Expert Systems With Applications 237 (2024) 121481

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

CoffeeNet: A deep learning approach for coffee plant leaves diseases recognition

Marriam Nawaz [a,1], Tahira Nazir [b,2], Ali Javed [a,3,*], Sherif Tawfik Amin [c,4], Fathe Jeribi [c,5], Ali Tahir [c,6]

[a] *Department of Software Engineering, University of Engineering and Technology-Taxila, 47050 Punjab, Pakistan*
[b] *Department of Computer Science, Riphah International University, Gulberg Green Campus Islamabad, Pakistan*
[c] *College of Computer Science and Information Technology Jazan University, Jazan 45142, Saudi Arabia*

A R T I C L E I N F O

Keywords:
CenterNet
Coffee plant disease
Classification
Deep learning
ResNet

A B S T R A C T

Coffee is regarded as the highest consumed drink around the globe and has accounted as a major source of income in the regions where it is cultivated. To meet the coffee marketplace's requirements around the globe, cultivators must boost and analyze its cultivation and quality. Several factors like environmental changes and plant diseases are the major hindrance to increasing the yield of coffee. The development in the field of computer vision has facilitated the earliest diagnostic of diseased plant samples, however, the incidence of various image distortions i.e., color, light, size, orientation changes, and similarity in the healthy and diseased portions of examined samples are the major challenges in the effective recognition of various coffee plant leaf infections. The proposed work is focused to overwhelm the mentioned limitations by proposing a novel and effective DL model called the CoffeeNet. Explicitly, an improved CenterNet approach is proposed by introducing spatial-channel attention strategy-based ResNet-50 model for the computation of deep and disease-specific sample characteristics which are then classified by the 1-step detector of the CenterNet framework. We investigated the localization and cataloging outcomes of the suggested method on the Arabica coffee leaf repository which contains the images captured in the more realistic and complicated environmental constraints. The CoffeeNet model acquires a classification accuracy number of 98.54%, along with an mAP of 0.97 that is presenting the usefulness of our technique in localizing and categorizing various sorts of coffee plant leaf disorders.

1. Introduction

The production of coffee, the most extensively traded commodity in the tropics, is produced by up to 25 million household farmers, who together account for up to 80% of global production. According to stats reported till 2021, Brazil was the highest exporter country of coffee, where Indonesia and India were in second, and third numbers, respectively (Aufar, Abdillah, & Romadoni, 2023). Moreover, studies indicate that Sumatra has produced the finest quality coffee yield ("ICO, Monthly coffee market report,"). The improvement of the agriculture industry including coffee production can be very beneficial for a nation's economy (Chang & Huang, 2021). Many other countries around the globe have put extraordinary effort into the area of agriculture to boost both the quantity and quality of various plants including coffee. It is impressive how Saudi Arabia's agriculture has developed recently, especially in Jazan, where huge deserts have been turned into farmland and numerous plants are being grown. Specifically, Jazan is famous for cultivating coffee plants that harvest the finest Khoulani coffee, one of

* Corresponding author.
E-mail addresses: marriam.nawaz@uettaxila.edu.pk (M. Nawaz), tahira.nazir@riphah.edu.pk (T. Nazir), ali.javed@uettaxila.edu.pk (A. Javed), samin@jazanu.edu.sa (S. Tawfik Amin), fjeribi@jazanu.edu.sa (F. Jeribi), agulsher@jazanu.edu.sa (A. Tahir).
[1] 0000-0002-2238-4645.
[2] 0000-0001-8130-3721.
[3] 0000-0002-1290-1477.
[4] 0009-0001-3788-0414.
[5] 0000-0002-8511-8002.
[6] 0000-0001-5685-3107.

https://doi.org/10.1016/j.eswa.2023.121481
Received 21 June 2023; Received in revised form 26 August 2023; Accepted 4 September 2023
Available online 7 September 2023

the highest-quality coffees in the globe. Besides, the majority of coffee is produced in poor countries, where it is an important means of income and contributes significantly to worldwide export revenues. Coffee is among the highest consumed beverages in the globe and one of the top frequently traded goods [3], with a market that is constantly expanding due to increasing consumption in developing countries and its significant impact on specialized and modern products in established nations. However, the world's rapidly changing environment is bringing in a number of coffee plant diseases that are harming its production quality and quantity and, eventually, lowering the revenue. The agricultural industry's reliance on people to physically examine the spots to identify such diseases presents a greater obstacle to coffee plant disease diagnosis as this process highly relies on the availability of domain experts which slowdowns the detection procedure (Jepkoech, Mugo, Kenduiywo, & Too, 2021). Moreover, such manual examination of plants is vulnerable to errors. Therefore, there exists a demand to develop an effective and reliable automated approach capable of locating and differentiating among various infections of coffee plant leaves.

The Fourth Industrial Revolution (FIR), the turning point of modern technological advancement, is the period in which physical and digital mechanisms can be interconnected via using advanced methods like deep learning (DL), machine learning (ML), artificial intelligence (AI), and big data (Hoosain, Paul, & Ramakrishna, 2020). Additionally, the FIR can boost production and expansion in a number of ways. Farming represents one of the industries undergoing scientific expansion and advancement, where innovations like AI, DL, and ML have an advantageous effect on agrarian expansion and manufacturing capacity (Dhanaraju, Chenniappan, Ramalingam, Pazhanivelan, & Kaliaperumal, 2022). An emerging technique called "smart agriculture" combines cutting-edge methods for raising crop output while simultaneously boosting inputs to farming in an ecologically sound way. Today, it is possible to cut expenditures and mistakes in order to attain environmentally and financially equitable farming (Hitimana & Gwun, 2014). Recent years have seen a number of initiatives to employ AI to aid growers in precisely identifying illnesses and pests that harm the farming industry, as well as to assess the nature of signs of various plant infections. By imitating human thinking operations, AI aims to give computers an understanding resembling that of humans. It presents new information while allowing for investigation, acquiring knowledge, and resolving issues. AI can revolutionize agribusiness by enabling producers to generate more effective outcomes with less effort while also offering a wide range of extra advantages (Albahli & Nawaz, 2022). ML applications have significantly increased in AI research during the past few years, especially the latest generation of models termed DL. Particularly, DL approaches have outperformed conventional ML technologies in several fields (Nawaz, Javed, & Irtaza, 2022; Nawaz, et al., 2023) due to their ability to effectively capture the structural information of a sample. In the previous few years, a variety of approaches have been employed to detect illnesses in plants, and DL techniques have been regarded as highly effective. Due to the promising performance of various DL methods in agriculture, several investigations have been carried out, proving that image-based evaluation is trustworthy for recognizing various types of plant disorders (Albattah, Nawaz, Javed, Masood, & Albahli, 2022). Many researchers have utilized various DL models like convolutional neural networks (CNN) for diagnosing various types of plant infections including coffee plant leaf abnormalities. CNNs are among the most optimistic DL-based strategies for autonomously discerning characteristics and effective model tuning, that employ numerous convolutional layers to encode various learning keypoints depending on suspected samples.

Vast volumes of pictorial samples are required for network modeling in DL, which is a downside (Saleem, et al., 2022). For instance, if there are fewer visuals in the databases, CNN's classification results suffer. Usually, transfer learning is employed to overcome this crucial limitation by offering a number of benefits, among them is that it doesn't need a lot of samples for network tuning because prior acquired knowledge from related activities can be applied to the present one. Although, significant works have been introduced for the timely recognition of coffee plant leaves, however, there is much room for improvement. Like many works employed the concept of applying object detection methods for locating and differentiating several types of plant leaf infections, however, little attention is paid to the employment of such approaches for recognizing the diseases of coffee plants. Further, historic approaches are not much competent to perform well for unseen cases and lack generalization ability. Moreover, approaches lack to diagnose numerous coffee plant infections or to identify multiple occurrences of an abnormality on a single image (Paul, et al., 2020). Further, the extensive resemblance in the structural information of various illnesses complicates the classification task. The digital samples are subject to various types of distortions as well like containing light variations, clutter, blurring, color, and angle alterations which also introduces a challenge to the effective determination of the diseased plant leaf portions. There also exists a huge variation in the size of both coffee plant leaves and associated infected portions which is also a major challenge to this field.

Agricultural losses can be kept under control by quickly identifying the coffee plant leaf virus reason and allowing the fast choice of the appropriate preventative measure. It also stands for the first and most important step in avoiding such infections. Our goal is to design an approach that can properly locate and categorize coffee plant leaf illnesses. An earlier diagnosis of such infections can lead to more curative measures and longer survival times. Therefore, an effort is put into the introduced work to overwhelm the above-mentioned problems of coffee plant leaf abnormalities recognition by suggesting an improved DL method called CoffeeNet. We have proposed an improved CenterNet approach by introducing an enhanced keypoints estimation ResNet-50 framework with an attention mechanism to extract the reliable and precise features of various coffee plant leaf disorders. The redesigned feature extractor base enhances the model's capacity to collect significant sample information in the context of complex backdrops and unpredictable environmental constraints. The proposed approach effectively operates in real-world conditions while concurrently performing identification and categorization utilizing an end-to-end training strategy. The nominative findings of the proposed technique are as follows:

1) We introduced a novel model called CoffeeNet with enhanced keypoints extraction CNN backbone for reliable identification and classification of coffee plant leaf syndromes.
2) A spatial and channel attention strategy is proposed in the feature engineering stage that computes inter-channel associations and pixels incorporations to nominate exact diseased regions of examined images with complex environment settings.
3) Improved determination of both diseased areas and associated classes of coffee plant leaves because of the better recognition capability of the proposed work.
4) A computationally efficient approach is suggested to recognize the various coffee plant leaf disorders as CenterNet utilized one step object recognition strategy.
5) We accomplished a vast assessment of the introduced method on a complex and publicly accessible data repository to show the robustness of our model. The attained model performance in the aspect of both numeric scores and pictorial representation ensures the improved results of our work even in complicated background settings like noise, blur attacks, and in alterations of lighting, and other sample distortions.

Our remaining manuscript follows the following section arrangements: Section 2 comprises related work, while the proposed approach is discussed in Section 3. Section 4 contains the result discussion, and the conclusion is explained in Section 5.

2. Related work

Various diseases have been affecting the yield of coffee production for years, therefore, the research community has put a huge effort to propose computerized solutions for the appropriate diagnostic of such abnormalities from the plants. The automated systems are based on the latest ML, and DL approaches to perform the identification of different coffee plant leaf infections. We investigate the existing approaches in this section to provide an overview of already performed work in this domain. One such work was proposed in (Alexandre Pereira Marcos, Natan Luis Silva Rodovalho, & André Ricardo Backes, 2019b) to locate the healthy and rust-effected coffee leaf images. The genetic procedure was applied to calculate the pertinent group of sample attributes focusing on the structure and colored aspects of the infected regions of samples to execute the categorization job. The approach (Marcos, et al., 2019b) is proficient to recognize the affected samples of varying colored attributes, however, classification accuracy needs more enhancement. Gutte et al. (Gutte & Gitte, 2016) employed 3 steps to recognize the monocot and dicot infections of coffee plant leaves. Firstly, the k-mean method was used to perform the segmentation of examined sample from which next the structural, chrominance, and shape information was computed to be passed as the feature vector to the next phase. Last, the classification was performed with the support vector machine (SVM) predictor. The work (Gutte & Gitte, 2016) performs well in classifying the diseased samples, however, not proficient in handling the distorted image samples. Esgarioa et al. (Esgario, Krohling, & Ventura, 2020) discussed a DL framework to measure the level of plant leaf infections caused by different diseases. The work (Esgario, et al., 2020) used a residual model to extract the dense information of examined images and apply the classification step. The method (Esgario, et al., 2020) improves the recognition ability to recognize different coffee plant leaf diseases, however, degradation in classification performance has been witnessed for images with multiple signs of abnormalities. An approach was discussed in [4] to classify the healthy and rust-affected samples. In (Marin, et al., 2021), structural and colored information of suspected images was computed to form the feature vector. Next, three different variants of the decision tree (DT) were used to distribute the input images into 2 groups. This work (Marin, et al., 2021) reports the best classification results for the Logistic DT, however, the model lacks to execute effectively for unseen cases.

Dogan et al. (Dogan, et al., 2023) proposed a hybrid technique to recognize the various classes of coffee bean abnormalities. First, a DL framework called GoogLeNet was applied to extract the dense set of sample attributes. Next, the extreme learning machine (ELM) was designated to achieve the categorization of samples. The work (Dogan, et al., 2023) also applied various swarm intelligence-based optimization approaches with the ELM predictor to boost the performance of model behavior. This work (Dogan, et al., 2023) shows better results, however, severe classification performance degradation has been reported for infected samples from different classes with a huge resemblance in the structure of the diseased portion. Another similar work was discussed in (Ruttanadech, et al., 2023) where various ML classifiers like SVM, KNN, linear discriminant analysis (LDA), DT, quadratic discriminant analysis (QDA), and Naive Bayes (NB) were designated for the categorization of different coffee bean abnormalities. The approach (Ruttanadech, et al., 2023) reports the best results for the DT method, however, the technique lacks to diagnose the early signs of coffee bean infections. Hasan et al. (Hasan, Yusuf, Rahim, & Alzubaidi, 2022) also employed a DL strategy to recognize the normal and infected leaf images of coffee plants. A sample augmentation step was used to enhance the sample size. Next, the approach (Hasan, et al., 2022) utilized the graph-cut technique to perform the segmentation of examined images, on which the color examination was utilized to differentiate the normal and diseased samples. This work (Hasan, et al., 2022) is also evaluated for the apple plant leaf diseases, however, the approach lacks to tackle the overlapped infection in a given image to tackle the presence of several infections in a single sample. A work was elaborated in (Yamashita & Leite, 2023) that utilized the MobileNet model to extract the sample information and perform the classification to locate various types of coffee plant leaf disorders. An augmentation phase was also used to increase the size of the data sample. This approach performs well in recognizing different coffee plant leaf infections; however, classification results degrade for unseen samples. Further, the work (Yamashita & Leite, 2023) also needs huge training data and is unable to tackle the presence of different diseased signs in a single sample. Lisboa et al. (Lisboa, Lima, & Queiroz, 2021) designed a CNN model to distribute the samples into two groups either normal or rust-affected images. After performing a preprocessing phase on the input images on which the designed CNN method was used to do the categorization job. The approach (Lisboa, et al., 2021) requires further enhancement in the classification results. Novtahaning et al. (Novtahaning, Shah, & Kang, 2022) presented a framework where numerous DL approaches were used to perform the classification of coffee plant leaf diseases. Further, an ensemble approach was also designed merging the features of three differing DL approaches to extract a dense set of sample characteristics. The work (Novtahaning, et al., 2022) reports effective results for ensemble technique, however, with an increased computing burden. A technique was elaborated in (Alexandre Pereira Marcos, Natan Luis Silva Rodovalho, & André R Backes, 2019a) that introduced a CNN framework to detect rust disease from the plant leaves of coffee. The work also cross-verified the attained results from domain experts. This approach (Marcos, et al., 2019a) shows better results in recognizing the coffee plant leaf infections, however, unable to better tackle the distorted images. Dann et al. (Javierto, Martin, & Villaverde, 2021) utilized a DL model called the YOLO-V3 framework to locate and categorize the infected regions from the examined images of the coffee plant. The work (Javierto, et al., 2021) is effective for coffee plant leaf illness diagnosis, however, unable to locate the small size infected regions. The work (Ventura, ESGARIO, & KROHLING, 2020) also employed different pertained DL approaches to categorizing the normal and infected samples of coffee plants. This work attained the best results for the residual approach; however, classification results need further improvements. Chang et al. (Chang & Huang, 2021) proposed an improved DL framework to distribute the coffee bean samples into numerous categories. First, a preprocessing step was applied to enhance the graphic representation of instances. Next, an improved AlexNet approach was suggested for dense sample characteristics computation and to execute the cataloging job. The method (Chang & Huang, 2021) shows improved generalization ability, however, classification performance needs evaluation for distorted samples as well.

Hsia et al. (Hsia, Lee, & Lai, 2022) also provided a DL approach for classifying coffee bean diseases. For this, firstly, a preprocessing stage was performed for improving the pictorial appearance of examined samples. Next, the processed samples were passed through an image resizing phase to meet the model requirements. Then, a dense CNN was proposed that employed the resized samples as input and accomplished the classification task by extracting the essential sample characteristics and assigning them to related groups. The approach (Hsia, et al., 2022) performs well for coffee bean infection recognition, however, unable to locate the early diseased signs effectively. Paulos et al. (Paulos & Woldeyohannis, 2022) designed a CNN approach to categorize the infections from the leaves images of coffee plants. After accomplishing a preprocessing phase, the designed CNN model was used to capture the detailed characteristics of examined samples and distribute them into relevant groups. The work (Paulos & Woldeyohannis, 2022) presents an efficient approach for coffee plant lead disorders categorization, however, the model needs evaluation on a complicated dataset. Ayikpa et al. (Ayikpa, Mamadou, Gouton, & Adou, 2022) elaborated an approach in which an improved CNN model is designed to recognize the input coffee plant leaf samples as being healthy or rust-affected. First, a preprocessing stage was applied to enrich the representation of examined samples. Next, the designed CNN approach was evaluated on the data sample and compared with different ML classifiers. The results (Ayikpa, et al., 2022) state that the suggested CNN model improves the classification results, however, the model needs evaluation on a large and diverse data sample to better explain its recall ability. Chowdhury et al. (Chowdhury & Burhan, 2021) also performed an analysis to recognize the infections of coffee plants. For this reason, the GIST feature descriptor was utilized to get the visual characteristics of samples. The computed features were used with numerous ML predictors like SVM, NB, and KNN methods to distribute the images in the relevant classes. This work (Chowdhury & Burhan, 2021) shows improved performance with the SVM predictor, however, the work requires the assessment of a complicated database. Based on the extensive analysis of already performed work on the diagnosis of coffee plant leaves it can be seen that even though huge effort has been put in by the scientist for the early recognition of various sample disorders, however, still there is a requirement for a better strategy. Existing approaches lack to perform well for distorted samples or are unable to locate multiple categories of coffee plant leaves infections robustly. Moreover, the presence of alterations in the color, size, mass, and alignment of infected regions also introduces challenges that need further investigation.

3. Proposed methodology: CoffeeNet

Our presented method called CoffeeNet consists of the following two major phases namely features extraction and recognition. Fig. 1 explains the complete flow and functionality of our proposed model. In the first phase, we prepared our dataset according to the model's requirements and perform annotation operation which is necessary for model training. The annotation process is essential for model training in which we have highlighted the affected regions in the form of a bounding box. The other phase comprises the improved deep learning model CenterNet with ResNet-50 having a Convolutional Block Attention module. From Fig. 1, it can be seen that the ResNet-50 with an attention unit is applied for deep feature calculation from the input images. As input, we passed the images and bounding box information to the network for model learning. After downsampling, the heatmap head identifies potential object center points for each class, the dimension head predicts the dimensions of the bounding box around each detected object, and the offset head refines the position of the bounding box based on the object's center. By combining the predictions from all three heads and performing post-processing steps like non-maximum suppression, the CoffeeNet is trained to recognize the affected regions of coffee plants. After the test phase, we employed evaluation metrics for the calculation of results.

3.1. Data preparation and annotations

In the initial step, we performed mandatory steps for dataset preparation and generated annotations from input samples with the help of available ground truths. For better training, it is essential to accurately identify the location and class of specified regions in the images. To generate the annotations, we used the freely available tool LabelImg [27], which gives us the details in the form of XML files along with each image.

3.2. CenterNet

For precise recognition of diseases or regions, it is essential to effectively calculate the deep features from images. However, feature extraction is still a challenging task due to some reasons: techniques may outcome in over-fitting by using the larger feature sets, and ii) the model can miss learning in the case of a smaller feature set.

It is essential to use an automatic keypoints engineering method rather than hand-crafted feature calculation to produce a discriminative and robust set of features. Due to the significant variations in the size, texture, color, and location of lesions, models using hand-crafted features are not reliable for appropriately diagnosing plant diseases. We used CoffeeNet, a DL-based framework, to address the issues because it can instantly calculate the useful features. The CenterNet convolution filters analyze the structure of the suspicious image to calculate its important points. The CenterNet [26] approach is preferred for identifying Coffee leaf diseases over RCNN, Fast-RCNN [28], Faster-RCNN [15], and [29] because these approaches classify objects using a 2-step object locator. In these techniques, to find the region of interest (RoIs) that may surround an object, the region Proposal Network (RPN) is employed that is originally used in [29]. The unconnected recognition heads of the technique then identify the output label of the detected region and construct a bounding box using the combined key points familiar to each RoI. These techniques are therefore complex in the aspect of computing burden and unreliable for real-time recognition demands. By providing both characteristics and position of RoI in an investigated sample concurrently, the CenterNet more effectively overcomes the shortcomings of RCNN, Fast, and Faster-RCNN. As a result, CenterNet is computationally efficient and more generally applicable to real-time object detection due to its one-stage object detection capability.

The following factors make it difficult to pinpoint the important features of interest for the classification of eye diseases: i) extreme brightness and shade changes, predicting the precise place of the diseased areas from the input image. ii) the class that each object belongs to. By using its heat maps and switching from a 2-phase to a 1-

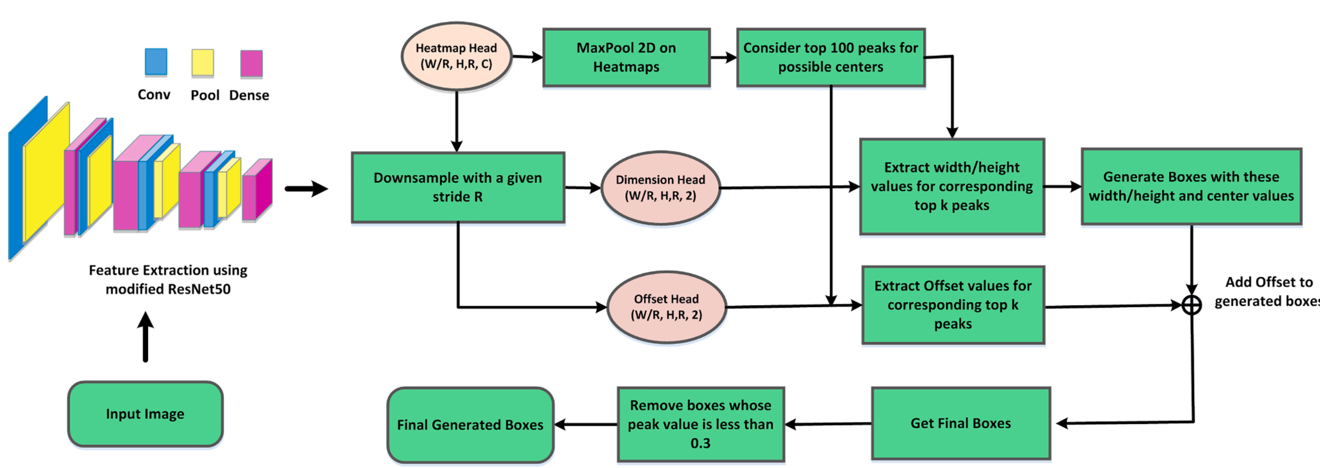

Fig. 1. Flow diagram of Proposed Technique.

phase object locator approach, the CenterNet technique can effectively identify and categorize afflicted regions of various classes. The Heatmap unit uses the center of key points and shows improved recall results, which aids to lower the proposed framework's feature computation cost.

3.3. Custom CenterNet: CoffeeNet

ResNet-101 and VGG16 were used in the traditional CenterNet [30] to compute image keypoints for sample examination. The VGG16 approach suffers from the model overfitting problem, while the ResNet-101 approach makes use of skip-links and identity methods to prevent non-linear transformations that, when utilizing the identity function, result in a straight gradient flow from the back to the front layers and cause huge computing burden. We have presented a customized CenterNet model that uses a lightweight model called the ResNet-50 together with the Convolutional Block Attention mechanism (CBAM) for keypoints selection from the Coffee plant leaves to address the shortcomings of the current framework. Since ResNet-50 contains fewer hyperparameters than backbone architectures, it is computationally more efficient. In order to recognize and classify various Coffee leaf ailments, the CoffeeNet system employs four processes called feature extractor, RPN, ROI pooling, and classification.

3.4. Features extraction

The centerNet approach utilizes a backbone network to calculate meaningful and semantic descriptions from input samples. These features play a crucial role in localizing RoI and performing cataloging in the detection head. The classification performance strongly relies on the quality of the learned keypoints (Nawaz, Javed, et al., 2022). Traditionally, CNN models, initially developed for image classification tasks, have been commonly used as the base model for CenterNet. Previous studies have combined the baseline CenterNet with different base models, i.e. VGG16 and ResNet101 (Masood, et al., 2023). However, research has shown that CNN models tend to prioritize irrelevant features rather than key features when faced with complex backgrounds during training. To address this issue and improve the learning of target representations from the input, attention mechanisms have recently been integrated into CNN networks (Alqahtani, et al., 2023). In the case of coffee disease spot feature representation learning, an attention unit has been incorporated into the base CNN to emphasize important features. This integration enhances the CenterNet model's ability to categorize Coffee leaf infections even in the existence of complex environment settings and practical complications. The attention mechanism employed in this context is based on the CBAM (Nazir, et al., 2022). CBAM adaptively refines features by understanding inter-channel linkages and spatial location properties. It achieves this by consecutively conjecturing attention maps alongside the channel and spatial alignment from the transitional keypoints map. The CBAM block then specifically multiplies the input keypoints map and the attention maps to enhance the process of feature representation learning.

The proposed backbone architecture consists of a CBAM attention block and a ResNet-50 CNN which is a well-known model that uses residual linkages between convolutional layers and identity shortcut connections to achieve high accuracy. While the Dense models, every layer typically passes its computed value to the layer below it, enabling it to create more enlightening keypoints. The gradient vanishing issue, however, makes the convergence outcomes worse as the network depth rises. Multiple residual blocks in the ResNet architecture enable the training process to skip some convolution layers with no degrading accuracy. Convolution layer results are coupled with the processed value of cutoff links to reduce training degradation, which is typically observed in truly deep networks, and get around the restriction of adding more layers. The ResNet model's overall design can be thought of as a collection of numerous CNNs, where several networks are connected by short-cut links and the best outcomes are obtained by averaging the output of all nets. The thorough layer-by-layer setup of the ResNet-50 with CBAM configuration, which is the backbone of the CoffeeNet network, is shown in Table 1. It has 48 convolutional layers broken up into 5 stages, each of which has numerous residual blocks stacked on top of one another. 3×3 convolutional layers, a normalization layer, ReLU activation, and a skip connection make up the residual block. The residual block's stacked layers provide fast connections that carry out identity mapping in order to complete residual mapping.

In order to designate the crucial elements from a given image in advance and utilize them as the input for the following network, we added the attention block to the model. It helps the model focus on virus-infected areas while reducing irrelevant data and improving diagnosis accuracy in situations like varying shade, brightness, and intensity that occur in the real world. To achieve adaptive keypoints improvement, the CBAM unit multiplies the computed keypoints map with the calculated attention map after sequentially inferring the attention map along with the spatial and channel dimensions using the intermediate keypoints map. To cut down on the computing burden, the channel for the substituted convolution layers is set to 64.

3.5. Heatmap head

In CenterNet, the Heatmap head is designated for estimating features heatmaps that show the existence and position of the object in an image. This head is a key component of the CenterNet model and is utilized for object detection and localization of objects.

The heatmap head takes the features acquired from the feature extractor model as input and processes them to construct a set of feature heatmaps. The outcome is feature maps set with the same dimensions but decreased channels. Each channel in feature maps signifies the heatmap for a specific feature. In training time, the head is optimized through the loss function which can be binary cross-entropy loss. This loss is calculated for every pixel in this heatmap and then accumulated from the whole heatmap loss. These heatmaps serve as a crucial component for subsequent steps in the CenterNet framework, such as object localization and dimension estimation.

3.6. Dimension head

In CenterNet, the dimension head is responsible for calculating the size of the objects identified in an image. Particularly, it predicts the width and height of each object based on the locations of features. The head is implemented as a convolutional layer or a series of convolutional layers. These layers take the features extracted from the features heatmaps as input and produce predictions for the object dimensions.

The output of this head is a set of feature maps with spatial

Table 1
ResNet-50 network.

Block/layer	Original	Modified
Conv1	$7 \times 7, 64$ 3×3 max pool	$[3 \times 3, 64] \times 3$ 7×7 Attention
Conv2	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
Conv4	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$
Conv5	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
Conv5	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$

dimensions that relate to the original input image. Each spatial location in these feature maps represents the location of a specific feature, and the dimension head forecasts the breadth and altitude of the object centered at that feature. The dimension head is usually designed to output two channels per spatial location, representing the width and height predictions. These channels are passed through an activation function to produce positive values for the predicted dimensions. During training, the dimension head is improved using a loss function, i.e., the smooth L1 loss, to determine the discrepancy in the projected dimensions and the original dimensions of the objects. The loss is calculated for each spatial location and then aggregated to form the overall dimension loss.

3.7. Offset head

The offset head is responsible for calculating the offsets between the features and the centers of the objects from an image. The offset head plays an essential role in localizing the objects precisely. The head is implemented as a convolutional layer, these layers take the features extracted from the backbone network as input and process them to produce a set of offset estimates. The result is a set of feature maps with the same spatial dimensions as the input image however decreased in channels. During training, the offset head is enhanced through a loss function, such as the smooth L1 loss. The loss is computed for each spatial location and then aggregated to form the overall offset loss.

3.8. Multitask loss

CenterNet is a technique that aims to accurately localize objects and classify them in an end-to-end manner. It accomplishes this by utilizing multi-task loss functions during the training. The multi-job loss method, denoted as L, is applied to each sampled head in the model.

$$L_{centernet} = L_{map} + \lambda_{\dim} L_{\dim} + \lambda_{off} L_{off} \tag{1}$$

The multi-task loss function L combines the losses from these other heads to establish a complete loss that gets both classification and localization information. By optimizing these tasks, the model can learn to accurately localize objects and predict their corresponding classes.

The specific form of the multi-task loss function L can vary depending on the implementation and the specific objectives of the method. Different components such as classification loss, regression loss, and other auxiliary losses may be combined in various ways to form the final loss function.

The L_{map} denotes Heatmap loss and is estimated with Equation (2):

$$\mathrm{L_{map}} = \frac{-1}{\mathrm{n}} \sum_{\mathrm{i,j,c,}} \begin{cases} \left(1-\widehat{\mathrm{o}}_{\mathrm{i,j,c}}\right)^{\alpha} \log\left(\widehat{\mathrm{o}}_{\mathrm{i,j,c}}\right) & if\, \widehat{\mathrm{o}}_{\mathrm{i,j,c}} = 1 \\ & otherwise \\ \left(1-\mathrm{O}_{\mathrm{i,j,c}}\right)^{\beta} \left(\widehat{\mathrm{o}}_{\mathrm{i,j,c}}\right)^{\alpha} & \\ \log\left(1-\widehat{\mathrm{o}}_{\mathrm{i,j,c}}\right) & \end{cases} \tag{2}$$

Here, n = total keypoints $O_{i,j,c}$ = center of real candidate key point, whereas $\widehat{o}_{i,j,c}$ is the projected. Also, α and β are the hyperparameters with the values of 2 and 4, respectively. Equation (3) shows the dimension head loss:

$$L_{\dim} = \frac{1}{n} \sum_{k=1}^{n} \left| \widehat{b}_k - b_k \right| \tag{3}$$

Here, $\widehat{b}_k$ is the projected bbox directs, while b_k is showing the genuine values of bboxes and n specifies the total images. Lastly, Equation (4) shows the Offset-head loss:

$$Loff = \frac{1}{n} \sum_{p} \left| \widehat{F}_{\widehat{p}} - \left(p/_{R} - \widehat{p} \right) \right| \tag{4}$$

Here, $\widehat{F}$ is the estimated offset, p is the real, $\widehat{p}$ is the down-sampled key point and R is the output stride.

3.9. Bounding box estimation

The bounding box (bbox) estimation refers to the process of predicting the coordinates or parameters that define the rectangular bounding box around an object in an image. The bounding box represents the spatial extent of the object and is commonly used for object detection and localization tasks. In the center point representation, the bounding box is typically parameterized by the center coordinates, width, and height. These parameters are predicted by a model based on the given input image.

During the inference stage of CenterNet, the heatmaps are processed independently for each class. To identify the center points of a specific category (c), a criterion is applied to the heatmap responses. If a response score is more than or identical to its eight-linked neighbors, it is considered a potential peak. This helps locate local maxima, which represent potential object center points. Once criteria are applied, the top 100 peak candidates are chosen based on their resultant values.

$$\widehat{Q} = \left\{ \left(\widehat{x}_j, \widehat{y}_j \right) \right\}_{j=1}^{N} \tag{5}$$

We have used the coordinates $(\widehat{x}_j, \widehat{y}_j)$ to denote the positions of each detected point, while $\widehat{Q}$ indicates a set of located center points. These features are associated with a confidence value Ô_(x,y,c), which indicates their recognition point. The bounding box is processed based on the keypoint values, using techniques such as regression. The confidence score of each keypoint provides the overall value of the detection, supporting a more robust and accurate localization. This approach allows us to influence the information provided by keypoints, rather than relying on a single point, increasing the performance of our method. The estimation of the bounding box is performed by using Equation (6).

$$\left(\widehat{x}_j + \widehat{\partial x_j} - \widehat{w_j}/2,\ \widehat{y}_j + \widehat{\partial y_j} - \widehat{h}_j/2,\ \widehat{x}_j + \widehat{\partial x_j} + \widehat{w_j}/2,\ \widehat{y}_j + \widehat{\partial y_j} + \widehat{h}_j/2 \right) \tag{6}$$

In this work, we have used the offset prediction $(\widehat{\partial x_j}, \widehat{\partial y_j})$ denoted by $\widehat{o}_{\hat{x},\hat{y}}$ to represent the predicted offsets. These offsets establish the change required to improve the location of the keypoints. Furthermore, we employed the size prediction $(\widehat{w_j}, \widehat{h_j})$ denoted as $\widehat{d}_{\hat{x},\hat{y}}$. to represent the predicted width and height of the bounding box related to each keypoint. To generate the final bbox, we directly apply the estimated keypoints deprived of using IoU-based non-maxima suppression (NMS). This means that instead of using a traditional method like NMS to eliminate redundant bounding boxes, we directly use the estimated features to determine the final bounding box.

3.10. Detection procedure

A DL-based system called CenterNet is unconstrained by strategies like selective search and proposal development. To determine the center points of the Coffee leaf affected part, offsets to the × and y directs, and the magnitudes of the bboxes together with the related output label, the CenterNet is fed the examined image and the bbox as input to the tuned framework. In our work, we use epochs with a 50 value and a learning rate of 0.001 as the hyper-parameters for model tuning.

4. Results

This section comprises the explanation of the utilized data repository along with the evaluation measures. A large experimental analysis comprising various investigations is also discussed to prove the efficacy of the presented solution for locating and cataloging the various types of coffee plant diseases.

4.1. Evaluation measures

To numerically elaborate the performance results of the presented work, various standard metrics are chosen like intersection over union (IoU), mean Average Precision (mAP), accuracy, precision, F1-score, and recall. The model categorization accuracy is computed by using Equation (7).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

Equation (8) exhibits the description of the mAP score, where the AP specifies the average precision computed for all groups, and t is denoting the processed sample. While T designates the total samples.

$$mAP := \sum_{i=1}^{T} AP(t_i)/T \quad (8)$$

Fig. 2 denotes the graphic form of precision, recall, and IOU, respectively.

4.2. Dataset

For network training and evaluation, we have selected a standard database called the Arabica coffee leaf dataset (Jepkoech, et al., 2021) comprises a total of 18,985 images from five classes of coffee plant leaves named Phoma, Cescospora, Rust, Healthy, and Miner. This dataset is further distributed as JMuBEN and JMuBEN2, where the JMuBEN part comprises samples of three types namely Phoma, Cescospora, and Rust with 8337, 7682, and 8337 images of mentioned categories. While JMuBEN2 comprises samples from healthy, and Minor classes with 18,985 and 16,979 instances. All images of the employed dataset are captured under real-world and complex settings which makes it a challenging dataset in this field. Samples contain several image distortions like color, size, brightness, and orientation variations. An example of the samples from the employed dataset is shown in Fig. 3.

4.3. Model assessment

To prove the efficacy of the CoffeeNet approach, we have executed numerous experimental evaluations. Initially, we tested the presented work to check its localization power and discuss the class-wise categorization ability of the CoffeeNet approach. Here, we also investigated the internal feature computation ability of the approach in the subsequent sections.

4.4. Model detection results

To propose an effective strategy for recognizing the coffee plant leaves abnormalities, it must be proficient in exactly locating the diseased region and the class associated with the detected area. For this reason, an experiment is presented here to show the localization capability of the framework by showing the achieved results on the test samples as provided in Fig. 4. The results in Fig. 4 are clearly explaining that the CoffeeNet framework is competent in exactly locating the diseased portion under the complex sample background setting with different artifacts like color, brightness, size, and position variations. To quantitatively discuss the results, we computed the mAP, and IOU scores for the CoffeeNet framework as these are the standard result computation measures utilized in the field of object detection and classification. Clearly, we attained the mAP, and IOU numbers of 0.970, and 0.973 which is clearly explaining the robustness of our technique.

4.5. Class-wise performance

After discussing the localization ability of our approach, next we performed an experiment to elaborate on the group-wise performance scores of our model. To do this, numerous standard performance measuring parameters like precision, recall, accuracy, confusion matrix, F1-score, and error rates for all categories are discussed in this part of the paper to thoroughly show the classification behavior of the CoffeeNet model. First, precision, and recall are calculated for all five classes of coffee plant diseases, and attained values are given in Fig. 5. The scores are clearly indicating the robustness of our technique for performing the categorization of coffee plant infections. Descriptively, we obtained precision scores of 97.21%, 96.94%, 96.97%, 97.38%, and 97% for Phoma, Cescospora, Rust, Healthy, and Miner classes while in the aspect of recall, the CoffeeNet approach shows numbers of 96.69%, 96.58%, 96.53%, 97.13%, and 96.47% for above-mentioned groups respectively.

Next, the group-wise results of the CoffeeNet model are calculated in the aspect of F1-score and error values for all 5 types of coffee plant diseases. The key cause to computing this evaluation metric is that the complete classification results of an approach cannot be fully captured by the precision, and recall values as some techniques can acquire better precision, with degraded recall values and inversely. So, the F1 metric is computed as it is empowered to take the contribution of both precision and recall and explain the classification results more appropriately. The attained F1-scores, and error values for all five sets of coffee plant leaves are given in Figs. 6 and 7. The scores in Figs. 6 and 7 clearly display that our methodology executes effectively in the aspect of the F1-score and shows small error rates for all five groups of coffee plant leaves disorders. Descriptively, we attained the highest error rate of 3.25% for the Minor class which can be due to the reason that this class contains very small sign of abnormalities and show huge color resemblance with the healthy regions of plants. Moreover, we acquired the lowest error value of 2.75% that is depicting the competency of our model to recognize the healthy class samples. Moreover, the CoffeeNet approach shows 96.95%, 96.76%, 96.75%, 97.25%, and 96.73% for Phoma, Cescospora, Rust, Healthy, and Miner classes, respectively.

Accuracy is another standard metric for reporting the categorization outcomes of a methodology. Here, we computed group-wise classification accuracy for all five types of coffee plant leaf abnormalities, and calculated values are given in Fig. 8. The classification scores are exhibited with the help of box graphs as these plots are effective to explain the results in more detail by elaborating the highest, minimum, and average values for all groups. The numbers in Fig. 8 are indicating that the CoffeeNet model is robust to perform well for all categories of coffee plant infections. Clearly, we obtained accuracy scores of 98.52%, 98.68%, 98.69%, 98.81%, and 98% for Phoma, Cescospora, Rust,

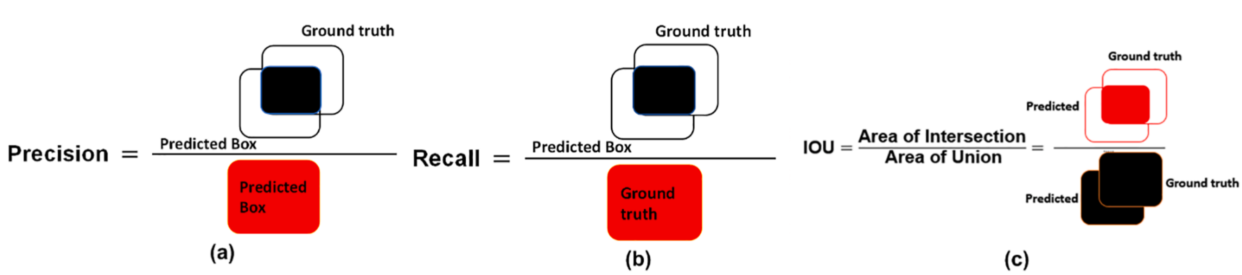

Fig. 2. Graphic depiction of (a) Precision, (b) Recall, and (c) IOU measures.

Fig. 3. Samples from the employed dataset.

Healthy, and Miner coffee plant disease groups. Further, the plot shows that for all groups the classification score remains from a minimum value of 98% to the highest score of 98.81%, thus a maximum variation of 0.81% has been observed among all five classes of coffee plant diseases. This also suggests an average error rate of 1.46% in the aspect of the classification score for all categories of the coffee plant leaf abnormalities.

Further, the confusion matrix attained for the CoffeeNet model is reported as this graph is effective to explain the classification capability of a technique by determining the scores in perspective of true positive rates. The confusion matrix for the presented approach is given in Fig. 9 that is proving the efficiency of our approach to recall all classes of coffee plant leaves due to its high recognition power. Descriptively, the CoffeeNet model has attained an average TPR of 96.68%. Additionally, the scores in Fig. 9 are indicating that the maximum error value is stated among the healthy and minor groups with an average value of 1.50% which is due to the reason that the miner class shows very few signs of leaf abnormalities and has a significant similarity with the healthy regions, however, still, both groups are well recognized by the proposed approach.

So, after performing numerous group-wise assessments of the CoffeeNet approach in this phase of the manuscript, we can conclude that our method is skilled and well-suited to recognize and categorize the numerous kinds of coffee plant leaf disorders. The fundamental factor for this improved classification behavior of the CoffeeNet model is its empowerment to capture the more related and nominative group of visual characteristics of the investigated samples. Further, the inclusion of spatial and channel attention strategy in the feature extractor assists the CoffeeNet approach to highlight the diseased-specific areas of the input images which as a result boosts the recognition power of the proposed approach.

4.6. Visualization and overall performance

In this phase of the paper, we have provided an evaluation of our approach by investigating its internal behavior and reporting the classification results entirely. To analyze the internal working of the model, we computed the heatmaps by using the Grad-Cam in the final convolution layer of the CoffeeNet model which assists to determine whether an approach is taking the accurate disease region of a sample to perform the recognition task. The attained visual results for samples from the employed dataset are reported in Fig. 10 in which the reddish color is signifying the area considered by the CoffeeNet approach to accomplish the classification task. So, the visual results given in Fig. 10 reveal that the CoffeeNet model is taking the right areas of diseased portion to execute the distribution of coffee plant leaves into relevant groups. The key factor which assists the CoffeeNet approach to attain improved explainability behavior is its high recognition power which permits it to separate the numerous groups of coffee plant leaves effectively and distribute the images into respective types.

Further, we also explain the overall classification performance of the CoffeeNet model by reporting its results on the test images of the employed data sample. To do so, we have taken several standard measures like AUC, accuracy, recall, F1-score, Precision, and an error value on the test sample, and obtained classification performance values are elaborated in Fig. 11. The results in Fig. 11 are proving the better classification behavior of the CoffeeNet approach for all performance measures. Descriptively, the CoffeeNet approach has reported an accuracy of 98.54%, along with the F1-score and error values of 96.89%, and 3.11%. Further, the CoffeeNet model has attained an AUC score of 98.40%, while for precision metric, a number of 97.10% is attained which is 96.68% in the aspect of recall measure. All these visual and numeric results are clearly indicating the high classification results of our approach which are clearly showing the efficacy of the CoffeeNet model in diagnosing and cataloging all five types of coffee plant leaf diseases.

4.7. Comparison with DL models

In this section, we have provided an investigation to compare the recognition performance of the presented model with various DL frameworks. To perform this, we used numerous latest DL approaches like VGG-16 (Nawaz, et al., 2021), Inception-V3 (Xia, Xu, & Nan, 2017), ResNet-152 (Nawaz, Javed, et al., 2022), Xception (Chollet, 2017), MobileNet-V2 (Chen, Zhang, Suzauddola, & Zeb, 2021), DenseNet (Nawaz, Nazir, et al., 2022), InceptionResNet-V2 (Ferreira, et al., 2018), and NASNetMobile (Saxen, et al., 2019). We have provided two types of comparison in the subsequent sections where initially, we have

Fig. 4. CoffeeNet model localization results.

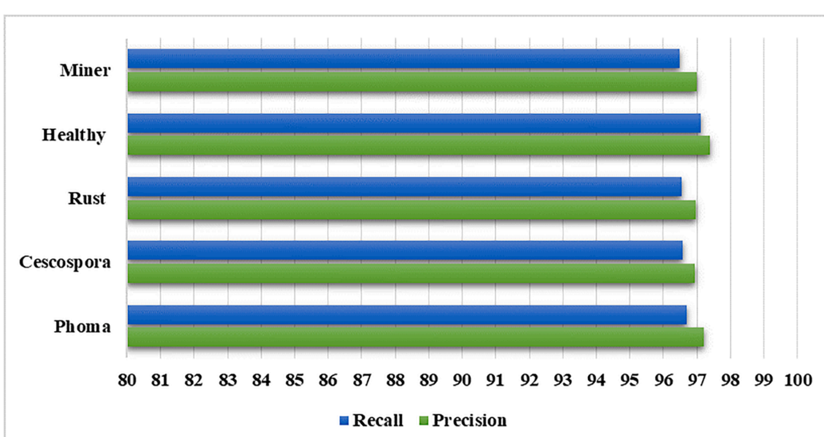

Fig. 5. Group-wise precision, and recall results for the CoffeeNet model.

compared the class-wise result of our approach with the DL approaches, while in the next phase, the results are discussed in terms of the entire dataset.

4.8. Class-wise performance comparison with DL models

Here, the performance of our approach is evaluated with the nominated DL approaches in the aspect of the AUC metric as it is the standard

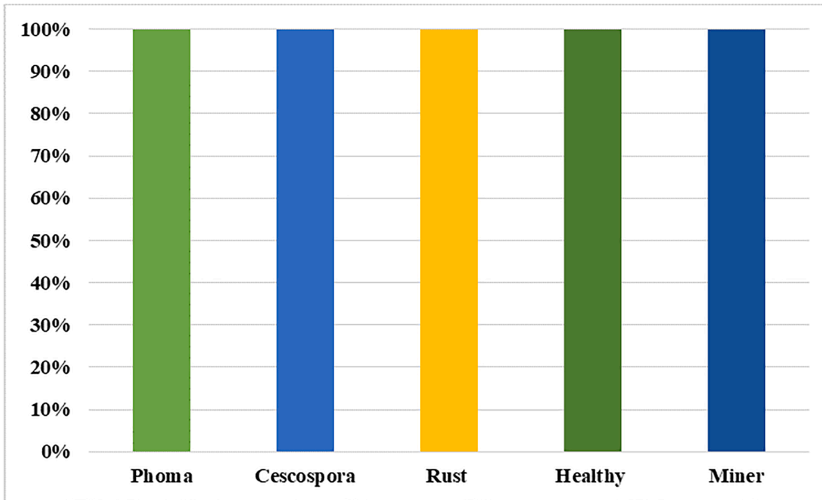

Fig. 6. Group-wise F1 for the CoffeeNet approach.

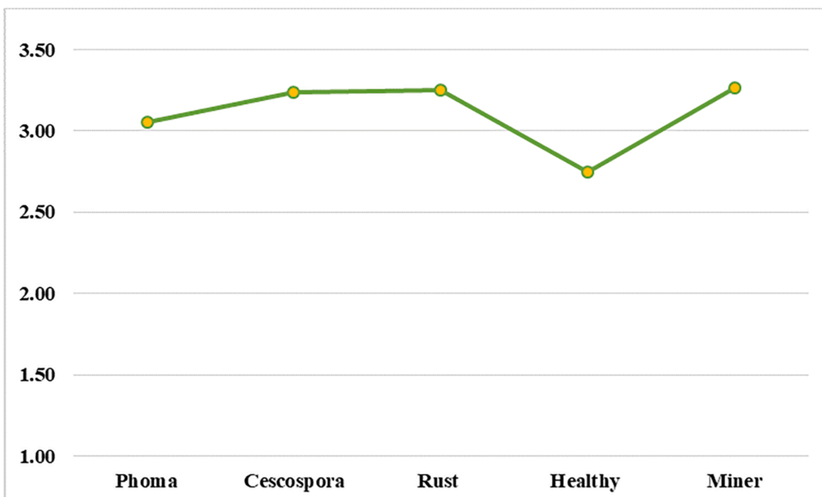

Fig. 7. Group-wise Error-rates for the CoffeeNet approach.

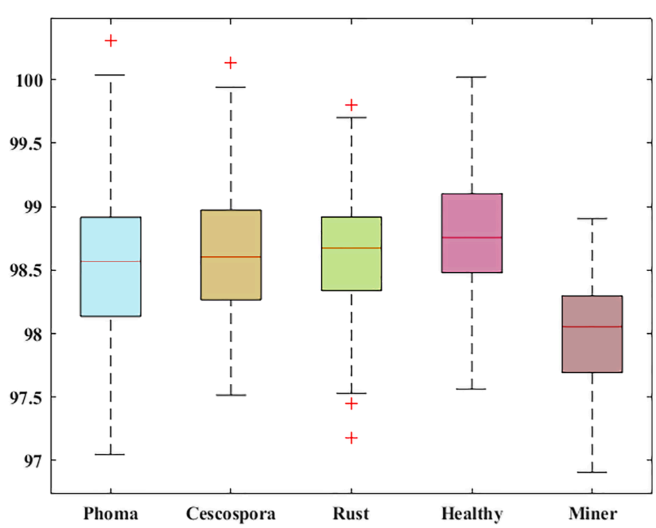

Fig. 8. Group-wise accuracy scores attained with the CoffeeNet model.

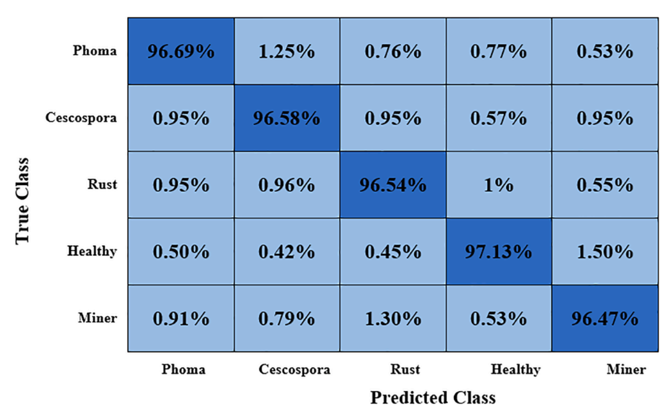

Fig. 9. Confusion matrix for the CoffeeNet model.

measure in the area of image classification. The AUC scores of VGG-16, Inception-V3, ResNet-152, Xception, MobileNet-V2, DenseNet, InceptionResNet-V2, and NASNetMobile are reported and compared for all five classes of coffee plant leaf diseases against the CoffeeNet approach as mentioned in (Novtahaning, et al., 2022). The obtained

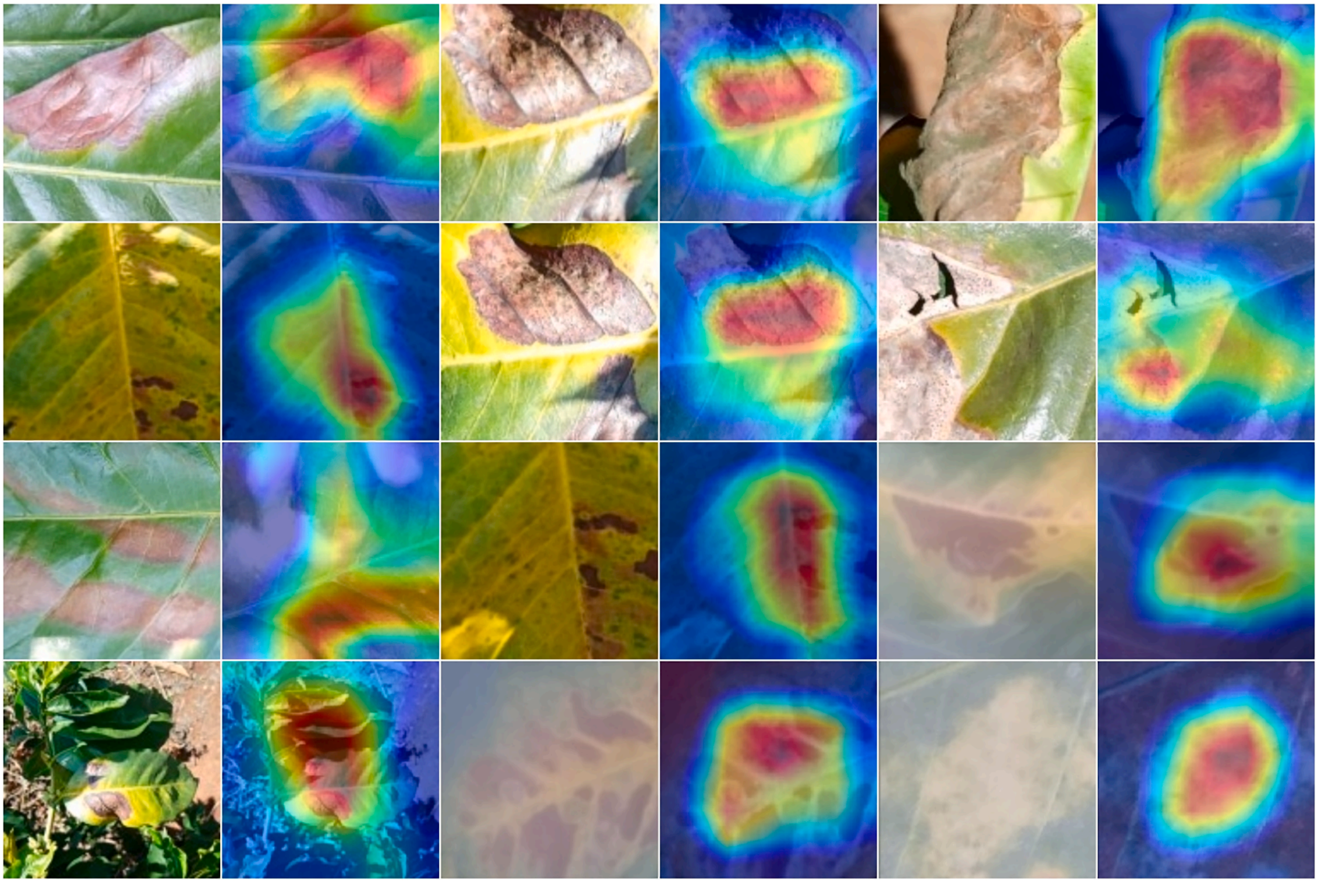

Fig. 10. Heatmaps attained with the CoffeeNet model.

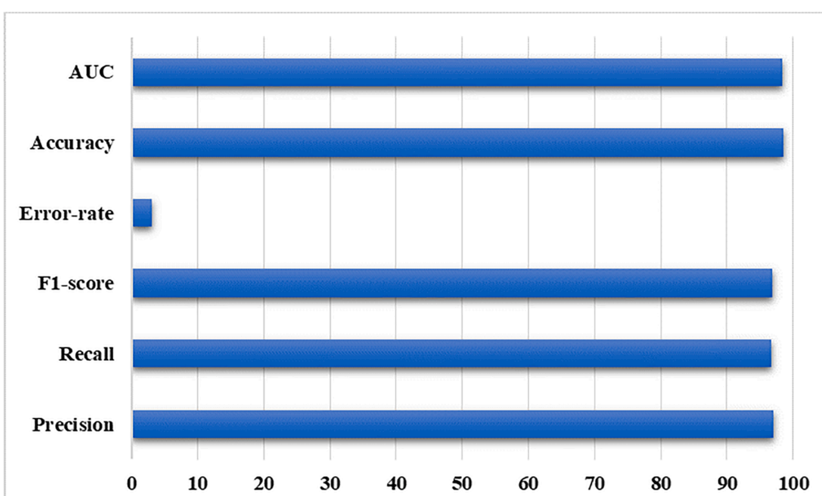

Fig. 11. Overall performance results of the CoffeeNet approach over the employed dataset.

comparison is given in Table 2 which clearly signifies that our work has attained the highest AUC scores for all five groups of coffee leaf disorders. Clearly, for the Phoma group, the comparison DL approaches have shown an average AUC number of 0.946 which is 0.99 for the proposed approach. Consequently, for the Phoma group, a performance gain of 4.37% is stated. Then, for the Cescospora group, the relative techniques attained an average AUC score of 0.92, while the CoffeeNet approach attained an average AUC score of 0.98 and exhibited a performance gain of 6%. Further, in terms of Rust, Healthy classes of coffee plant leaf disorders, the comparison DL techniques show average AUC numbers of 0.898 and 0.896 which is 0.99, and 1 for the proposed work, and presented performance gains of 9.25% and 10.37% for the mentioned groups respectively. Finally, for the Miner class, the selected DL approaches have attained an AUC number of 0.856 which is 0.97 for the proposed approach. Hence, for the Miner class, the CoffeeNet approach has exhibited a performance gain of 11.37% clearly signifying the efficacy of our model. The given performance comparison clearly validates that the proposed CoffeeNet approach effectively computes the coffee

Table 2
Group-wise comparison of the CoffeeNet model with DL frameworks in terms of AUC.

Model	Phoma	Cescospora	Rust	Healthy	Miner
VGG-16	0.98	0.97	0.95	0.97	0.95
Inception-V3	0.93	0.86	0.88	0.74	0.83
ResNet-152	0.97	0.98	0.97	0.99	0.90
Xception	0.91	0.95	0.92	0.90	0.86
MobileNet-V2	0.95	0.94	0.67	0.82	0.81
DenseNet	0.95	0.90	0.91	0.89	0.83
InceptionResNet-V2	0.92	0.85	0.98	0.97	0.84
NASNetMobile	0.96	0.91	0.90	0.89	0.83
Proposed	0.99	0.98	0.99	1	0.97

plant leaf abnormalities-specific characteristics in the model training phase and consequently correctly recalls them from examined images. Further, regardless of the similarity in the healthy and infected sections of leaves and the complex surrounding environmental conditions, the suggested methodology is competent in differentiating all groups of leaf infections with a high recall rate. Comparatively, the other DL approaches are not much competent to robustly extract the fine-level attributes of the examined samples under such diverse environmental constraints and eventually result in performance degradation. This comparison proves that the introduction of the ResNet-50 feature extractor with added attention strategy boosts the finer-grained disease-specific feature extraction power of the CoffeeNet approach by allowing the propagation of highly significant sample information at pixel and channel levels.

4.9. Overall performance comparison with DL models

Next, the classification results of the CoffeeNet model are equated with the DL frameworks in terms of the entire data sample. This experiment is conducted by comparing the results of our approach with the peer models via taking various performance measures like accuracy, F1, error rate, recall, and precision, and values are given in Table 3. The results in Table 3 are clearly explaining the robustness of our approach in comparison to all other DL models as we outperformed the competitor models in the aspect of all performance measures and achieved the highest classification values. More explicitly, the minimum classification performance is reported by the MobileNet-V2 approach with accuracy, and error rates of 74.60%, and 26.50%. While, the second lowest results are acquired by the DenseNet, and NASNetMobile approaches with accuracy and error values of 83.80% and 16.70%. The VGG-16 approach shows comparable results with accuracy and error values of 94.20%, and 5.90%, however, with an increased computing burden. Comparatively, the CoffeeNet technique exhibits the maximum accuracy score of 98.54%. Descriptively, for precision metric, the relative techniques exhibit an average score of 86.34% which is 97.10% for the proposed model and give a performance gain of 10.76%. Then, in the aspect of the recall measure, the competitor models attain an average number of 85.50%, while the CoffeeNet approach attains a value of 96.68% and exhibits a performance gain of 11.18%. Next, for the F1-score measure, the peer works have reported an average number of 85.29%, which is 96.89%, and attains a performance gain of 11.60. While, for accuracy measure, the selected DL works have shown an average score of 85.80%, which is 98.54% for our model and we have attained a performance gain of 12.74%. Moreover, the peer works have reported an average error rate of 14.71%, which is 3.11% for our approach and clearly indicates the efficacy of our model for recognizing all types of coffee plant infections.

Table 3
Comparison of the CoffeeNet model with DL approaches on the entire data sample.

Model	Precision	Recall	F1-score	Error-rate	Accuracy
VGG-16	94.40%	94%	94.10%	5.90%	94.20%
Inception-V3	83.50%	85.10%	83.50%	16.50%	83.90%
ResNet-152	94%	93.20%	93.30%	6.70%	93.80%
Xception	85.50%	85.30%	85.20%	14.80%	85.40%
MobileNet-V2	76.80%	74.10%	73.50%	26.50%	74.60 t%
DenseNet	84.70%	83.20%	83.30%	16.70%	83.80%
InceptionResNet-V2	86.70%	86%	86.10%	13.90%	86.90%
NASNetMobile	85.10%	83.10%	83.30%	16.70%	83.80%
Proposed	97.10%	96.68%	96.89%	3.11%	98.54%

The key characteristic of our approach is its improved recall capability which enhances the recognition results of the CoffeeNet approach. The inclusion of the attention mechanism-based ResNet-50 approach in the proposed work improves the reliable visual characteristics computation capability of our approach which boosts its learning behavior in the existence of various sample distortions like noise, clutter, color, size, and orientation variations, and assists the approach to acquire infection-specific signs in the complicated background settings. Whereas, the comparative techniques lack such behavior, and hence the comparison provided both in the aspect of classes and on the whole data sample proves that the CoffeeNet approach is more reliable in both locating and classifying the investigated samples.

4.10. Comparison with base models

In this phase, an experiment is performed to analyze the localization and classification performance of the proposed CoffeeNet approach in contrast to various other object recognition approaches. The test samples from the employed dataset are used to check the performance of all approaches in the presence of various sample artifacts like light, color, size, and position variations of infected image areas.

For this reason, two categories of object identification approaches are taken which are known as 1-step, and 2-step methods. The main distinction in both types of approaches is that the 2-phase object detection approaches perform by initially determining the location of the area of interest in the examined samples and then the categorization job is accomplished. In comparison, the one-phase object recognition approaches perform both tasks in a single step. For a fair comparison, we have taken both types of approaches that have exhibited effective results on the COCO data sample (Lin, et al., 2014) and evaluated the performance of the Arabica coffee leaf dataset. For the 2-phase approaches, we have taken the Fast-RCNN (Girshick, 2015) and Faster-RCNN (Ren, He, Girshick, & Sun, 2016) model while for the 1-phase approaches, we have taken the You Only Look Once (YOLO)(Redmon & Farhadi, 2018) and single shot detector (SSD) (Liu, et al., 2016), and RetinaNet (Lin, Goyal, Girshick, He, & Dollár, 2017) models. All these frameworks are trained with the following parameters: a batch size of 32, a learning rate with a value of 0.001, and a momentum score of 0.9 respectively. The attained performance analysis is given in Table 4.

The mAP, and IOU scores for all techniques are estimated as these measures have been heavily investigated in the area of object recognition and marked as standard measures. Further, the test time of all

Table 4
Performance analysis of the CoffeeNet model with other object recognition approaches.

Model	Feature extractor	mAP	IOU	Test time (sec/img)
Two-Stage Models				
Fast-RCNN	VGG-16	0.80	0.830	0.30
Faster-RCNN	VGG-16	0.85	0.853	0.28
Faster-RCNN	ResNet-101	0.93	0.938	0.26
One-Stage Models				
YOLO-V3	DarkNet-53	0.82	0.830	0.26
SSD	ResNet-101	0.81	0.838	0.29
RetinaNet	ResNet-101	0.90	0.901	0.25
Proposed	DenseNet-77	0.97	0.973	0.23

approaches is also investigated to examine them from the perspective of processing complexity. The scores in Table 4 validated the effectiveness of our approach both in the aspect of running time and recognition performance. Descriptively, the VGG16-based Fast-RCNN approach has shown the lowest classification results with mAP, and IOU scores of 0.80, and 0.830. This model is also more computationally expensive and takes 0.30 s to process a sample. The second minimum identification values are reported by the YOLO-V3 model with scores of 0.82, and 0.830. While in terms of processing time, the SSD approach displays the second lowermost performance with a time of 29 s in processing a given sample. Comparatively, the proposed CoffeeNet approach performs better in the aspect of both coffee plant leaf infections recognition and processing time with mAP, IOU values of 0.970, and 0.973 with a sample processing time of 0.23 s. The main reason for this better performance of the CoffeeNet approach is due to its high recognition ability, comparatively, the Fast-RCNN approach employed the hand-coded features and is unable to tackle the image distortions. Moreover, the ResNet101-based Faster-RCNN approach shows better results with mAP, and IOU numbers of 0.93, and 0.938, however, with an increased processing time of 0.26 s per sample because of its 2-phase object recognition framework. Further, the SSD and YOLO approaches are not proficient to recognize the infected regions of small sizes and resulting in performance degradation. Moreover, although the RetinaNet can handle small regions, however, the approach is incompetent to draw reliable anchors over the acentric keypoints. The proposed CoffeeNet model better handle such issues of all approaches by presenting a better feature computation network by incorporating the spatial and attention mechanism which empowers the approach to nominate the more disease-specific keypoints which are later effectively localized and categorized by the presented network. Moreover, the 1-phase detector of the CoffeeNet approach also provides a processing advantage as well. So, our approach is more effective and proficient than all other comparative object recognition approaches.

4.11. Performance analysis with state-of-the-art

We accomplished an investigation to analyze the classification performance of the CoffeeNet model with various state-of-the-art approaches to compare our results with them. To perform this, we executed two types of evaluation where initially, we investigated class-wise results and then compared the classification performance on the entire dataset.

The class-wise result comparison with the latest approaches defined in (Atila, Uçar, Akyol, & Uçar, 2021; Novtahaning, et al., 2022) in terms of AUC metric is given in Table 5, whereas the comparative analysis over the entire data sample is provided in Table 6. The comparative investigation in Table 5 and Table 6 is showing that our proposed work is more proficient in recalling all groups of coffee plant infections as compared to new works. The (Atila, et al., 2021) presents a DL model called EfficientNet-B0 which is tuned to classify the abnormalities from the coffee plants with an accuracy value of 95%, whereas the method in (Novtahaning, et al., 2022) presented an ensembling approach by joining three different CNN models for dense features computation which are later classified to five groups of coffee plant leaf abnormalities and gained an accuracy value of 97.30%. The models in (Atila, et al., 2021; Novtahaning, et al., 2022) show better classification results, however, our model outperforms both techniques in terms of both group-wise and entire sample evaluations. The comparative approaches deploy more dense networks for keypoints computation and are unable to learn a more disease-specific feature vector of examined samples. Comparatively, our approach performs reliably due to the more effective visual characteristics computation ability which enables it to prohibit the model over-fitting issue and better learn the sample information in the feature engineering phase. The inclusion of an attention mechanism allows the CoffeeNet approach to reliably extract the fine-level infection information both along the pixel and channel orientation of samples, which eventually permits the framework to correctly identify the area of focus in the given image and enhances the overall recognition power of the model in the incidence of complex sample background settings. So, we can conclude that our model is more reliable to identify and classify all five types of coffee plant leaf diseases.

Table 5
Comparison with new works in terms of AUC.

Model	Phoma	Cescospora	Rust	Healthy	Miner
EfficientNet-B0 (Atila, et al., 2021)	0.96	0.98	0.98	0.99	0.93
Ensemble Model (Novtahaning, et al., 2022)	0.97	0.98	0.98	1	0.92
Proposed	0.99	0.98	0.99	1	0.97

Table 6
The CoffeeNet model comparison with the latest works on the entire data sample.

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
EfficientNet-B0 (Atila, et al., 2021)	95.20	94.80	94.90	95
Ensemble Model (Novtahaning, et al., 2022)	95.70	95.20	95.10	97.30
Proposed	97.10	96.68	96.89	98.54

5. Conclusion

This research work has presented an effective and improved DL model named the CoffeeNet framework to locate and categorize different types of coffee plant leaf abnormalities. Descriptively, a customized CenterNet framework is introduced by proposing the spatial-channel attention strategy-based ResNet50 model as its backbone network to extract a denser and disease-specific set of sample keypoints vector. The extracted characteristics are then localized and categorized by the 1-phase locator of the CenterNet approach. A huge experimental analysis accompanying various evaluations has been accomplished on a complex data sample called the Arabica coffee leaf to show the efficacy of the suggested approach. We have proved through experiments that the proposed CoffeeNet approach is more effective than several latest approaches and capable of better handling various types of sample distortions. We acquired an accuracy value of 98.54%, along with an mAP number of 0.97 that is clearly signifying the reliability of our approach to diagnosing the various types of coffee plant leaf contaminations. As a future concern, we are willing to investigate other DL nets to further boost the categorization performance.

CRediT authorship contribution statement

Marriam Nawaz: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Tahira Nazir:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Ali Javed:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Writing – review & editing. **Sherif Tawfik Amin:** Methodology, Investigation, Writing – original draft, Writing – review & editing. **Fathe Jeribi:** Formal analysis, Data curation, Writing – review & editing, Visualization. **Ali Tahir:** Resources, Writing – review & editing, Visualization.

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.

Data availability

We have used publicly available dataset and link is provided in the paper.

Acknowledgements

The authors would like to thank the research facilities and support provided by the Multimedia Signal Processing (MSP) Research Lab at University of Engineering and Technology-Taxila, Pakistan for conducting this research.

References

Albahli, S., & Nawaz, M. (2022). DCNet: DenseNet-77-based CornerNet model for the tomato plant leaf disease detection and classification. *Frontiers Plant Science, 13*.
Albattah, W., Nawaz, M., Javed, A., Masood, M., & Albahli, S. (2022). A novel deep learning method for detection and classification of plant diseases. *Complex Intelligent Systems*, 1–18.
Alqahtani, Y., Nawaz, M., Nazir, T., Javed, A., Jeribi, F., & Tahir, A. (2023). An improved deep learning approach for localization and Recognition of plant leaf diseases. *Expert Systems with Applications, 120717*.
Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics, 61*, Article 101182.
Aufar, Y., Abdillah, M. H., & Romadoni, J. (2023). Web-based CNN Application for Arabica Coffee Leaf Disease Prediction in Smart Agriculture. *Jurnal RESTI, 7*, 71–79.
Ayikpa, K. J., Mamadou, D., Gouton, P., & Adou, K. J. (2022). Experimental Evaluation of Coffee Leaf Disease Classification and Recognition Based on Machine Learning and Deep Learning Algorithms. *Journal of Computer Science, 18*, 1201–1212.
Chang, S.-J., & Huang, C.-Y. (2021). Deep learning model for the inspection of coffee bean defects. *Applied Sciences, 11*, 8226.
Chen, J., Zhang, D., Suzauddola, M., & Zeb, A. (2021). Identifying crop diseases using attention embedded MobileNet-V2 model. *Applied Soft Computing, 113*, Article 107901.
Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).
Chowdhury, M., & Burhan, U. (2021). Coffee leaf disease recognition using gist feature. *International Journal of Information Engineering Electronic Business, 13*, 55–61.
Dhanaraju, M., Chenniappan, P., Ramalingam, K., Pazhanivelan, S., & Kaliaperumal, R. (2022). Smart Farming: Internet of Things (IoT)-Based Sustainable Agriculture. *Agriculture, 12*, 1745.
Dogan, M., Taspinar, Y. S., Cinar, I., Kursun, R., Ozkan, I. A., & Koklu, M. (2023). Dry bean cultivars classification using deep cnn features and salp swarm algorithm based extreme learning machine. *Computers Electronics in Agriculture, 204*, Article 107575.
Esgario, J. G., Krohling, R. A., & Ventura, J. A. (2020). Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers Electronics in Agriculture, 169*, Article 105162.
Ferreira, C. A., Melo, T., Sousa, P., Meyer, M. I., Shakibapour, E., Costa, P., & Campilho, A. (2018). Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15* (pp. 763-770): Springer.
Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
Gutte, V. S., & Gitte, M. A. (2016). A survey on recognition of plant disease with help of algorithm. *International Journal of Engineering Science, 7100*.
Hasan, R. I., Yusuf, S. M., Rahim, M. S. M., & Alzubaidi, L. (2022). Automated masks generation for coffee and apple leaf infected with single or multiple diseases-based color analysis approaches. *Informatics in Medicine Unlocked, 28*, Article 100837.
Hitimana, E., & Gwun, O. (2014). Automatic estimation of live coffee leaf infection based on image processing techniques. *arXiv preprint arXiv:*.
Hoosain, M. S., Paul, B. S., & Ramakrishna, S. (2020). The impact of 4IR digital technologies and circular thinking on the United Nations sustainable development goals. *Sustainability, 12*, 10143.
Hsia, C.-H., Lee, Y.-H., & Lai, C.-F. (2022). An Explainable and Lightweight Deep Convolutional Neural Network for Quality Detection of Green Coffee Beans. *Applied Sciences, 12*, 10966.
ICO, Monthly coffee market report. In.
Javierto, D. P. P., Martin, J. D. Z., & Villaverde, J. F. (2021). Robusta Coffee Leaf Detection based on YOLOv3-MobileNetv2 model. In *In 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)* (pp. 1–6). IEEE.
Jepkoech, J., Mugo, D. M., Kenduiywo, B. K., & Too, E. C. (2021). Arabica coffee leaf images dataset for coffee leaf disease detection and classification. *Data in brief, 36*, Article 107142.
Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740–755). Springer.
Lisboa, E., Lima, G., & Queiroz, F. (2021). Coffee Leaf Diseases Identification and Severity Classification using Deep Learning. In *Anais Estendidos do XXXIV Conference on Graphics, Patterns and Images* (pp. 201–205). SBC.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37). Springer.
Marcos, A. P., Rodovalho, N. L. S., & Backes, A. R. (2019a). Coffee leaf rust detection using convolutional neural network. In *In 2019 XV Workshop de Visão Computacional (WVC)* (pp. 38–42). IEEE.
Marcos, A. P., Rodovalho, N. L. S., & Backes, A. R. (2019b). Coffee leaf rust detection using genetic algorithm. In *In 2019 XV Workshop de Visao Computacional (WVC)* (pp. 16–20). IEEE.
Marin, D. B., Santana, L. S., Barbosa, B. D. S., Barata, R. A. P., Osco, L. P., Ramos, A. P. M., & Guimarães, P. H. S. (2021). Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. *Computers Electronics in Agriculture, 190*, Article 106476.
Masood, M., Nawaz, M., Nazir, T., Javed, A., Alkanhel, R., Elmannai, H., … Bourouis, S. (2023). *MaizeNet: A deep learning approach for effective recognition of maize plant leaf diseases*. IEEE Access.
Nawaz, M., Javed, A., & Irtaza, A. (2022). ResNet-Swish-Dense54: A deep learning approach for deepfakes detection. *The Visual Computer*, 1–22.
Nawaz, M., Masood, M., Javed, A., Iqbal, J., Nazir, T., Mehmood, A., & Ashraf, R. (2021). Melanoma localization and classification through faster region-based convolutional neural network and SVM. *Multimedia Tools Applications, 80*, 28953–28974.
Nawaz, M., Nazir, T., Baili, J., Khan, M. A., Kim, Y. J., & Cha, J.-H. (2023). CXray-EffDet: Chest Disease Detection and Classification from X-ray Images Using the EfficientDet Model. *Diagnostics, 13*, 248.
Nawaz, M., Nazir, T., Masood, M., Ali, F., Khan, M. A., Tariq, U., … Damaševičius, R. (2022). Melanoma segmentation: A framework of improved DenseNet77 and UNET convolutional neural network. *International Journal of Imaging Systems Technology, 32*, 2137–2153.
Nazir, T., Nawaz, M., Javed, A., Malik, K. M., Saudagar, A. K. J., Khan, M. B., … AlKhathami, M. (2022). COVID-DAI: A novel framework for COVID-19 detection and infection growth estimation using computed tomography images. *Microscopy Research Technique, 85*, 2313–2330.
Novtahaning, D., Shah, H. A., & Kang, J.-M. (2022). Deep Learning Ensemble-Based Automated and High-Performing Recognition of Coffee Leaf Disease. *Agriculture, 12*, 1909.
Paul, A., Ghosh, S., Das, A. K., Goswami, S., Choudhury, S. D., & Sen, S. (2020). A review on agricultural advancement based on computer vision and machine learning. In *Emerging technology in modelling and graphics* (pp. 567–581). Springer.
Paulos, E. B., & Woldeyohannis, M. M. (2022). Detection and Classification of Coffee Leaf Disease using Deep Learning. In *In 2022 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)* (pp. 1–6). IEEE.
Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:.02767*.
Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis Machine Intelligence, 39*, 1137–1149.
Ruttanadech, N., Phetpan, K., Srisang, N., Srisang, S., Chungcharoen, T., Limmun, W., … Kongtragoul, P. (2023). Rapid and accurate classification of Aspergillus ochraceous contamination in Robusta green coffee bean through near-infrared spectral analysis using machine learning. *Food Control, 145*, Article 109446.
Saleem, M. F., Shah, S. M. A., Nazir, T., Mehmood, A., Nawaz, M., Khan, M. A., … Thinnukool, O. (2022). *Signet Ring Cell Detection from Histological Images Using Deep Learning*. Computers, Materials & Continua.
Saxen, F., Werner, P., Handrich, S., Othman, E., Dinges, L., & Al-Hamadi, A. (2019). Face attribute detection with mobilenetv2 and nasnet-mobile. In *In 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)* (pp. 176–180). IEEE.
Ventura, J., ESGARIO, J., & KROHLING, R. (2020). Deep learning for classification and severity estimation of coffee leaf biotic stress.
Xia, X., Xu, C., & Nan, B. (2017). Inception-v3 for flower classification. In *In 2017 2nd international conference on image, vision and computing (ICIVC)* (pp. 783–787). IEEE.
Yamashita, J. V. Y. B., & Leite, J. P. R. (2023). Coffee disease classification at the edge using deep learning. *Smart Agricultural Technology, 4*, Article 100183.