be examined. Additionally, methods including data augmentation, transfer learning, and ensemble learning can assist improve RNNs' performance in tasks requiring the identification of plant leaf diseases. Table 2 summarizes the deep learning techniques used for plant disease identification.

2.2.4. Transformers in plant disease identification

Transformers are an advanced type of deep neural network architecture widely used to perform natural language processing (NLP), computer vision (CV), and speech processing (SP) tasks. Their distinctive feature is the use of a self-attention mechanism to accurately process sequential input data [66]. What sets transformers apart is their unique ability to process entire input sequences simultaneously to facilitate a more comprehensive understanding of context and relevance compared to traditional models [67]. This characteristic proves valuable in overcoming challenges associated with longer sequences to effectively address issues like the vanishing gradients problem commonly encountered by recurrent neural networks (RNNs). Transformer models include Bidirectional Encoder Representations from Transformers (BERT), Generative Pre-trained Transformer 3 (GPT-3), Vision Transformer (ViT) and DEtection Transformer (DETR). BERT and GPT-3 are used for NLP while ViT and DETR are used in Computer vision.

VisonTransformer models usually represent an image as a sequence of non-overlapping fixed-size patches that are then linearly embedded into 1D vectors (Fig. 6). The Transformer model uses the vectors as input tokens. When processing the input data, the self-attention mechanism is used to assist the model to determine the relative relevance of various tokens in the sequence. Since the self-attention mechanism allows the model to capture global contextual information, it can learn long-range dependencies and relationships between image patches. Vison Transformer models comprise of an encoder, which contains several layers of self-attention and feed-forward neural networks, and a decoder that generates the final output.

Currently, transformers are being explored in the domain of plant disease identification [67]. They treat images as a sequence of patches to capture fine-grained details and spatial relationships. Their self-attention mechanism enables focused analysis which allow the model to discover relevant features associated with disease symptoms [66]. Transformers excel in handling longer sequences which make them skilled in processing complex plant images. Therefore, their adaptability, combined with interpretability through attention maps, offers promise in accurately identifying diseases, especially with limited labeled data.

Issues with the implementation of transformers for plant disease identification

Specific challenges may be experience during the implementation of transformers to identify plant diseases.

- One major problem when training transformer models for plant disease identification is the lack of adequate labeled data [68]. Collecting massive, labeled datasets for various plant diseases can be resource-intensive and time-consuming. There may be possibility that the overall performance of the model may be limited as it may not be able to generalize well to the different plant species and diseases. This problem can be confronted by using transfer learning that pre-trains models on a big dataset of different image classes, then fine-tuning it to fit the smaller dataset of plant diseases.
- Vision transformers, though powerful, often require a high processing power during training [69]. This arises from the extensive parameters and the necessity to process every part of the image, even when not essential for the given task. The large pixel sizes of images also contribute to their high computational cost. To alleviate this, model architectures optimized for efficiency can be explored. The goal of these architectures is to minimize computational demands without sacrificing performance. Likewise, by taking advantage of cloud-based resources, training may be scaled efficiently and more cost-effectively due to distributed computing capabilities.

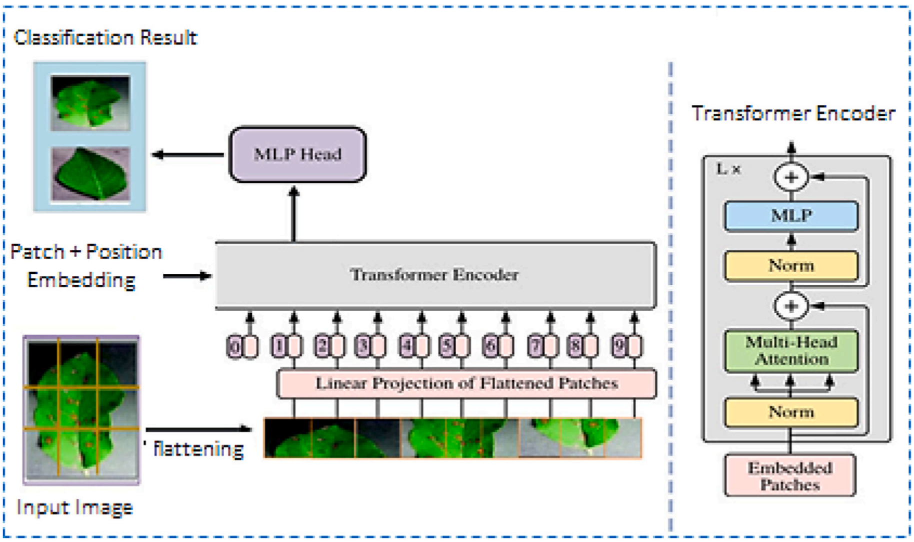

Fig. 6. A Vision Transformer Model for plant disease identification.

- Plant diseases often exhibit spatial patterns and localized symptoms, failure to capture these nuances may hinder the accuracy of model in identifying diseases [66]. To capture this problem, hybrid models that integrate spatial information can be explored, or the transformer may be augmented with extra layers designed to capture spatial dependencies.

CNNs are well-suited for image-based disease identification, while GANs are powerful in plant disease identification context to augment dataset with synthetic samples. RNNs are effective in handling sequential data and temporal dependencies. Transformers provide special attention mechanism that assist in capturing relevant features associated with disease symptoms. Depending on the nature of the data and the specific task at hand, either CNNs, GANs, RNNs, and Transformers or a combination of any of them can be applied to improve plant disease identification accuracy and provide valuable insights for disease management. Table 3 provides different instances of research studies where robust algorithms have been applied for Plant disease Identification.

3. Traditional and content-based filtering techniques for providing treatment recommendations for plant diseases

3.1. Traditional techniques for providing treatment recommendations for plant diseases

A rule-based technique in plant disease treatment recommendation uses traditional methods to suggest an approprite remedy for plant diseases. It is a recommendation strategy that is based recognized diseases and specific contextual elements. It perates by examining the characteristics of the disease, plant species, environmental conditions, and other relevant factors to provide personalized recommendations for disease control. For instance, Abu-Nasser et al. [75] suggested an expert system that could identify and offer remedies for various types of identified watermelon diseases. Alajrami and Abu-Naser [76] equally design an expert system which helped farmers to diagnose and provide appropriate treatments for onion plant diseases. In rule-based system for plant disease treatment recommendation, the rules are usually created by domain experts, who has knowledge about effective treatment methods and management practices. For instance, a rule-based system for plant disease treatment recommendation might be as follows.

Rule 1: **IF** the plant leaves show yellow spots **AND** the spots are encircled by a brown ring **AND** the leaves are curly **THEN** it implies the likelihood of Tomato Yellow Leaf Curl Virus. **Treatment**: Detach the unhealthy plants from the healthy ones to avoid further spread of the disease.

Table 3
Some cases of robust algorithms implemented for plant disease identification.

Model	Study	Year	Technique	Results
CNN	Zhang et al. [27]	2019	Dilated CNN with global pooling	The proposed model effectively identified cucumber diseases.
	Nanehkaran et al. [70]	2020	Used hue, saturation and intensity-based and LAB-based hybrid segmentation algorithm and CNN	Accuracy obtained was 15.51 % higher than the traditional technique for plant disease identification
	Chen et al. [71]	2021	MobileNet-V2 pre-trained on ImageNet as backbone network and the attention mechanism along with a classification activation map (CAM) were utilized	Achieved an average accuracy of 99.14 % for Crop pest recognition
	Hassan et al. [33]	2021	Different CNN architectures such as InceptionV3, InceptionResNetV2, MobileNetV2, and EfficientNetB0 were used	A disease-classification accuracy rates of 99.56 % was obtained
	Al-Gaashani et al. [39]	2023	Ensemble of MobileNetv2, NasNetMobile and a simple CNN	Achieved 98 % accuracy in identifying plant diseases
GAN	Bi and Hu [36]	2020	Wasserstein generative adversarial network with gradient penalty (WGAN-GP)	Improved the overall classification accuracy of plant diseases by 24.4 %
	Wang and Cao [49]	2023	GAN combined with a classifier	Achieved recognition accuracy of 99.78 % and 86.52 %, respectively on PlantVillage and AI Challenger 2018 datasets
	Song et al. [50]	2023	Generative adversarial network (GAN) and attention mechanism	High accuracy of 97 % in maize disease detection tasks
	Stephen et al. [48]	2024	Optimized deep generative adversarial network (DGAN) and 3D 2D CNN	Classified rice leaf diseases with an improved accuracy of 98.7 %.
RNN	Nandhini et al. [54]	2022	RNN and CNN were combined named Gated-Recurrent Convolutional Neural Network (G-RecConNN)	The model correctly classified the diseases of plantain trees with between 2.8 % and 4.2 % increase in the recall measure
	Tanwar and Singh [72]	2023	Hybrid of ResNext50-LSTM	Accuracy of 95.44 % was attained for plant disease classification
	Rajalakshmi et al. [73]	2024	Gated recurrent multi-attention neural network (GRMA-Net).	Achieved 99.16 % accuracy in identifying agricultural diseases
Transformers	Yu et al. [67]	2023	Inception convolution and vision transformer	Had 99.22 % accuracy on ibean dataset for plant disease identification
	Thai et al. [68]	2023	ViT + Least Important Attention Pruning (LeIAP) algorithm and sparse matrix-matrix multiplication (SPMM)	Evaluation results on the cassava leaf disease dataset indicate that the algorithms reduced the model size to 28 % and increase the training and inference speed by 10 % and 15 %
	Thakur et al. [74]	2023	CNN combined with ViT model	It achieved accuracy of 98.86 % and precision of 98.9 % on PlantVillage' dataset in dentifying plant diseases.

Rule 2: IF the plant leaves show water-soaked lesions that turns dark and hollowed **AND** the stems display a blackish stain **THEN** it implies the likelihood of a bacterial infection named Bacterial Blight. **Treatment:** Use a bactericide that has copper or detach the unhealth plant parts to reduce the spread of the disease.

When a plant disease is detected, the rule-based system evaluate the appropriate facts against the predefined rules to verify the suitable treatment. The system confirms which rules are met by the disease features and contextual factors; it then offers recommendations according to the activated rules. The basic structure of a rule-based system is depicted in Fig. 7.

Plant disease treatment recommendation rule-based systems have several benefits. They can provide quick, automated ideas for disease prevention and give farmers practical advice on how to mitigate the effects of diseases. Rule-based systems can also be transparent and comprehensible [77], as the fundamental principles can be evaluated and adjusted by domain experts to enhance the correctness and significance of recommendations. Rule-based systems for plant disease treatment recommendation also face challenges [78] such as, limited adaptability to uncommon diseases, difficulty in managing and keeping numerous rules, Lack of ambiguity in managing overlapping symptoms and difficulty in describing intricate relationships between symptoms. investigating complementary techniques, which include data-driven models or hybrid systems, can be explored to overcome these challenges and enhance the efficiency of plant disease treatment recommendation systems. Rule-based systems are often applied in diverse domains, such as expert systems [79] and decision support systems. Table 3 shows the strength and the weaknesses of rule-based technique for plant disease treatment recommendation.

3.2. Content-based filtering techniques for providing treatment recommendations for plant diseases

Content-based filtering is a recommendation technique that can be applied to provide treatment for plant disease. It uses diverse types of techniques that makes use of the unique features of plant diseases and other vital related information to provide treatment recommendations for plant diseases [80]. The following are some content-based filtering techniques that can be used to provide treatment recommendations for plant diseases:

Hybrid Filtering: They can combine various recommendation techniques, such as content-based filtering, collaborative filtering, and other approaches [81]. Also, they can enhance content-based filtering by incorporating extra data sources such as past treatment records, feedback of user, or collaborative filtering that is based on similar user's experiences. This type of integration will allow the creation of reliable and accurate recommendation systems. As such, different learning techniques were used by Isinkaye and Erute [82] to create a user-friendly smartphone-based plant disease detection and treatment recommendation system. CNN was utilized for feature extraction and the ANN and KNN to categorize the plant diseases. A content-based filtering recommendation algorithm was used to suggest appropriate treatment for the discovered plant diseases after classification.

Feature-Based Filtering: Feature-based filtering focuses on extracting relevant features from plant and disease data and representing them in a feature space [83]. These features could include plant characteristics, disease symptoms, geographical factors, or genetic information. Machine learning algorithms are then applied to learn patterns and similarities within the feature space. This approach can provide reliable and efficient recommendations by capturing complex associations between features. For example, Patil

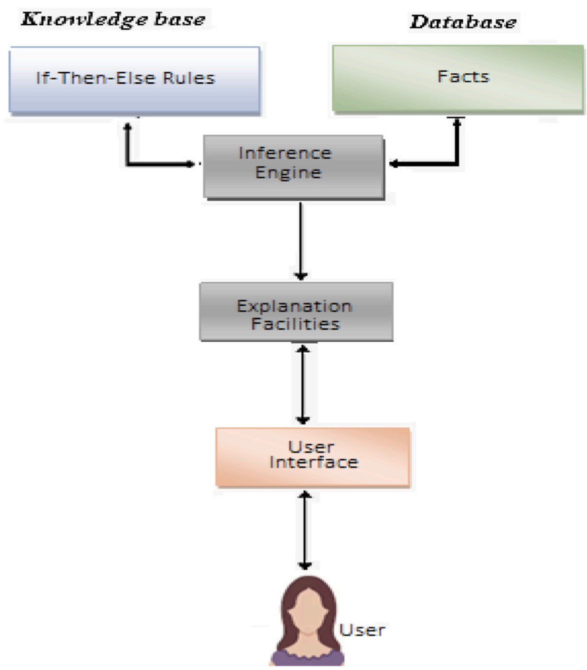

Fig. 7. Basic rule-based system.

et al. [84] utilized a sequential and VGG-16 deep convolutional neural network model to identify diseases in crops while they used content-based filtering to suggest suitable crops to users based on factors such as location and the period of the year for cultivation. Patil et al. [84] also ustilized a sequential and VGG-16 deep convolutional neural network models to identify diseases in crops, they used content-based filtering to suggest suitable crops to users based on parameters like location and the period for cultivation.

3.2.1. Limitations of content-based filtering techniques in plant disease treatment recommendations

While Content-Based Filtering (CBF) offers valuable benefits in plant disease treatment recommendation, it also has certain limitations to consider when using it. For instance, CBF heavily depends on the availability and quality of data. Inaccurate or incomplete data can lead to substandard recommendations [85]. It is imperative to get a reliable and complete dataset for the effectiveness of the system. The technique mainly depend on existing knowledge and prior accomplishments [8], it always have issues in detecting novel treatments that have not been previously explored. Overemphasis on Attribute-Based Similarity [86] is another difficulty associated with CBF, it majorly always compares the attributes and characteristics of plants and diseases. While this approach can be useful in certain circumstances, it may neglect critical factors such as genetic variations, developing diseases, or environmental changes that can impact disease management approaches. Content-based filtering does not directly consider the experiences and recommendations of other users or experts [8]. Collaborative filtering, which draws on the knowledge of other users or experts, might reveal important information that content-based filtering alone cannot capture. Therefore, it is necessary to understand these limitations and consider them in the design and implementation of a plant disease treatment system that will utilize CBF. Combining CBF with other techniques, incorporating user feedback, and accounting for contextual factors can help mitigate some of these limitations and enhance the overall effectiveness of the system. Table 4 Summarizes the strengths and the weaknesses of rule-based and content-based filtering technique for plant disease treatment recommendation.

4. Possible approaches to combine deep learning techniques with content-based filtering techniques to provide tailored and efficient treatment recommendations for plant diseases

A thorough and efficient system for managing plant leaf diseases can be created by combining deep learning approaches for disease identification with content-based filtering for treatment recommendations. The deep learning component focuses on accurately diagnosing diseases, and content-based filtering takes advantage of similarities across disease cases to make recommendations for individual treatments. Therefore, to integrate deep learning and content-based filtering for disease identification and treatment recommendation, the following techniques could be considered.

Hybrid Technique [87]: This develops a hybrid model that combines deep learning for disease identification and content-based filtering for treatment recommendations [88]. The model can consist of two interconnected modules: a deep learning module that classifies plant leaf images and extracts disease-related features, and a content-based filtering module that matches the identified disease with appropriate treatment based on similarity measures. This hybrid integration, will produce a robust and flexible system that can incorporate correct disease identification with tailored treatment recommendations, thereby supporting in effective management of plant diseases.

Feature Extraction and Similarity Computation [89] CNN as a variation of deep learning model can be trained to extract useful features from plant leaf images. These features can be the descriptions of disease-related patterns and characteristics. The extracted features can be used to compute similarity scores between diagnosed leaves and a database of known treatments. Content-based filtering techniques, such as cosine similarity [90] or TF-IDF [91], can be utilized for recommedation function.

Embedding Learning [92]: A deep learning model, such as Siamese or triplet networks can be trained, to learn embeddings [93] of plant leaf images and treatment descriptions. The model can be trained using pairs or triplets of examples, where similar leaves or treatments are encouraged to have similar embeddings. The learned embeddings can then be used to compute similarity scores between leaves and treatments, and recommendations can be made based on the highest scores.

Multi-modal Data Combination [94]: Different data such as images, textual descriptions, and environmental factors can be introduced into the integrated system [95]. Deep learning models, such as multi-modal networks or fusion architectures, can then be employed to process and integrate these modalities. The combined information can enhance disease identification correctness and enhance the content-based filtering for treatment recommendations.

Active Learning and Feedback Loop [96]: Active learning techniques can be integrated to selectively label and acquire new data

Table 4
Summary of strengths and the weaknesses of rule-based and content-based filtering technique for plant disease treatment recommendation.

Technique	Strengths	Weaknesses
Rule-based [78]	They offer quick suggestion for disease control, they can be customized, they are transparent and interpretable.	Limited adaptability to uncommon diseases, complexity in managing and maintaining many rules, subjectivity and variability in symptom interpretation, Lack of uncertainty when handling overlapping symptoms.
Content-based filtering [85]	Tailored recommendations for managing plant diseases, enhances the accuracy and reliability of recommendations, can provide advanced and sophisticated recommendations by capturing complex relationships between features, multiple recommendation techniques	Inaccurate or incomplete data can lead to suboptimal recommendation, limitations in discovering innovative or novel treatments, may overlook important factors, ack of joint filtering.

points for training the learning models [97]. This can lower the reliance on large pre-labeled datasets and allows the models to adjust and improve over time. Also, user feedback on treatment effectiveness can be integrated to refine the content-based filtering component and enhance the recommendations generated to users.

Transfer Learning: Pre-trained deep learning models can be trained on large-scale image datasets like ImageNet and used as a starting point for disease identification [87]. The pre-trained models can be fine tunned using a smaller labeled dataset of plant leaf images specific to diseases. The fine-tuned models can then be used for disease identification, and the extracted features can be used in content-based filtering for treatment recommendations.

5. Datasets for training and testing the deep learning and content-based filtering models in the context of identifying and treating plant diseases

Datasets on plant diseases are essential for advancing plant pathology research and development [98]. The identification, categorization, and control of many plant diseases are all made possible by the knowledge and insights provided by these datasets. These datasets can be analyzed and studied to help create more precise and effective disease detection procedures, disease management plans, and crop improvement methods.

In the fields of computer vision and machine learning, there are several publicly accessible plant disease datasets that are frequently used for training and evaluation purposes. These datasets are important tools for creating and evaluating algorithms for plant disease detection and classification since they provide labeled images of diseased plants along with corresponding healthy plant images. Fig. 8 shows some samples of diseased and healthy plant images. The section following provides a brief description of some datasets that have been used in plant disease detection and classification.

The *Plant Pathology dataset* (https://www.kaggle.com/competitions/plant-pathology-2020-fgvc7/data) is a collection of images created for a Kaggle competition focused on classifying foliar diseases of apples. It comprised 1200 images of apple scab, 1399 of cedar apple rust, 187 of complex disease symptoms (i.e., more than one disease on the same leaf) and 865 of healthy leaves. The plant images were taken with the aid of a Canon Rebel T5i DSLR and smartphones under different illumination, angle, surface and noise conditions, directly from the field [99]. The dataset contained a substantial number of images for each category and provided participants with diverse samples for training and evaluation.

The *Plant Village dataset* (https://github.com/spMohanty/PlantVillage-Dataset) consists of 54303 healthy and unhealthy leaf images divided into 38 categories by species and disease. The images span 14 crop species: apple, blueberry, cherry, corn, grape, orange, peach, bell pepper, potato, raspberry, soybean, squash, strawberry and tomato. It contains images of 17 fungal diseases, 4 bacterial diseases, 2 mold (oomycete) diseases, 2 viral disease and 1 disease caused by a mite. Twelve crop species also include images of healthy leaves that are not visibly affected by a disease [100]. Most of the images were acquired under controlled lab conditions with uniform backgrounds. Also, the dataset in Plant Village is unbalanced and they are not representative of real-field conditions, hence they always find it difficult to generalize with higher accuracy during model training [101].

The *Rice Leaf dataset* (https://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases) consists of three classes of diseases which are bacterial leaf blight, brown spot and leaf smut, each having 40 images and a total of 120 images collected from a village called Shertha near Gandhinagar, Gujarat, India, captured with a white background using a Nikon D90 digital SRL camera with 12.3 megapixels in November 2015. The authors collected leaves with varying degrees of disease spread, where all images have a resolution of 2848 × 4288 pixels [100].

PlantDoc dataset (https://github.com/pratikkayal/PlantDoc-Object-Detection-Dataset https://github.com/pratikkayal/PlantDoc-Dataset) was released by Researchers at Indian Institute of Technology in the fall of 2019, it consists of images of 2598 dataset across 13 plant species and 27 classes (17 disease; 10 healthy), most of which were acquired under field conditions for the purpose of image classification and object detection.

The *robusta coffee leaf images dataset* (RoCoLe) (https://doi.org/10.17632/c5yvn32dzg.2) contains 1560 leaf images with visible red mites and spots (denoting coffee leaf rust presence) Tofor infection cases and images without such structures for healthy cases. Also, the dataset has annotations about objects (leaves), state (healthy and unhealthy) and the severity of disease (leaf area with spots). The images were collected in real-world conditions in the same coffee plants field using a smartphone camera. RoCoLe dataset facilitates the evaluation of the performance of machine learning algorithms used in image segmentation and classification problems

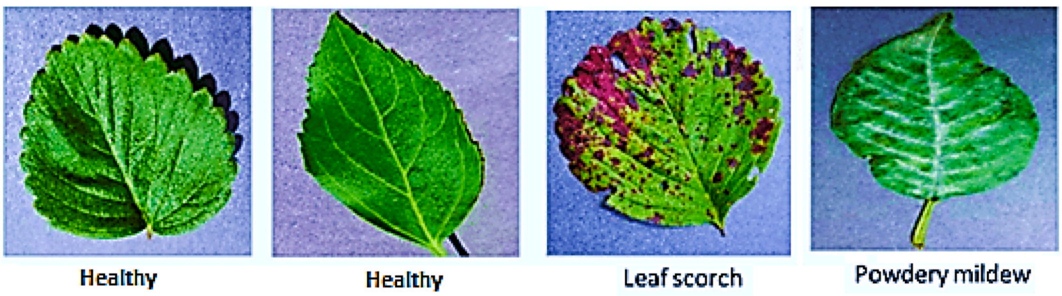

Fig. 8. Examples of healthy plant images.

related to plant diseases identification [102].

BRACOL (Brazilian Arabica Coffee Leaf) (https://data.mendeley.com/datasets/yy2k5y8mxg/1/files/c16b08ee-3ca6-4bf0-8f4e-4285a53a4a24) images dataset is used to identify coffee diseases. It has 1747 leaves images which are affected by biotic stresses such as leaf miner, leaf rust, brown leaf spot and cercospora leaf spot [103]. The images were taken from the abaxial side of the leaves under partially controlled conditions and placed on a white background with the aid of five different smartphones. The annotations on the images were made using the tool VGG Image Annotator, abbreviated by VIA [104].

New Plant Diseases Dataset available at (https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset), is composed of 87 K images of healthy and diseased crop leaves, categorized into 38 different classes. The datasets is divided into 80 % for training and 20 % for validation. Additionally, a directory consisting of 33 test images is created for prediction purposes. The dataset was generated through offline augmentation techniques applied to the original dataset.

Plant disease datasets are useful tools that support research and innovation in plant pathology. They aid in building more accurate disease detection techniques, management plans, and predictive models. Therefore, efforts should be directed toward improving the quality, diversity, and availability of these datasets to further facilitate advancements in plant disease research and support sustainable agriculture practices. Numerous datasets suitable for plant disease research are available in the repository of Kaggle at https://www.kaggle.com/datasets?search=plant+disease+datasets.

6. Performance measures to evaluate the accuracy and efficiency of the deep learning and content-based filtering techniques

The evaluation metrics that are frequently employed to measure treatment recommendations for plant diseases usually offer insightful information about their effectiveness. The specific intentions and dataset properties should guide the choice of metrics to use. A blend of metrics can give a thorough understanding of the strengths and weaknesses of a system. The following evaluation metrics relevant to plant disease treatment recommendation are emphasized as prominent in the domain of study.

Accuracy: Accuracy measures the overall correctness of the model's predictions [105] for disease treatment recommendations. It calculates the ratio of correctly predicted treatment recommendations to the total number of cases.

$$Accuracy = \frac{\text{Number of correct treatments recommendation}}{\text{Total number of treatments recommendations}}$$

Precision: Precision evaluates the accuracy of the positive disease treatment recommendations made by the model. It measures the fraction of correctly predicted positive cases over all the positive predictions [106].

$$Precision = \frac{\text{Number of relevant treatments recommended}}{\text{Total number of treatments recommended}}$$

Recall (Sensitivity or True Positive Rate): Recall calculates the model's ability to identify all positive disease cases and recommended treatments. It measures the ratio of true positive cases and recommendations to the total number of actual positive cases [106].

$$Recall = \frac{\text{Number of relevant treatments recommended}}{\text{Total number of relevant treatments}}$$

F1 Score: The F1 score is the harmonic mean of precision and recall [107]. It provides a balanced evaluation metric that considers both precision and recall, which is important for disease treatment recommendation tasks.

$$F1\ score = 2 * \frac{(Precision * Recall)}{(Precision * Recall)}$$

Mean Absolute Error (MAE): It is the measure of the average absolute difference between the predicted value, y and the actual value, $\bar{y}$ given by users to items [108]. However, in the context of plant disease treatment recommendations, Mean Absolute Error (MAE) can be used to evaluate the accuracy of the system's predictions for the effectiveness of different treatments for specific plant diseases. The MAE formula for plant disease treatment recommendations is calculated as:

$$MAE = {}^{1}\!/_{N} \sum |y - \bar{y}|$$

where "N" represents the total number of plant disease treatment cases. A lower MAE specifies that the recommendation system's predictions are more accurate and closer to the actual effectiveness, while a higher MAE suggests less accuracy in the predictions.

Confusion Matrix: The confusion matrix provides a detailed breakdown of the model's predictions for disease treatment recommendations. It shows the number of true positive, true negative, false positive, and false negative cases. It helps in understanding the types of errors made by the model and its performance across different classes [109].

Execution time: The execution time or CPU time in plant disease identification is the amount of time required for a learning model to finish the training and inference processes (time taken for the model to make predictions on new, unobserved data). That is, the execution time, ET of the operation corresponds to the difference between the end time, T_E and the start time, T_S as depicted:

$$ET = T_E - T_S$$

Table 5 highlights the strength and the weaknesses of evaluation metrics used for plant disease treatment recommendation and possible improvement that can enhance their performances.

7. Gaps and areas of improvement in current research, and the likely future directions for advancing the field

This section discusses the gaps, areas primed for improvement, and promising avenues for future development in the existing plant disease identification and treatment recommendation systems.

7.1. Gaps and improvements

Despite improvements in image recognition and machine learning approaches, current systems can still have trouble in correctly identifying plant diseases. The quality and diversity of the training dataset, which may not always capture the full spectrum of disease

Table 5
Summary of the Strength and the weaknesses of evaluation metrics used for plant disease treatment recommendation.

Metric	Strength	Weaknesses	Improvements
Accuracy	Simplicity as it is easy to understand and calculate, it gives a direct interpretation of the performance of RSs in terms of how well it is making correct treatment suggestions, when the dataset is well-balanced, accuracy metric appropriately reflects the overall model performance.	Imbalance datasets can lead to high accuracy which can be misleading, as the model may be biased towards predicting the majority class, leading to poor performance on the minority class, it ignores severity of errors as it treats all recommendations as equally important, regardless of the treatment's effectiveness, it does not consider the degree of relevance or effectiveness of treatments	use data augmentation techniques to enhance dataset class imbalance problem, include treatment effectiveness score or confidence levels in the recommendation process, use accuracy in conjunction with other evaluation metrics
Precision	It Focuses on the correctness of treatment recommended, useful for decision making, especially in situations where the cost of false positives (e.g., recommending irrelevant treatments) is significant,	Precision does not consider false negatives; therefore, a model can achieve high precision by being conservative and rarely recommending treatments, this might result in ignoring critical cases where speedy treatment is required.	a filter that verifies recommendations based on additional information or expert validation before presenting them to the end-users, in the case of imbalanced datasets where one class (e.g., disease) is more prevalent than the other (e.g., healthy plants), use class weights during model training to assign higher weights to the minority class (disease) to emphasize its importance, which can lead to improved precision.
Recall	It captures all positives cases, for example, in plant disease treatment recommendation, recall indicates the ability of the model to recommend treatments for as many cases of the disease as possible, it ensures comprehensive recommendations in situations where missing a positive case (a disease) is costly.	A model with high recall might generate many treatment recommendations, including some false positives, which could lead to unnecessary expenses and resource wastage, models designed to be highly sensitive (high recall) may be more likely to recommend treatments even when there is uncertainty, leading to an increased number of false positives.	Incorporate data from various sources to supplement traditional plant disease data, this can improve the model's ability to detect diseases, leading to higher recall, set a lower probability threshold for making treatment recommendations to increase sensitivity. This can lead to more false positives which can help identify a higher number of true positive cases, thereby improving recall.
Mean Absolute Error (MAE)	MAE is easy to understand and interpret, less sensitive to outliers compared to other error metrics like Mean Squared Error (MSE), emphasis on accuracy hence, it penalizes large errors proportionally to their magnitude.	MAE treats all errors equally, regardless of the severity level, as MAE is not differentiable at zero, it may not be ideal for some optimization approaches.	Properly preprocess the data to handle missing values, outliers, and noise, removing or imputing problematic data points can improve the model's ability to learn from the clean data, augment the dataset with additional synthetic samples to improve the model's ability to generalize to different disease severity levels and conditions.
Confusion Matrix	It allows for a more distinctive analysis beyond simple accuracy, confusion matrix helps identify the types of errors a model makes, this can help reveal which diseases are misclassified or missed by the model, leading to insights for improvement.	Confusion matrix treats all misclassifications equally, irrespective of the severity of the disease, majorly confusion matrix is used for binary classification problems so it might not thrive on multi-class classification.	Experiment with different confidence thresholds for making disease predictions, adjusting the threshold can impact the trade-off between false positives and false negatives.
Execution time	It gives a practical estimation of the performance of the system in real-world situation, Extreme execution times can indicate performance bottlenecks which can guide optimization efforts	consistency across different datasets can be affected because of varied execution time, which is based on the input data features, it may not provide in-depth understanding of the underlying behavior of specific computational complexities of algorithms. The complete picture of the performance of the system may not be captured by execution time alone, therefore factors such as responsiveness or efficiency could be ignored	To have a better understanding of the algorithm's performance, combine algorithm-specific diagnostics tools with execution time metric. For a thorough assessment, combine execution time metrics with other metrics like memory usage and responsiveness. Use diverse datasets for evaluation.

variations or account for regional variances, have a significant impact on the accuracy of these systems. Efforts should be focused on establishing bigger and full datasets that contain diverse conditions and geographical areas. For disease identification, many current systems rely on static photos or periodic data, which may not adequately reflect the dynamic character of plant diseases. For detecting the spread of diseases and speedy intervention, real-time monitoring is essential, but it is frequently absent in current systems. Real-time monitoring and early disease identification made possible by advances in sensor, IoT, and remote sensing technologies can lead to prompt intervention. There is room for improvement in the field of developing prediction models that can predict disease outbreaks based on historical data, weather patterns, and other variables. It is possible to give farmers timely warnings by modeling disease spread using sophisticated machine learning and statistical methodologies. If these models are successfully implemented, crop losses might be decreased, interventions could be made sooner, and resource usage in agriculture could be improved.

Systems for identifying plant diseases frequently need a lot of computing power to process and analyze big datasets or execute sophisticated machine learning algorithms. Deploying and utilizing these systems may be hampered by a lack of access to high-performance computing infrastructure, particularly in distant agricultural areas or locations with limited resources. To address this gap, domain knowledge can be utilized to extract and choose relevant features from input data, thereby lowering the input space and computational effort without noticeably affecting accuracy. Deep neural networks are frequently referred to as "black boxes," as it is always challenging interpreting the inner workings of these models and comprehend the justification for their decisions. Lack of model interpretability can undermine reliability, openness, and the capacity to verify and improve the recommendations generated by the system. It is possible to create reliable user interfaces that permit users to interact with and examine model explanations, predictions, and insights. This hands-on exploration enhances accurate understanding of the system.

7.2. Future directions

Future research in the domain of plant disease identification and treatment recommendation systems can focus on several promising areas to enhance their accuracy and effectiveness. For example, developing more interpretable and transparent machine learning models is crucial for building trust and understanding in plant disease identification systems. Research can explore approaches that can make complex models more understandable, so that users can decrypt the reasons behind the system's recommendations and thus enhance better decision-making. Diverse data modalities can also be combined to provide a clear understanding of plant disease. In other words, when various data sources are combined with current systems, disease detection, prediction, and treatment recommendation accuracy can be improved. Internet of Things (IoT) devices and sensor networks can also enable real-time data collection and monitoring of environmental conditions, soil moisture, pest populations, and plant health parameters. These data sources can be combined with disease identification systems to improve their accuracy and enable timely interventions. Additionally, the detection of disease-resistant trait in plant variations and the provision of tailored recommendations based on certain genetic features can be made possible by integrating genetic and genomic data into disease identification systems.

8. Conclusion

Plant diseases have adverse effects on crop yields, quality, and economic stability in agricultural systems worldwide. Agricultural sector has been faced with the challenge of protecting yields against the threat of these diseases. Traditional methods for identification of plant diseases have many limitations. To reduce plant disease imposed-threat to crop growth and hence sustainable food production, there is need to identify alternative techniques that could provide a better remedy to plant disease problems. Therefore, in this review, we have described the different strengths and weaknesses of both traditional and the state of the arts techniques in identifying and treating plant diseases and suggested their improvement stategies, we proposed possible approaches to combine content-based filtering with deep learning to provide tailored and efficient treatment recommendations for plant diseases, the features of different datasets used in the domain were also examined, also the strength and the weaknesses of evaluation metrics used for plant disease identification and treatment recommendation were also investigated and possible enhancement to improve their performances were suggested. Finally, gaps associated with the domain were identified and areas of improvement in current research, and the likely future directions for advancing the field were suggested. Acting on these improvement startegies and future directions suggested will empower researchers, experts and policy makers to continue to advance technologies that will better help farmers preserve their crops and carry out agricultural activities with ease and hence sustainable and resilient agriculture.

Funding

The authors declare that no financial support was received for this research.

Data availability statement

No data was used for the research described in the article.

CRediT authorship contribution statement

Folasade Olubusola Isinkaye: Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Michael Olusoji Olusanya:** Writing – review & editing, Supervision. **Pramod Kumar Singh:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the research support received from the National Research Foundation (NRF) of South under Grant Number [PSTD2204113035], through the Innovation Postdoctoral Fellowship award, and the hosting institution,Sol Plaatje University, South Africa.

References

[1] B.K. Singh, M. Delgado Baquerizo, E. Egidi, E. Guirado, J. E Leach, H. Liu, P. Trivedi, Climate change impacts on plant pathogens, food security and paths forward, Nat. Rev. Microbiol. 24 (2023) 1–17, https://doi.org/10.1038/s41579-023-00900-7.
[2] Y. Fang, R.P. Ramasamy, Current and prospective methods for plant disease detection, Biosens 5 (3) (2015) 537–561, https://doi.org/10.3390/bios5030537.
[3] A. Khakimov, I. Salakhutdinov, A. Omolikov, S. Utaganov, Traditional and current-prospective methods of agricultural plant diseases detection: a review, in: IOP Conference Series: Earth Environ. Sci, vol. 951, 2022 012002, https://doi.org/10.1088/1755-1315/951/1/012002, 1.
[4] U. Grote, A. Fasse, T.T. Nguyen, O. Erenstein, Food security and the dynamics of wheat and maize value chains in Africa and Asia, Front. Sustain. Food Syst. 4 (2021) 617009, https://doi.org/10.3389/fsufs.2020.617009.
[5] S.N. Hlophe-Ginindza, N.S. Mpandeli, The role of small-scale farmers in ensuring food security in Africa, Food Security in Africa (2021) 1–12, https://doi.org/10.5772/intechopen.91694.
[6] K. Patel, H.B. Patel, Machine learning and recommendation system in agriculture: a survey and possible extensions, in: International Conference on Innovations in Computer Science and Engineering, 2022, pp. 665–674, https://doi.org/10.1007/978-981-19-7455-7_53.
[7] H. Yu, L.T. Yang, Q. Zhang, D. Armstrong, M.J. Deen, Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives, Neurocomputing 444 (2021) 92–110, https://doi.org/10.1016/j.neucom.2020.04.157.
[8] F. O Isinkaye, Y.O. Folajimi, B. A, Ojokoh, Recommendation systems: principles, methods and evaluation, Egypt. Inform. J 16 (3) (2015) 261–273, https://doi.org/10.1016/j.eij.2015.06.005.
[9] L. Kaur, S.G. Sharma, Identification of plant diseases and distinct approaches for their management, Bull. Natl. Res. Cent. 45 (2021) 1–10, https://doi.org/10.1186/s42269-021-00627-6.
[10] R. Yakkundimath, G. Saunshi, S. Palaiah, Automatic methods for classification of visual based viral and bacterial disease symptoms in plants, Int. J. Inf. Technol. 14 (1) (2022) 287–299, https://doi.org/10.1007/s41870-021-00701-2.
[11] G. Hbock, Poole, P.E. Parker, T.R. Gottwald, Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging, CRC Crit. Rev. Plant Sci. 29 (2) (2010) 59–107, https://doi.org/10.1080/07352681003617285.
[12] A.Y. Khaled, S. Abd Aziz, S.K. Bejo, N.M. Nawi, I.A. Seman, D.I. Onwude, Early detection of diseases in plant tissue using spectroscopy–applications and limitations, Appl. Spectrosc. Rev. 53 (1) (2018) 36–64, https://doi.org/10.1080/05704928.2017.1352510.
[13] A. Rao, A. And, S. B Kulkarni, A hybrid approach for plant leaf disease detection and classification using digital image processing methods, Int. J. Electr. Eng. Educ. (2020), https://doi.org/10.1177/0020720920953126, 0020720920953126.
[14] V. Singh, A. K Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, Inf. Process. Agric 4 (1) (2017) 41–49, https://doi.org/10.1016/j.inpa.2016.10.005.
[15] R. Paul, A.C. Saville, J.C. Hansel, Y. Ye, C. Ball, A. Williams, Q. Wei, Extraction of plant DNA by microneedle patch for rapid detection of plant diseases, ACS Nano 13 (6) (2019) 6540–6549, https://doi.org/10.1021/acsnano.9b00193.
[16] M. Ahmad, M. Abdullah, H. Moon, D. Han, Plant disease detection in imbalanced datasets using efficient convolutional neural networks with stepwise transfer learning, IEEE Access 9 (2021) (2021) 140565–140580, https://doi.org/10.1109/access.2021.3119655.
[17] A. Cellini, S. Blasioli, E. Biondi, A. Bertaccini, I. Braschi, F. Spinelli, Potential applications and limitations of electronic nose devices for plant disease diagnosis, Sensors 17 (11) (2017) 2596, https://doi.org/10.3390/s17112596.
[18] R. Bhagwat, Y. Dandawate, A review on advances in automated plant disease detection, Int. j. eng. innov. technol 11 (4) (2021) 251–264, https://doi.org/10.46604/ijeti.2021.8244.
[19] K. Jha, A. Doshi, P. Patel, M. Shah, A comprehensive review on automation in agriculture using artificial intelligence, Artif Intell Agric 2 (2019) 1–12, https://doi.org/10.1016/j.aiia.2019.05.004.
[20] H. Farman, J. Ahmad, B. Jan, Y. Shahzad, M. Abdullah, A. Ullah, Efficientnet-based robust recognition of peach plant diseases in field images, Comput. Mater. Contin 71 (2022) 2073–2089, https://doi.org/10.32604/cmc.2022.018961.
[21] Y. Guo, H. Zhao, S. Zhang, Y. Wang, D. Chow, Modeling and optimization of environment in agricultural greenhouses for improving cleaner and sustainable crop production, J. Clean. Prod. 285 (2021) 124843, https://doi.org/10.1016/j.jclepro.2020.124843.
[22] J.G.A. Barbedo, Expert systems applied to plant disease diagnosis: survey and critical view, IEEE Lat. Am. Trans 14 (4) (2016) 1910–1922, https://doi.org/10.1109/TLA.2016.7483534.
[23] M.M. Taye, Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions, Computation 11 (3) (2023) 52, https://doi.org/10.3390/computation11030052.
[24] I.H. Sarker, Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions, SN comput. sci 2 (6) (2021) 420, https://doi.org/10.1007/s42979-021-00815-1.
[25] S. Shamshirband, M. Fathi, A. Dehzangi, A.T. Chronopoulos, H. Alinejad-Rokny, H. A, Review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues, J. Biomed. Inform 113 (2021) 103627, https://doi.org/10.1016/j.jbi.2020.103627.
[26] Y. Toda, F. Okura, How convolutional neural networks diagnose plant disease, Plant Phenomics (2019), https://doi.org/10.34133/2019/9237136.
[27] S. Zhang, S. Zhang, C. Zhang, X. Wang, Y. Shi, Cucumber leaf disease identification with global pooling dilated convolutional neural network, Comput. Electron. Agric. 162 (2019) 422–430, https://doi.org/10.1016/j.compag.2019.03.012.
[28] A. Fuentes, S. Yoon, S.C. Kim, D.S. Park, A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition, Sensors 17 (9) (2017) 2022, https://doi.org/10.3390/s17092022.
[29] I. Ahmed, P.K. Yadav, Plant disease detection using machine learning approaches, Expert Syst 40 (5) (2023) e13136, https://doi.org/10.1111/exsy.13136.
[30] R. Sujatha, J.M. Chatterjee, N.Z. Jhanjhi, S.N. Brohi, Performance of deep learning vs machine learning in plant leaf disease detection, Microprocess. Microsyst. 80 (2021) 103615, https://doi.org/10.1016/j.micpro.2020.103615.
[31] M. Agarwal, S.K. Gupta, K.K. Biswas, Development of Efficient CNN model for Tomato crop disease identification, Sustain. Comput.: Inform. Syst 28 (2020), https://doi.org/10.1016/j.suscom.2020.100407.
[32] S. Shamshirband, M. Fathi, A. Dehzangi, A.T. Chronopoulos, H. Alinejad-Rokny, H. A, Review on deep learning approaches in healthcare systems: taxonomies, challenges, and open issues, J. Biomed. Inform 113 (2021) 103627, https://doi.org/10.1016/j.jbi.2020.103627.
[33] S.M. Hassan, A.K. Maji, J. Jasiński, Z. Leonowicz, E. Jasińska, Identification of plant-leaf diseases using CNN and transfer-learning approach, Electronics 10 (12) (2021) 1388, https://doi.org/10.3390/electronics10121388.

[34] K.P. Ferentinos, K.P. Deep, Learning models for plant disease detection and diagnosis, Comput. Electron. Agric. 145 (2018) 311–318, https://doi.org/10.1016/j.compag.2018.01.009.

[35] E.C. Too, L. Yujian, S. Njuki, L. Yingchun, A comparative study of fine-tuning deep learning models for plant disease identification, Comput. Electron. Agric. 161 (2019) 272–279, https://doi.org/10.1016/j.compag.2018.03.032.

[36] L. Bi, G. Hu, Improving image-based plant disease classification with generative adversarial network under limited training set, Front. Plant Sci. 11 (2020) 583438, https://doi.org/10.3389/fpls.2020.583438.

[37] M.G. Lanjewar, K.G. Panchbhai, Convolutional neural network-based tea leaf disease prediction system on smart phone using paas cloud, Neural Comput. Appl. 35 (3) (2023) 2755–2771, https://doi.org/10.1007/s00521-022-07743-y.

[38] J. Chen, A. Zeb, Y.A. Nanehkaran, D. Zhang, Stacking ensemble model of deep learning for plant disease recognition, J Ambient Intell Humaniz Comput (2022) 1–14, https://doi.org/10.1007/s12652-022-04334-6.

[39] M.S. Al-Gaashani, F. Shang, A.A. Abd El-Latif, Ensemble learning of lightweight deep convolutional neural networks for crop disease image detection, J. Circuits, Syst. Comput 32 (5) (2023) 2350086, https://doi.org/10.1142/S021812662350086X.

[40] P. Karpov, G. Godin, I.V. Tetko, Transformer-CNN: Swiss knife for QSAR modeling and interpretation, J. Cheminf. 12 (1) (2020) 1–12, https://doi.org/10.1186/s13321-020-00423-w.

[41] Q. Xu, M. Zhang, Z. Gu, G. Pan, Overfitting remedy by sparsifying regularization on fully connected layers of CNNs, Neurocomputing 328 (2019) 69–74, https://doi.org/10.1016/j.neucom.2018.03.080.

[42] G. Geetharamani, A. Pandian, Identification of plant leaf diseases using a nine-layer deep convolutional neural network, Comput. Electr. Eng. 76 (2019) 323–338, https://doi.org/10.1016/j.compeleceng.2019.04.011.

[43] N. Ganatra, A. Patel, Performance analysis of fine-tuned convolutional neural network models for plant disease classification, Int. J. Control Autom. Syst. 13 (3) (2020) 293–305. http://sersc.org/journals/index.php/IJCA/article/view/24998.

[44] M. Durgadevi, Generative adversarial network (gan): a general review on different variants of gan and applications, in: 2021 6th International Conference on Communication and Electronics Systems (ICCES), IEEE, 2021, pp. 1–8, https://doi.org/10.1109/icces51350.2021.9489160.

[45] R. Katafuchi, T. Tokunaga, Image-based plant disease diagnosis with unsupervised anomaly detection based on reconstructability of colors, arXiv preprint arXiv:2011.14306 (2020), https://doi.org/10.48550/arXiv.2011.14306.

[46] S. Sookpong, T. Kasetkasem, T. Phatrapornnant, J. Yu, A unhealthy plant identification system using a generative adversarial network, in: 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2022, pp. 1–4, https://doi.org/10.1109/ecti-con54298.2022.9795542.

[47] Y. He, K.P. Seng, L.M. Ang, Generative adversarial networks (GANs) for audio-visual speech recognition in artificial intelligence IoT, Information 14 (10) (2023) 575.

[48] A. Stephen, A. Punitha, A. Chandrasekar, Optimal deep generative adversarial network and convolutional neural network for rice leaf disease prediction, Vis. Comput. 40 (2) (2024) 919–936, https://doi.org/10.1007/s00371-023-02823-z.

[49] X. Wang, W. Cao, GACN: generative adversarial classified network for balancing plant disease dataset and plant disease recognition, Sensors 23 (15) (2023) 6844, https://doi.org/10.3390/s23156844.

[50]] Y. Song, H. Zhang, J. Li, R. Ye, X. Zhou, B. Dong, L. Li, High-accuracy maize disease detection based on attention generative adversarial network and few-shot learning, Plants 12 (17) (2023) 3105, https://doi.org/10.3390/plants12173105.

[51] Q. Cao, N. Balasubramanian, A. Balasubramanian, MobiRNN: efficient recurrent neural network execution on mobile GPU, in: Proceedings of the 1st International Workshop on Deep Learning for Mobile Systems and Applications, 2017, pp. 1–6, https://doi.org/10.48550/arXiv.1706.00878.

[52] L. Mou, P. Ghamisi, X.X. Zhu, Deep recurrent neural networks for hyperspectral image classification, IEEE Trans Geosci Remote Sens 55 (7) (2017) 3639–3655, https://doi.org/10.1109/TGRS.2016.2636241.

[53] S.H. Lee, H. Goëau, P. Bonnet, A. Joly, Attention-based recurrent neural network for plant disease classification, Front. Plant Sci. 11 (2020) 601250, https://doi.org/10.3389/fpls.2020.601250.

[54] M. Nandhini, K.U. Kala, M. Thangadarshini, S.M. Verma, Deep Learning model of sequential image classifier for crop disease detection in plantain tree cultivation, Comput. Electron. Agric. 197 (2022) 106915, https://doi.org/10.1016/j.compag.2022.106915.

[55] T. Daniya, S. Vigneshwari, Deep neural network for disease detection in rice plant using the texture and deep features, Comput. J. 65 (7) (2022) 1812–1825, https://doi.org/10.1093/comjnl/bxab022.

[56] E.A. Pool, J.F. Kooij, D.M. Gavrila, Context-based cyclist path prediction using recurrent neural networks, in: 2019 IEEE Intelligent Vehicles Symposium (IV), 2019, pp. 824–830, https://doi.org/10.1109/IVS.2019.8813889.

[57] S. Tanwar, J. Singh, ResNext50 based convolution neural network-long short term memory model for plant disease classification, Multimed. Tools. Appl (2023) 1–19, https://doi.org/10.1007/s11042-023-14851-x.

[58] M. Shoaib, B. Shah, S. Ei-Sappagh, A. Ali, A. Ullah, F. Alenezi, F. Ali, An advanced deep learning models-based plant disease detection: a review of recent research, Front. Plant Sci. 14 (2023) 1158933, https://doi.org/10.3389/fpls.2023.1158933.

[59] S. Bodapati, H. Bandarupally, M. Trupthi, COVID-19 time series forecasting of daily cases, deaths caused and recovered cases using long-short term memory networks, in: 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 525–530, https://doi.org/10.1109/ICCCA49541.2020.9250863.

[60] Y.L. Kong, Q. Huang, C. Wang, J. Chen, J. Chen, D. He, Long short-term memory neural networks for online disturbance detection in satellite image time series, Remote Sens 10 (3) (2018) 452, https://doi.org/10.3390/rs10030452.

[61] A. Rehmer, A. Kroll, On the vanishing and exploding gradient problem in Gated Recurrent Units, IFAC-PapersOnLine 53 (2) (2020) 1243–1248, https://doi.org/10.1016/j.ifacol.2020.12.1342.

[62] C. Subakan, M. Ravanelli, S. Cornell, M. Bronzi, J. Zhong, Attention is all you need in speech separation, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 21–25, https://doi.org/10.1109/icassp39728.2021.9413901.

[63] M. Abdel-Nasser, K. Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN, Neural Comput. Appl. 31 (2019) 2727–2740, https://doi.org/10.1007/s00521-017-3225-z.

[64] X. Zhang, C. Zhong, J. Zhang, T. Wang, W.W. Ng, Robust recurrent neural networks for time series forecasting, Neurocomputing 526 (2023) 143–157, https://doi.org/10.1016/j.neucom.2023.01.037.

[65] P. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, in: International Conference on Machine Learning, 2013, pp. 1310–1318, https://doi.org/10.48550/arXiv.1211.5063.

[66] S. Parez, N. Dilshad, N.S. Alghamdi, T.M. Alanazi, J.W. Lee, Visual intelligence in precision agriculture: exploring plant disease detection via efficient vision transformers, Sensors 23 (15) (2023) 6949, https://doi.org/10.3390/s23156949.

[67] S. Yu, L. Xie, Q. Huang, Inception convolutional vision transformers for plant disease identification, Internet of Things 21 (2023) 100650, https://doi.org/10.1016/j.iot.2022.100650.

[68] H.T. Thai, K.H. Le, N.L.T. Nguyen, FormerLeaf: an efficient vision transformer for Cassava Leaf Disease detection, Comput. Electron. Agric. 204 (2023) 107518, https://doi.org/10.1016/j.compag.2022.107518.

[69] S. Khan, M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, M. Shah, Transformers in vision: a survey, ACM Comput. Surv. 54 (10s) (2022) 1–41, https://doi.org/10.1145/3505244.

[70] Y.A. Nanehkaran, D. Zhang, J. Chen, Y. Tian, N. Al-Nabhan, Recognition of plant leaf diseases based on computer vision, J Ambient Intell Humaniz Comput (2020) 1–18, https://doi.org/10.1007/s12652-020-02505-x.

[71] J. Chen, W. Chen, A. Zeb, D. Zhang, Y.A. Nanehkaran, Crop pest recognition using attention-embedded lightweight network under field conditions, Appl. Entomol. Zool 56 (2021) 427–442, https://doi.org/10.1007/s13355-021-00732-y.

[72] S. Tanwar, J. Singh, ResNext50 based convolution neural network-long short-term memory model for plant disease classification, Multimed. Tools. Appl 82 (2023) 29527–29545, https://doi.org/10.1007/s11042-023-14851-x.
[73] B. Rajalakshmi, S. Kumar, B.K. Devi, B.P. Kavin, G.H. Seng, Single and multi-crop species disease detection using ITSO based gated recurrent multi-attention neural network, J. auton. intell 7 (4) (2024) 1–18, https://doi.org/10.32629/jai.v7i4.1126.
[74] P.S. Thakur, S. Chaturvedi, P. Khanna, T. Sheorey, A. Ojha, Vision transformer meets convolutional neural network for plant disease classification, Ecol Inform 77 (2023) 102245, https://doi.org/10.1016/j.ecoinf.2023.102245.
[75] B.S. Abu-Nasser, S.S. Abu Naser, Rule-based system for watermelon diseases and treatment, International. J. Acad. Inf. Syst. Res 2 (7) (2018) 1–7. https://ssrn.com/abstract=3242597.
[76] M.A. Alajrami, S.S. Abu-Naser, Onion rule based system for disorders diagnosis and treatment, Int. J. Acad. Pedagog. Res 2 (8) (2018) 1–9. http://dstore.alazhar.edu.ps/xmlui/handle/123456789/360.
[77] X. Gu, P.P. Angelov, Highly interpretable hierarchical deep rule-based classifier, Appl. Soft Comput. 92 (2020) 106310, https://doi.org/10.1016/j.asoc.2020.106310.
[78] K. Khandelwal, D.P. Sharma, Hybrid reasoning model for strengthening the problem-solving capability of expert systems, Int. J. Adv. Comput. Sci. Appl. 4 (10) (2013) 88–94, https://doi.org/10.14569/IJACSA.2013.041014.
[79] F.O. Isinkaye, J. Soyemi, S.O. Awosupin, A mobile based expert system for disease diagnosis and medical advice provisioning, Int. J. Comput. Sci. Inf. Secur. 15 (1) (2017) 568–572.
[80] P. Akulwar, A recommended system for crop disease detection and yield prediction using machine learning approach, Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries (2020) 141–163, https://doi.org/10.1002/9781119711582.ch8.
[81] Y. Afoudi, M. Lazaar, M. and Al Achhab, Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network, Simul Model Pract Theory 113 (2021) 102375, https://doi.org/10.1016/j.simpat.2021.102375.
[82] F.O. Isinkaye, E.D. Erute, A smartphone-based plant disease detection and treatment recommendation system using machine learning techniques, Transactions on Machine Learning and Artificial Intelligence 10 (1) (2022) 1–8, https://doi.org/10.14738/tmlai.101.11313.
[83] J. Xu, B. Xu, S. Men, Feature-based similarity retrieval in Content-based image retrieval, in: 2010 Seventh Web Information Systems and Applications Conference, IEEE, 2010, pp. 215–219, https://doi.org/10.1109/wisa.2010.46.
[84] N. Patil, S. Kelkar, M. Ranawat, M. Vijayalakshmi, Krushi sahyog: plant disease identification and crop recommendation using artificial intelligence, in: 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1–6, https://doi.org/10.1109/incet51464.2021.9456114.
[85] H. Wang, Y. Zhao, Q. Wang, B. Zhou, Latent factor models fusing user & item attributes, in: IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 3201–3206, https://doi.org/10.1109/ssci44817.2019.9002724.
[86] J. Son, S.B. Kim, Content-based filtering for recommendation systems using multiattribute networks, Expert Syst. Appl. 89 (2017) 404–412, https://doi.org/10.1016/j.eswa.2017.08.008.
[87] M. Subramanian, K. Shanmugavadivel, P.S. Nandhini, On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves, Neural Comput. Appl. 34 (16) (2022) 13951–13968, https://doi.org/10.1007/s00521-022-07246-w.
[88] S. Wang, C. Xu, A.S. Ding, Z. Tang, A novel emotion-aware hybrid music recommendation method using deep neural network, Electronics 10 (15) (2021) 1769, https://doi.org/10.3390/electronics10151769.
[89] R. Zhu, X. Tu, J.X. Huang, Deep learning on information retrieval and its applications, in: Deep Learning for Data Analytics, Academic Press, 2020, pp. 125–153, https://doi.org/10.1016/B978-0-12-819764-6.0000.
[90] M. Kashif, G. Raja, F. Shaukat, An efficient content-based image retrieval system for the diagnosis of lung diseases, J. Digit. Imag. 33 (4) (2020) 971–987, https://doi.org/10.1007/s10278-020-00338-w.
[91] M. Sharma, R. Mittal, A. Bharati, D. Saxena, A.K. Singh, Emotional information-based hybrid recommendation system, in: Soft Computing for Problem Solving: Proceedings of the SocProS 2022, 2023, pp. 249–267, https://doi.org/10.1007/978-981-19-6525-8_20.
[92] A. Smetanin, A. Uzhinskiy, G. Ososkov, P. Goncharov, A. Nechaevskiy, Deep learning methods for the plant disease detection platform, AIP Conf. Proc. 2377 (1) (2021), https://doi.org/10.1063/5.0068797 (2021.
[93] A. Maheswari, K. Arunesh, A recommendation system based on COVID-19 prediction & analyzing using ensemble boosted machine learning algorithm, SN comput. sci 4 (5) (2023) 478, https://doi.org/10.1007/s42979-023-01907-w.
[94] Z. Zhao, Z. Zhao, Q. Li, J. Zhuang, X. Ju, Multi-modal cross-attention learning on detecting anomalies of electrical equipment, in: 2023 International Conference on Pattern Recognition, Machine Vision and Intelligent Algorithms (PRMVIA), 2023, pp. 137–142, https://doi.org/10.1109/prmvia58252.2023.00030.
[95] I. Afyouni, Z. Al Aghbari, R.A. Razack, Multi-feature, multi-modal, and multi-source social event detection: a comprehensive survey, Inf. Fusion 79 (2023) 279–308, https://doi.org/10.1016/j.inffus.2021.10.013.
[96] H. Raghavan, O. Madani, R. Jones, Active learning with feedback on features and instances, J. Mach. Learn. Res. 7 (2006) 1655–1686. https://dl.acm.org/doi/10.5555/1248547.1248608.
[97] A. Tamkin, D. Nguyen, S. Deshpande, J. Mu, N. Goodman, Active learning helps pretrained models learn the intended task, Adv. Neural Inf. Process. Syst. 35 (2022) 28140–28153, https://doi.org/10.48550/arXiv.2204.08491.
[98] M. Shoaib, B. Shah, S. Ei-Sappagh, A. Ali, A. UllahF.Alenezi, F. Ali, An advanced deep learning models-based plant disease detection: a review of recent research, Front. Plant Sci. 14 (2023) 1158933, https://doi.org/10.3389/fpls.2023.1158933.
[99] R. Thapa, N. Snavely, S. Belongie, A. Khan, The plant pathology 2020 challenge dataset to classify foliar disease of apples, arXiv preprint arXiv:2004.11958, https://doi.org/10.48550/arXiv.2004.11958, 2020.
[100] G. Fenu, F.M. Malloci, DiaMOS plant: a dataset for diagnosis and monitoring plant disease, Agronomy 11 (11) (2021) 2107, https://doi.org/10.3390/agronomy11112107.
[101] A. Ahmad, D. Saraswat, A. El Gamal, A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools, Smart Agric. Technol 3 (2023) 100083, https://doi.org/10.1016/j.atech.2022.100083.
[102] J. Parraga-Alava, K. Cusme, A. Loor, E. Santander, E. RoCoLe, A robusta coffee leaf images dataset for evaluation of machine learning based methods in plant diseases recognition, Data Brief 25 (2019) 104414, https://doi.org/10.1016/j.atech.2022.100083.
[103] L.M. Tassis, J.E.T. de Souza, R.A. Krohling, A deep learning approach combining instance and semantic segmentation to identify diseases and pests of coffee leaves from in-field images, Comput. Electron. Agric. 186 (2021) 106191, https://doi.org/10.1016/j.compag.2021.106191.
[104] A. Dutta, A. Zisserman, The VIA annotation software for images, audio and video, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 2276–2279, https://doi.org/10.1145/3343031.3350535.
[105] H.B. Prajapati, J.P. Shah, J.P. And, V.K. Dabhi, Detection and classification of rice plant diseases, Intell. Decis. Technol. 11 (3) (2017) 357–373, https://doi.org/10.3233/IDT-170301.
[106] S. Jiang, J. Song, Evaluation metrics for personalized recommendation systems, J. Phys. Conf. 1920 (1) (2021) 012109, https://doi.org/10.1088/1742-6596/1920/1/012109.
[107] S.S. Islam, M.S. Haque, M.S.U. Miah, T.B. Sarwar, R. Nugraha, Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study, PeerJ Comput. Sci 8 (2022) e898, https://doi.org/10.7717/peerj-cs.898.
[108] Y.M. Afify, I.F. Moawad, N.L. Badr, M.F. Tolba, An enhanced distance-based similarity measure for user-based recommendations, in: Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016, vol. 2, 2017, pp. 42–52, https://doi.org/10.1007/978-3-319-48308-5_5.
[109] M. Heydarian, T.E. Doyle, R. Samavi, MLCM: multi-label confusion matrix, IEEE Access 10 (2022) 19083–19095, https://doi.org/10.1109/access.2022.3151048.