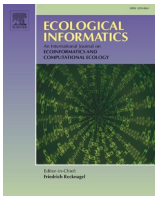
Ecological Informatics 77 (2023) 102213

Contents lists available at ScienceDirect

Ecological Informatics

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



Inception-based global context attention network for the classification of coffee leaf diseases

R. Karthik [a,*], J. Joshua Alfred [b], J. Joel Kennedy [b]

[a] Centre for Cyber Physical Systems, School of Electronics Engineering, Vellore Institute of Technology, Chennai, India
[b] School of Electronics Engineering, Vellore Institute of Technology, Chennai, India

ARTICLE INFO

Keywords:
Coffee leaf disease
Deep learning
Inception
Global context block
Multi-head attention

ABSTRACT

Coffee is a significant global commodity that is consumed in large quantities on a daily basis, and it ranks as the second-most important product in global trade. The leaves of coffee plants are vulnerable to fungal and pest attacks that can harm their photosynthetic regions, leading to the production of low-quality, diseased beans. Accurate diagnosis of coffee disease is crucial in determining suitable remedial action. Different pathogens require different pesticides or other forms of treatment, and a precise diagnosis can guarantee that the necessary steps are taken to prevent further damage to the plant. The major objective of this research is to develop a disease detection approach for coffee leaves based on deep learning techniques. The proposed network is developed with inception modules, a global context module, and a multi-head attention module to achieve precise classification. Through the simultaneous application of different filter sizes and pooling operations on the input, inception modules facilitate the proposed network in extracting features at multiple scales and acquiring significant feature maps at different levels of abstraction. The Global Context Block employs a channel attention mechanism to generate a single feature vector that modulates the input feature maps to obtain high-level contextual information. Finally, the multi-head attention module captures complex relationships between features from different subsets and aggregates them to form a more powerful representation of the input. The experimental results indicated that the proposed network, which was trained on the BRACOL dataset, outperformed existing networks in detecting coffee leaf disease. It achieved an accuracy of 98.57% and a F1 score of 98.55%.

1. Introduction

Coffee production is a widespread practice in 80 tropical countries, mainly in the regions of Africa, Asia, and Latin America, yielding an annual output of 167.2 million bags of coffee beans (Krishnan 2017). As per an annual review by the International Coffee Organization, global coffee production consists of 93.97 million bags of Arabica beans and 73.2 million bags of Robusta beans (Annual Review n.d.). Brazil is acknowledged as the foremost coffee producer and exporter on a global scale, contributing to roughly 29.6% of the world's coffee exports.

Arabica and Robusta represent the primary species among the 100 coffee varieties, and they serve as the major drivers of the global coffee industry and consumption. Arabica coffee is grown in approximately 80% of countries worldwide and is more prevalent on the American continent than in Asian countries, mainly due to favorable temperature and humidity conditions. Arabica coffee is widespread in the south western highlands of Ethiopia, a country whose economy relies heavily on coffee production (Hailu et al. 2017). Moreover, the south eastern regions of Brazil have been identified as particularly suitable for Arabica coffee cultivation and bean productivity. Unlike Arabica coffee, Robusta coffee has demonstrated a lower tolerance toward colder temperatures. Consequently, the lowlands of Espirito Santo and Rondonia States have emerged as major centers for the large-scale cultivation of Robusta coffee (Volsi et al. 2019). However, the coffee crops of every country, including Brazil, are susceptible to pests and diseases.

Widespread diseases, such as Cercospora leaf spots, leaf miners, and fungal phoma, are known to adversely affect the surface of coffee leaves, causing a reduction in the overall photosynthetic area and consequently leading to a decline in coffee production. Among these, coffee leaf rust is recognized as the primary disease responsible for coffee crop damage. This fungal pathogen spreads through spores present on the surface of coffee leaves and enters their stomatal openings, leading to a reduction in leaf area and energy accumulation (Silva et al. 2022). The efficiency of coffee production can be significantly diminished by leaf rust, which

* Corresponding author.
E-mail addresses: r.karthik@vit.ac.in (R. Karthik), joshuaalfred.j2019@vitstudent.ac.in (J. Joshua Alfred), joelkennedy.j2019@vitstudent.ac.in (J. Joel Kennedy).

https://doi.org/10.1016/j.ecoinf.2023.102213
Received 1 May 2023; Received in revised form 10 July 2023; Accepted 12 July 2023
Available online 18 July 2023

can be influenced by various factors, including crop management practices and climatic conditions (Laércio Zambolim and Eveline Teixeira Caixeta n.d.). Cercospora, or brown leaf spots, is another prevalent disease caused by *Cercospora coffeicola*. It manifests as circular brown necrotic spots surrounded by a yellowish halo, with the fungus entering the adaxial surface of coffee leaves and causing significant harm to adjacent tissues. This ultimately results in a reduction in both plant productivity and bean quality (Andrade et al. 2021). Coffee leaf miner disease is caused by moth infestation, which affects the mesophyll of coffee leaves and inhibits photosynthesis, resulting in reduced yields (Dantas et al. 2021). Another coffee leaf disease, Phoma leaf spots, germinates and infects the photosynthetic area of coffee leaves under optimal temperature and humidity conditions (Ventura et al. 2019). Furthermore, coffee production is significantly impacted by several other pests, including coffee stem borers and coffee berry borers, as well as diseases such as berry disease, bacterial blight, and coffee wilt.

Accurate detection and identification of a plant disease or pest is a critical step prior to treatment. The use of pesticides without proper detection can be ineffective or may even compromise the plant's resistance to the pathogen. However, manually identifying the disease by visual inspection of infected leaves is a difficult and time-consuming task. Automated disease diagnosis systems offer a solution to this problem by providing a low-cost, efficient, and accurate method for detecting the type of disease. A computer-aided application utilizing machine learning and deep learning techniques has the potential to benefit farmers by providing quick and accurate identification of diseases along with relevant information. This research proposes an automatic disease detection system for coffee leaves using a custom deep learning network. The proposed network can be practically implemented on drones, facilitating the remote analysis of coffee plantations. The deployed network can effectively process and categorize the image data obtained by drones, thereby promoting an effective remote diagnosis of coffee leaf diseases.

The proposed work is a customized 7-layered Convolutional Neural Network (CNN) with Inception, Global Context, and Multi-head Attention modules. The integration of these modules improved the performance of the network when compared with recent works on coffee leaf disease classification. This proposed network is trained on the BRACOL dataset, consisting of image samples of healthy and diseased Arabica coffee leaves. The contributions of this research are summarized as follows:

- A novel customized CNN is developed to accurately classify the specific disease affecting the leaves of Arabica coffee plants.
- With the integration of Inception, Global Context, and Multi-head Attention modules, the network can capture complex representations of the features, enhancing the network's ability to distinguish between diseases while maintaining computational costs.

The remainder of the paper is structured as follows. Section 2 provides an elaboration on existing works on coffee leaf disease detection. Section 3 outlines the proposed model and its components. Section 4 presents the environmental setup, the experiments conducted, and a discussion of the acquired results. Finally, Section 5 concludes the paper.

2. Related works

Numerous research works have focused on the development of automated disease detection systems for plant leaves of various species (Sharma et al. 2023). The following section analyses existing works conducted in the field of coffee leaf disease detection, and it is classified into two categories: (1) Machine Learning methods and (2) Deep Learning methods.

2.1. Machine learning-based methods

The integration of machine learning methods, combined with feature engineering and model optimization techniques, facilitates advancements in the field of plant leaf disease classification (Singh et al. 2023). Recent literature on coffee leaf disease detection utilizing supervised learning methods encompasses prevalent algorithms, including Naive Bayes (NB), Decision Trees (DT), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forests (RF), and others. Such methods involve the extraction of features from input images of coffee leaves and the utilization of machine learning algorithms to achieve an accurate diagnosis.

Syahputra et al. employed the NB algorithm to discriminate among the symptoms exhibited by coffee plants and to create a system that enables the detection of pests and diseases (Syahputra et al. 2020). Marin et al. presented a method for detecting coffee leaf rust using images obtained using an unmanned aerial vehicle. They used several decision tree models, including the logistic model tree, J48, ExtraTree, REPTree, Feature tree, Random tree, and RF, for classification and evaluated their performance. It was inferred that the logistic model tree method was more effective in predicting coffee leaf rust than other methods (Marin et al. 2021). Essoh et al. utilized a range of texture descriptors obtained using a three-dimensional gray-level co-occurrence matrix on color images to classify diseases. Principal component analysis was employed to select relevant descriptors and combined multiclass SVM and KNN algorithms for classification (Essoh et al. 2022). Abrham et al. utilized the median filtering technique to remove noise, the K-means technique was utilized for image segmentation, and performance was evaluated using the Back-Propagation Neural Network (BPNN) classifier (Abrham et al. 2018). Oliveira et al. applied different supervised machine-learning algorithms, such as RF, KNN, and U-Net, to detect nematode infection in coffee plantations (Oliveira et al. 2019). Miranda et al. preprocessed satellite images of coffee plants and utilized RF, Naïve Bayes, and Multilayer perceptron algorithms to detect coffee berry necrosis. Experiments showed that Naïve Bayes exhibited higher performance than the other methods (Miranda et al. 2020). Chowdhury et al. developed a coffee leaf recognition system utilizing Gabor filter-based feature descriptors. Classification was performed using SVM, NB, AdaBoost, and KNN algorithms, with SVM exhibiting better performance compared to the other classifiers (Chowdhury and Burhan 2021).

The efficacy of machine learning techniques depends heavily on the appropriate selection of discriminative features. Classical machine learning algorithms necessitate the use of hand-engineered features, which entails the need for domain expertise. In contrast, deep learning algorithms learn features incrementally through hidden layers, eliminating the need for complex feature engineering.

2.2. Deep learning-based methods

Deep learning algorithms can be applied directly to input data without the need for hand-engineered features. It is now possible to train deep learning models efficiently by leveraging parallelism with the help of high-performance computing and graphics processing units. Recent works in designing deep learning networks for the classification of plant leaf diseases are extensive and continually advancing (Ganguly et al. 2022; Pandey and Jain 2022).

Various deep learning models have been proposed for training leaf images to identify diseases, and many studies have utilized state-of-the-art architectures, such as AlexNet, ResNet, VGG16, and GoogleNet to detect infections in coffee leaves. Kumar et al. employed transfer learning to enhance the accuracy and robustness of a convolutional neural network model used to classify cropped coffee leaf images. The model utilized the InceptionV3 model in conjunction with dense and flattened layers for classification and was trained on an augmented dataset (Kumar et al. 2020). Similarly, Suparyanto et al. detected

Hemileia vastatrix (fungus responsible for coffee leaf rust) in coffee leaves using ResNet18 (Suparyanto et al. 2022). Montalbo et al. utilized a VGG16 network to classify Barako coffee leaf images into four categories: healthy, rust, infested by pest, and brown spot leaves. The performance of the network was evaluated by varying its hyperparameters (Montalbo and Hernandez 2020a). Hasan et al. proposed an automated leaf disease diagnosis method that utilized an extended Gaussian kernel density estimation approach to evaluate the probability of the nearest shared neighborhood to cluster symptoms. This approach was combined with the ResNet50 classifier, and the experimental results displayed reduced classification errors (Hasan et al. 2023). Binney et al. conducted a study on the automatic classification of diseases in Kenyan Arabica coffee leaves, employing three pre-trained architectures, namely ResNet50, DenseNet-121, and VGG19. It was inferred that the performance of DenseNet-121 surpassed the other architectures examined (Binney and Ren 2022). Barbedo et al. used a diverse collection of leaf images, including coffee and 11 other plant species, and used GoogleNet architecture and 10-fold cross-validation for disease classification. The model was capable of identifying several disease classes among the coffee species, such as bacterial blight, brown eye spot, brown leaf spot, leaf miner, rust, and blister spot (Barbedo 2018). Javierto et al. presented a MobileNetV2 model combined with YOLOv3 for the classification of coffee leaf diseases (Javierto et al. 2021). Similarly,

Faisal et al. proposed a hybrid feature fusion methodology that incorporates MobileNetV3, Swin Transformer, and variational autoencoder for disease detection in robusta coffee leaves (Faisal et al. 2023b). Yamashita et al. employed the MobileNet network in two ways: a single-stage model for classifying an input coffee leaf image into one of six categories, and a cascaded two-stage model. The latter starts by classifying the input image into one of four categories. If the category was 'Brown spots', the input is passed to the second stage, which utilizes the same architecture to differentiate between Cercospora, Leaf Miner, and Phoma spots (Bordin Yamashita and Leite 2023). Likewise, Waldamichael et al. proposed an HSV color segmentation algorithm to segment coffee leaf portions from the background and employed a MobileNetV2 classifier to detect Rust, Phoma, Cercospora, or Miner in such images (Waldamichael et al. 2021). Similarly, Faisal et al. proposed a novel dense fusion CNN (DFNet), employing multiple automated feature extractors, MobileNetV2 and NASNetMobile, to enhance plant leaf disease classification for the disease detection of robusta coffee leaves (Faisal et al. 2023a). Ramamurthy et al. presented an improvised EfficientNet-B0 by incorporating a ghost module at the end of the architecture to effectively classify Leaf miner and Rust in Arabica coffee plants (Ramamurthy et al. 2023). Fan et al. introduced a novel framework that integrates the features of InceptionV3 and the Histogram of Gradient feature extraction algorithm for the classification of coffee leaf diseases (Fan et al. 2022). Montalbo et al. used images of Barako coffee leaves to train three pre-trained models, namely VGG16, Xception, and ResNetV2–152, each with different hyperparameter settings, and their performances were analyzed (Montalbo and Hernandez 2020b). Novtahaning et al. proposed an ensemble-based deep learning approach that utilizes three fine-tuned state-of-the-art architectures - EfficientNet-B0, VGG16, and ResNet-152, for the classification of coffee leaf diseases (Novtahaning et al. 2022).

Esgario et al. employed UNet and PSPNet architectures for semantic segmentation to extract leaf boundaries and then utilized pre-trained models (AlexNet, GoogleNet, VGG16, and ResNet50) to classify diseases from segmented leaf images. It was concluded that the UNet architecture outperformed PSPNet for segmentation and that the ResNet50 architecture was the most accurate for the classification task among the analyzed models (Esgario et al. 2022). Tassis et al. suggested a three-stage approach for detecting coffee leaf disease. First, they utilized Mask R-CNN for instance segmentation, followed by UNet and PSPNet for the semantic segmentation of diseased spots on leaves. Finally, ResNet was employed for disease classification (Tassis et al. 2021). Yebasse et al. utilized two approaches to classify robusta coffee leaf images as healthy or diseased: a naive approach using ResNet and a guided approach combining U2Net and ResNet (Yebasse et al. 2021). Similarly, Martinez et al. proposed a three-stage approach for coffee leaf disease detection, comprising a leaf detection unit, a preprocessing unit to identify diseased patches on a leaf, and a ResNet for the exact identification of disease on the leaf (Martinez et al. 2022). Montalbo proposed the Triple-Deep Convolutional Neural Network (DCNN) method, utilizing an ensemble of three aggregated DCNN models to enhance accuracy and mitigate bias. The approach utilized a three-stage process to refine classification options and diversify the feature pool, which yielded better results than other state-of-the-art DCNN models (Montalbo 2022).

Few works have employed few-shot and multi-task learning methods to train their networks for the classification of coffee leaf diseases. Tassis et al. employed five pre-trained models, namely ResNet50, MobileNetv2, VGG16, DenseNet-121, and EfficientNet-B4, as the backbones of TripletNet and ProtoNet, to classify coffee leaf diseases. The experimental results indicated that ProtoNet outperformed TripletNet in terms of classification accuracy, and EfficientNet-B4 was the most effective backbone, exhibiting higher metric scores than the other four models (Tassis and Krohling 2022). Afifi et al. employed a few-shot learning approach and utilized three pre-trained architectures (ResNet18, ResNet34, and ResNet50) to construct two baseline models: a Triplet network and a deep adversarial Metric Learning (DAML) approach (Afifi et al. 2020). Esgario et al. proposed single-task and multi-task learning methods using various pre-trained architectures to classify coffee leaf disease and assess its severity (Esgario et al. 2020).

In addition to the existing pre-trained architectures, some research works have suggested novel custom architectures for classifying coffee leaf diseases. De Vita et al. implemented a CNN consisting of four convolutional layers and two fully connected layers to identify four classes: healthy, miner, phoma, and rust (De Vita et al. 2020). Similarly, Divyashri et al. designed a CNN to detect brown eye spots, leaf blight, rust, and miner in coffee leaves (Divyashri et al. 2021). Madhukar et al. designed a CNN of four pairs of convolutional and max pooling layers, followed by two dense layers, to diagnose the diseases on coffee leaf images (Madhukar et al. 2022). Wallelign proposed a five-layered CNN to detect rust and wilt on coffee leaves (Wallelign 2020). Sorte et al. performed texture extraction on images of coffee leaves and used a simplified version of the AlexNet model with five convolutional layers and three fully connected layers to detect rust and brown eye spots on the leaves (Sorte et al. 2019). Marcos et al. manually collected coffee leaf images and built a custom CNN model to detect rust. In addition to detecting rust, the model also performed segmentation and highlighted the regions of rust within the images (Marcos et al. 2019). Shubhashini preprocessed coffee images using contrast enhancement and resizing techniques. They proposed a customized multi-layered CNN named 'LeNet', based on the AlexNet architecture for classification (Shubhashini pal. 2021).

The above-mentioned methods demonstrate the effectiveness of deep learning models in detecting coffee leaf diseases. However, this research has identified certain limitations, such as class imbalance, inadequate class-wise adaptability, and insufficient attention to important leaf features. The following subsection discusses these issues and the steps taken to address these research gaps in the proposed work.

2.3. Research gaps and motivation

Despite the availability of various methods to detect diseases in coffee leaves, there are still a few challenges that need to be overcome in this field.

1) Many studies have been conducted using limited datasets that often exhibit class imbalance, leading to poor generalization of the network. Hence, to improve the effectiveness of the trained model, it

is recommended that the model be evaluated using larger and more diverse datasets.
2) Several deep learning models employ widely used and validated architectures, such as GoogleNet and ResNet. These models are frequently characterized by their significant size and complexity, which can pose challenges for deployment on low-resource devices. The linear arrangement of convolutional layers in these models tends to result in intricate computations and a higher number of trainable parameters, thus hindering their practical utility in real-world applications.
3) Studies related to coffee leaf disease detection have proposed models that exhibit limited feature learning based on the long-range dependencies of image pixels. Enhancing the network by considering a broader context beyond local image patches can contribute to a comprehensive understanding of features and improve the network's diagnostic ability.
4) The majority of deep learning models assign equal importance to all features derived from various levels. However, to enhance the model's sensitivity in classification, it is necessary to incorporate feature weighting at each stage. This approach enables the identification and propagation of significant features into deeper layers of the network, facilitating more accurate classification.

2.4. Research contributions

The following are the research contributions presented in the proposed work.

1) To address the issue of class imbalance and improve the generalizability of the network, a range of extensive data augmentation techniques were employed to increase the number of images in the dataset. This approach ensured the reliability of the proposed network and helped prevent overfitting.
2) The proposed network employs inception modules for the parallel computation of input feature maps, resulting in better performance while maintaining computational efficiency. In comparison to the other deep learning techniques discussed in the literature, the proposed network learned approximately 7.6 M parameters for disease classification.
3) The global context block present in the proposed network extracts global contextual information from the features, thereby promoting an overall understanding of the image. This block aggregates information across the entire spatial extent of the feature map and employs channel attention to convert this information into a solitary feature vector. This vector is modulated into the input feature map, thereby incorporating a richer understanding of the overall structure of the input.
4) The multi-head attention module employed in the proposed network partitions the input feature map into subsets, calculates the interactions between the features, and subsequently combines them to form a comprehensive representation of essential features that contribute to the appropriate disease class. The important features were provided more weightage and passed onto succeeding layers for precise classification.

3. Methodology

This section presents the methodology proposed for disease classification in coffee leaves. The forthcoming sections provide a detailed explanation of the proposed network architecture, including the significance of each integrated module.

3.1. Architectural overview of the proposed network

In this research, a linear 2D CNN network that incorporates Inception modules, Global Context (GC), and Multi-Head Attention (MHA) is proposed. The main objective of using inception modules is to enable the network to learn representations at different spatial scales while simultaneously limiting the number of parameters required by the network. The GC block is a computational module that utilizes channel attention to produce a solitary feature vector. This vector is then used to adjust the input feature maps and extract global contextual details. The purpose of these modulated feature maps is to improve downstream task performance by providing pertinent contextual information. The multi-head attention module integrated into the proposed network captures complex relationships between features from different subsets and aggregates them to form a more powerful representation of the input. The architectural sketch of the proposed network is presented in Fig. 1, providing information on the layer arrangement and computational procedure. The algorithm of the proposed network is given as follows:

Algorithm for the Proposed System
Input: Image Dataset (X), Labels (Y).
Output: Trained Proposed Model

1) Data Preprocessing:

- Splitting X and Y:
  - X_{train}, Y_{train} – Training set
  - X_{val}, Y_{val} – Validation set
  - X_{test}, Y_{test} – Testing set
- Data augmentation techniques:
  - Rotation: 0° – 20°
  - Horizontal & Vertical flipping
  - Random Brightness adjustment: 0.2–1.5
  - Contrast: 14–20
  - Saturation Adjustment: 14–20
  - Noise: Up to 35%

2) Model Definition:
   a Define the proposed model (*ProposedNet*):
   - Apply a 3 × 3 convolutional layer with ReLU activation to input X.
   - Pass the output to an Inception module, which includes different filter sizes (1 × 1, 3 × 3, 5 × 5) and max pooling.
   - Apply a 3 × 3 convolutional layer with ReLU activation and batch normalization.
   - Pass the output to another Inception module.
   - Apply a 3 × 3 convolutional layer with ReLU activation and batch normalization.
   - Apply max pooling with a pool size of (3,3) and apply dropout.
   - Apply a 3 × 3 convolutional layer with ReLU activation and batch normalization.
   - Compute attention weights by applying a 1 × 1 convolution and softmax function. Use a 1 × 1 convolution for channel-wise attention mapping (squeeze-and-excitation).
   - Concatenate the output of the attention module with the original feature map from the GC module.
   - Apply a 3 × 3 convolutional layer with ReLU activation and batch normalization.
   - Apply max pooling with a pool size of (3, 3) and apply dropout.
   - Apply a 3 × 3 convolutional layer with ReLU activation and batch normalization.
   - Pass the output to the multi-head attention module. Compute query (Q), key (K), and value (V) matrices for each segmented region of the feature map using learnable weights. Calculate attention scores using a scaled dot product attention function. Obtain the weighted sum for each attention head and merge the outputs by projecting

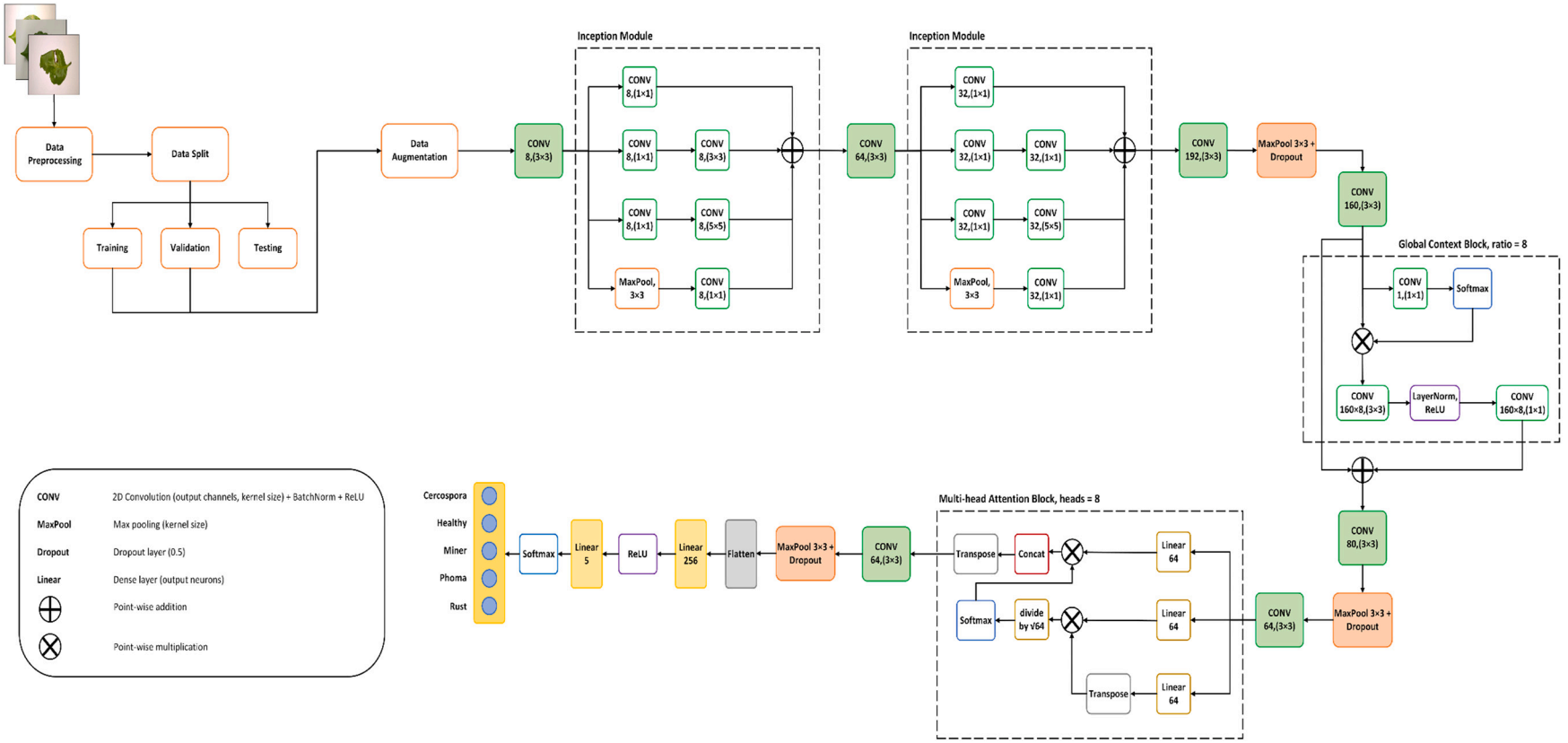


Fig. 1. Architectural sketch of the proposed Inception-based Global Context Attention CNN for Coffee leaf disease detection.

them back to the original dimension through concatenation.
- Apply a 3 × 3 convolutional layer with ReLU activation and batch normalization.
- Apply max pooling with a pool size of (3, 3) and apply dropout.
- Flatten the feature matrix into a one-dimensional vector.
- Apply a dense layer to generate a 256-dimensional output vector with ReLU activation.
- Apply a second fully connected layer to generate a 5-dimensional output vector, followed by the Softmax function to obtain class probabilities.

3) Hyperparameter tuning and Model Compilation:

- Optimization algorithm: Adaptive Moment estimation (Adam)
- Learning rate = 0.0001, Weight decay: 0.0001
- Loss function: Categorical cross-entropy

4) Model Training:

- Initialize the model weights.
  - ➢ Repeat the following steps until the model reaches convergence:
    - For each training example (X_{train}, Y_{train}):
      - Perform forward propagation:
        - train_outputs = *ProposedNet*(X_{train})
      - train_loss = *CategoricalCrossEntropy*(train_outputs, Y_{train})
      - Perform backpropagation:
        - train_loss.backward()
        - optimizer.step()
      - Compute validation accuracy by comparing train_outputs & Y_{train}
    - For each validation example (X_{val}, Y_{val}):
      - Perform forward propagation:
        - val_outputs = *ProposedNet*(X_{val})
      - val_loss = *CategoricalCrossEntropy*(val_outputs, Y_{val})
      - Compute validation accuracy by comparing val_outputs & Y_{val}
    - Record training and validation accuracy and loss values.

5) Model Evaluation:

- For each testing example (X_{test}, Y_{test}):
- Perform forward propagation on X_{test}
- Y_{pred} = *ProposedNet*(X_{test})
- Using Y_{pred} & Y_{test}, compute the evaluation metrics such as Accuracy, Precision, Recall, F1-Score, Specificity, and Cohen's Kappa Score

End Algorithm

3.2. Inception module

Conventional deep neural networks generally consist of a linear stack of layers, but this design presents a trade-off between network performance and computational costs. Adding more layers and modules to a network can improve its accuracy, but may also increase computational expense, resulting in decreased efficiency. On the other hand, a simpler network design may achieve higher computational efficiency, but it could suffer from reduced consistency and accuracy in its results. To achieve an optimal balance between performance and computational efficiency, the proposed network architecture incorporates Inception modules (Szegedy et al. 2015).

By employing filters of various sizes, such as 1 × 1, 3 × 3, and 5 × 5, in parallel within the same layer, the Inception module is designed to extract features at multiple scales. Smaller filters (e.g., 1 × 1 or 3 × 3) are good at capturing local details and patterns, while larger filters (e.g., 5 × 5) can capture more global information. Parallel convolutions with different filter sizes ensure that the network can effectively capture features at multiple scales. By using this design, the module can apprehend the characteristics of dissimilar spatial dimensions and reduce the number of parameters acquired in the network. The Inception module comprises a bottleneck layer that uses 1 × 1 filters to perform dimensionality reduction on the input feature maps. These bottleneck layers reduce the number of input channels, which helps reduce computational costs. Additionally, they enable the network to learn combinations of features from different channels, allowing for richer and more diverse representations. After passing through the bottleneck layer, the resulting feature maps are sent through parallel branches. Each branch applies a distinct set of convolution filters to the feature maps. These branches capture different types of patterns and spatial scales, enabling the model to capture both local and global information, leading to richer and more expressive representations. The schematic overview of the inception module is presented in Fig. 2.

3.3. Global context block

GC block is a novel, flexible, and efficient computation component that enhances the performance of a network designed for an image classification task (Cao et al. 2019). It incorporates global context information into feature representations, enabling the network to have a better understanding of the overall structure of the input. This module consists of three steps: context modelling, transformation, and fusion.

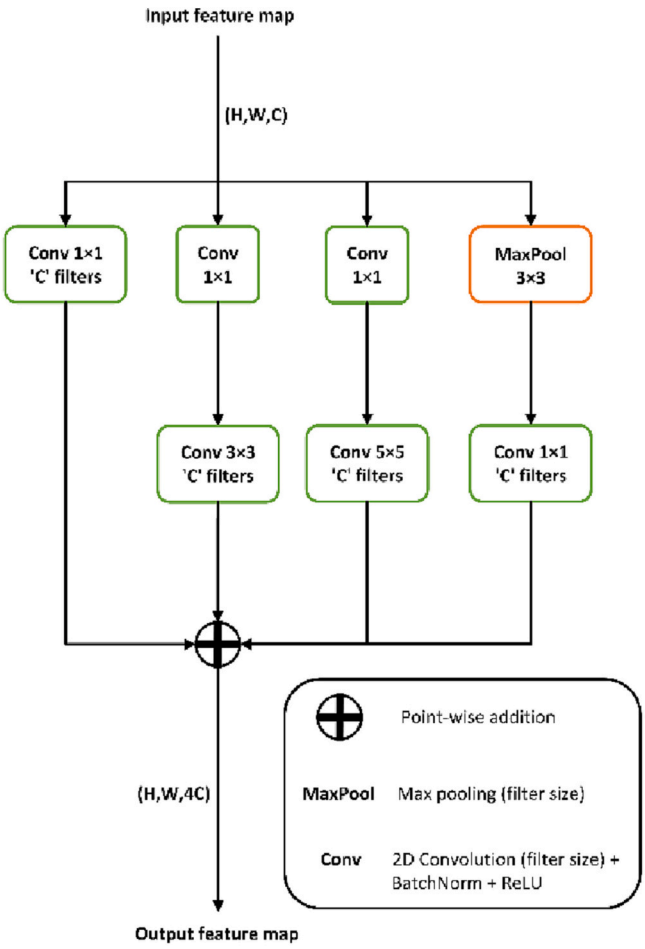


Fig. 2. Schematic overview of the Inception module used in the proposed system.

An overview of the GC block is presented in Fig. 3.

Long-range dependencies refer to the relationships between distant elements or features in the input data. Typical CNNs calculate features by aggregating information from local pixel neighborhoods, which can capture local information but may not capture long-range dependencies between distant elements or features in the input data. To represent such long-range dependencies between pixels, the non-local network in the GC block computes the output at each position by considering a weighted aggregation of features from all other positions in the feature map. The simplified non-local block in the proposed system has three main steps. First, a 1×1 convolution and a softmax function are applied to obtain attention weights and global context features via global attention pooling. Next, the features are transformed using a 1×1 convolution. Finally, the global context features are combined with the features of each input via feature aggregation. The Transform component of the GC block applies squeeze-and-excitation to the feature map obtained from the non-local network, resulting in channel attention. This component consists of a 1×1 convolution that spatially squeezes the feature map with a bottleneck ratio to obtain a channel descriptor. It is then passed through layer normalization, ReLU, and a 1×1 convolution to produce a channel-wise attention map. This map is used to reweight the feature maps and enhance important features.

The Fusion module adds global context information to the feature map by concatenating the solitary feature vector of the GC module with the original feature map. This improves accuracy and efficiency by making the network use global context information. The GC module is a lightweight subnetwork that is added to the main network, and the fusion module integrates information from the GC module into the main network.

3.4. Multi-head attention block

Recently, there has been a growing trend in the use of attention mechanisms in convolutional neural networks (CNNs). This trend has gained significant attention due to its potential to enhance network performance. The MHA mechanism enables a network to concentrate on the most significant regions of an input image (Vaswani et al. 2017). It does this by assigning attention weights to the features that are relevant to identifying the disease class of the image. When presented with an input sequence 'X' in the form ($batch\_size, sequence\_length, d_{model}$), where '$d_{model}$' denotes the dimension of the hidden representation prior to attention calculation, the MHA mechanism follows the subsequent sequence of steps. A graphic depiction of the MHA module employed in the proposed system is presented in Fig. 4.

The initial step involves the computation of the query (Q), key (K), and value (V) matrices, which are presented in Eqs. 1, 2, and 3, as documented in (Vaswani et al. 2017).

$$\mathrm{Q_i} = \mathrm{XW_q} \tag{1}$$

$$\mathrm{K_i} = \mathrm{XW_k} \tag{2}$$

$$\mathrm{V_i} = \mathrm{XW_v} \tag{3}$$

Here, 'i' signifies the computation carried out at the i^{th} attention head, while W_q, W_k, and W_v denote the learnable weights. The dimensionality of each attention head 'd_v', is determined as the result of dividing the dimensionality of the entire input feature map 'd_{model}', by the number of attention heads, denoted by 'n'. Each head is responsible for attending to a distinct region of the feature map. The attention scores for each head (A_i for the i^{th} head) were calculated using a scaled dot-product attention function defined by Eq. 4.

$$\mathrm{Ai} = Softmax\left(\frac{QK^T}{\sqrt{d_v}}\right) \tag{4}$$

The attention mechanism computes the attention scores for each region of the feature map, which are utilized to generate a weighted sum of the corresponding feature map values. The weighted sum for each head denoted as MA_i is determined according to the relation presented in Eq. 5.

$$\mathrm{MA_i} = \mathrm{A_i} \times \mathrm{V_i} \tag{5}$$

Lastly, the outputs from all heads are concatenated, and the resultant tensor is projected back into the original dimension through a concatenation operation, followed by an output calculation, as specified in Eq. 6.

$$\mathrm{MA_{output}} = Concat(\mathrm{MA_1}, \mathrm{MA_2}, \mathrm{MA_3}, \ldots, \mathrm{MA_h}) \times \mathrm{W_0} \tag{6}$$

Here, the matrix W_0 denotes the weight matrix of shape ($n \times d_{head}$, d_{model}). The multi-head attention process enables the network to grasp diverse relationships between features present in the data, thereby enhancing the expressive power of the overall network.

The proposed network integrates this module to generate attention maps that encapsulate a more comprehensive representation of image features. The module enables the network to attend selectively to the

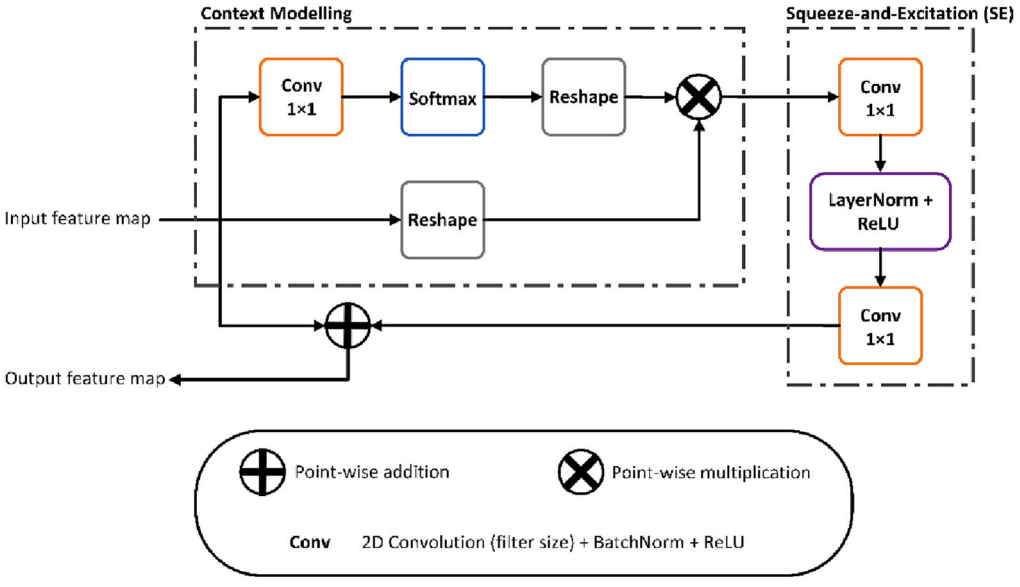


Fig. 3. Overview of the GC block used in the proposed system.

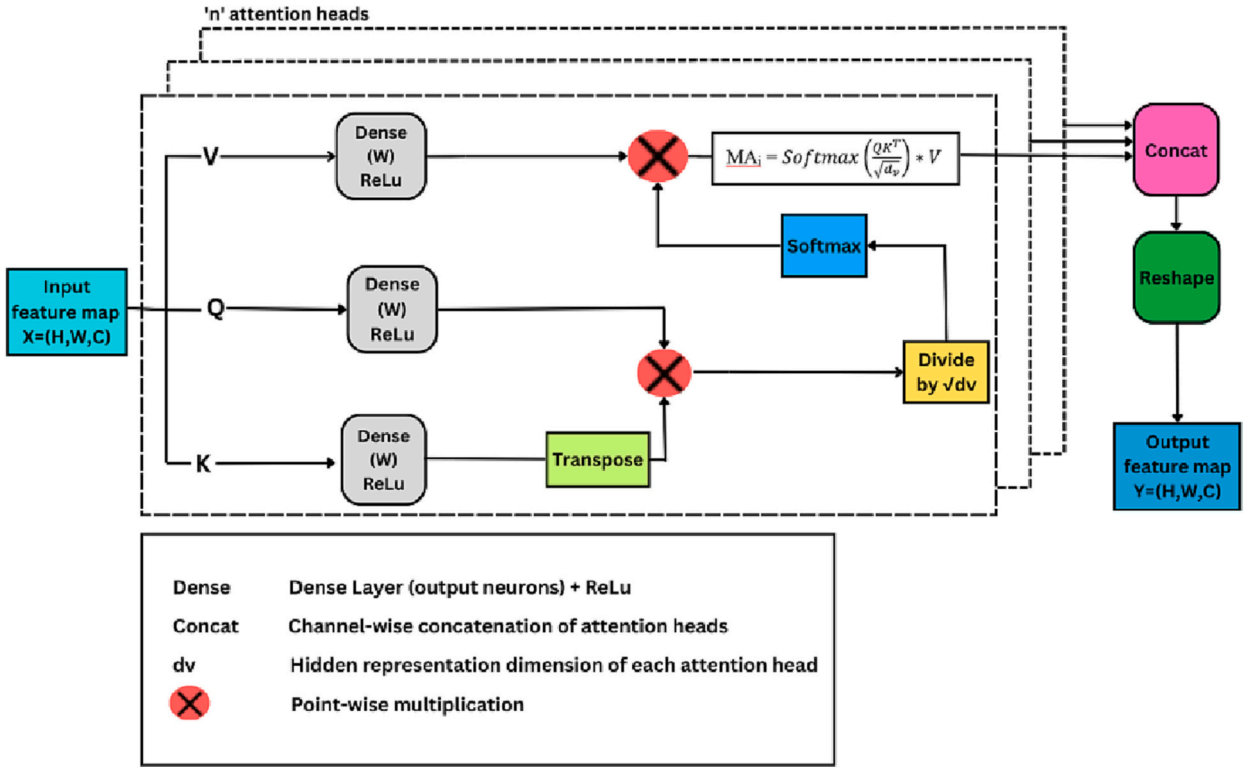


Fig. 4. Schematic overview of the Multi-head attention module used in the proposed system.

relevant regions of the image, thereby resulting in better classification accuracy and precision.

4. Results and discussion

This section provides an overview of the dataset used to train the proposed network, the data augmentation approach, model experimentation, and validation. Additionally, the last subsection evaluates the model's performance using various performance metrics.

4.1. Dataset description

This research employed the publicly available BRACOL dataset, which comprises 1747 images of Brazilian Arabica coffee leaves (Krohling et al. 2019). These images have a dimension of 2048 × 1024 pixels. The dataset utilized in this study includes labeled images that are classified into two distinct categories: healthy and diseased. The diseased group is further subcategorized into four specific diseases: Rust, Miner, Cercospora, and Phoma. The images, comprising one leaf per image, were captured under uniform lighting and background conditions. Additionally, these images were gathered at diverse times throughout the year. The sample images of the different categories are presented in Table 1.

4.2. Data augmentation

The presence of imbalanced data within the dataset can result in overfitting and suboptimal convergence during the training process. To address this issue, a set of augmented images was generated using various data augmentation techniques implemented via the Keras and Pillow libraries. The following methods were employed: (1) Random rotation of images around their center by an angle ranging from 0 to 20°, (2) Random horizontal and vertical flipping, (3) Random brightness adjustment of the image within a range of 0.2 to 1.5, (4) Contrast adjustment of 14 to 20, (5) Saturation adjustment of 14 to 20, and (6) Addition of noise up to 35%. The dataset originally consisted of 1747 images and was partitioned into three subsets: a training set comprising 60% of the data, a validation set comprising 20% of the data, and a test set comprising the remaining 20% of the data. Following the dataset split, an independent augmentation process was conducted on the training and validation sets. Consequently, the combined number of images included in both the training and validation sets increased to 7500, with 5000 images allocated to the training set and 2500 images allocated to the validation set.

4.3. Environmental setup

The proposed experiments in this research were executed utilizing PyTorch, an open-source framework based on Python, and the Torch library. All experiments were carried out on Azure Virtual Machines and Google Colab, which provided NVIDIA Tesla P40 GPU for training the deep learning network. The source code files used in the upcoming experiments, along with the proposed network, are available at: https://github.com/Joshua-Alfred/Inception-GC-MHA-based-CNN.

4.4. Hyperparameter tuning

The optimization and analysis of hyperparameters are crucial steps in developing deep learning networks, as they significantly impact a model's capacity to learn and generalize input data. This study focuses on enhancing the model's performance by optimizing the hyperparameters using the grid search algorithm: (1) dropout rates in all dropout layers, (2) learning rate, (3) weight decay, (4) batch size, and (5) gradient update optimization algorithm. The study conducted a grid search algorithm using the Ray tune framework to find the best set of hyperparameters from the search space of dropout probability values ranging from 0.4 to 0.7, learning rate and weight decay parameters ranging from 0.0001 to 0.1, and gradient optimizers, namely Adam, Batch Gradient Descent (BGD), and Stochastic Gradient Descent (SGD). The results indicated that a dropout rate of 0.5 yielded the best model

Table 1
Sample healthy and affected coffee leaf images in the BRACOL dataset.

Input Class	Image Samples
Healthy	
	
	
Leaf Rust	
	
Phoma	
	
	
Miner	
	
	
Cercospora	
	

performance. Furthermore, the learning rate and weight decay parameters led to maximum network convergence at a value of 0.0001. The study also showed that the Adam optimizer outperformed SGD and BGD in obtaining better results. Table 2 presents the optimal parameter values acquired from the hyperparameter tuning experiment.

4.5. Evaluation metrics

This research used different evaluation metrics to determine the performance of the trained model, including Accuracy, F1-score, Recall, Precision, Specificity scores, and the confusion matrix. A confusion matrix was used to display the model's distribution of actual and predicted disease classes. By examining the matrix, performance metrics for the model were derived, including True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) for each class. Accuracy is the ratio of correctly predicted instances (TP and TN) to the total number of instances, providing an overall measure of the model's correctness. Precision is the proportion of correctly predicted positive instances (TP) to the total number of instances predicted as positive (TP and FP), indicating the model's ability to avoid false positives. F1-Score is the harmonic mean of precision and recall, balancing both metrics and providing a single value that represents the model's overall performance. Specificity: The proportion of correctly predicted negative instances (TN) to the total number of actual negative instances (TN and FP), reflecting the model's ability to identify all negative instances while avoiding false positives. Moreover, these metrics are used to derive Cohen's Kappa score, which is a statistical measure that quantifies the level of agreement between two evaluators or raters in a classification task while accounting for the potential of agreement by chance. The score is normalized to a range between −1 and 1, where a value of 1 indicates perfect agreement, 0 represents the level of agreement expected by chance, and values below 0 indicate that the level of agreement is worse than expected by chance. Eq. 12 represents the equation used to calculate Cohen's kappa score, as documented in (Cohen 1960).

The AUC score was also calculated, representing the model's ability to predict the correct disease class. This score was determined by the area under the two-dimensional Receiving Operating Characteristics (ROC) curve, which describes the probabilistic variation between Recall and False Positive rates. The corresponding relations for each metric are presented below.

$$Accuracy = \frac{TP + TN}{Total\ number\ of\ data\ points} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1\ Score = \frac{2{*}Precision{*}Recall}{Precision + Recall} \tag{10}$$

$$Specificity = \frac{TN}{TN + FP} \tag{11}$$

$$Cohen\ Kappa\ score = \frac{(P_o - P_e)}{(1 - P_e)} \tag{12}$$

Table 2
Hyperparameter experimentation results.

Parameters	Search Space	Optimal value
Dropout	[0.4, 0.5, 0.6, 0.7]	0.5
Learning Rate	[0.0001, 0.001, 0.01, 0.1]	0.0001
Weight decay	[0.0001, 0.001, 0.01, 0.1]	0.0001
Batch Size	[16, 32, 64]	32
Optimization Algorithm	[SGD, BGD, Adam]	Adam

Where P_o is the observed agreement probability between the raters, and P_e is the probability of chance agreement. The BRACOL dataset, consisting of Brazilian Arabica coffee leaf images, was utilized for training and testing the proposed network and other ablation experiments, as described below.

4.6. Ablation studies

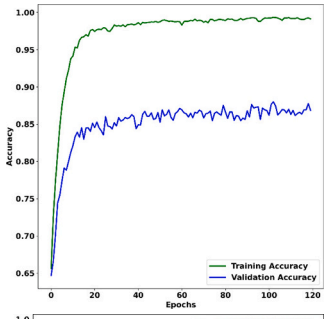
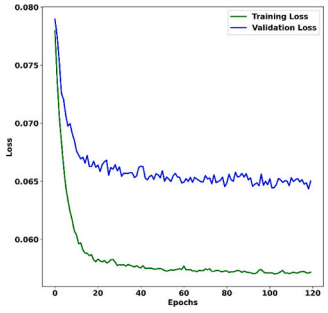
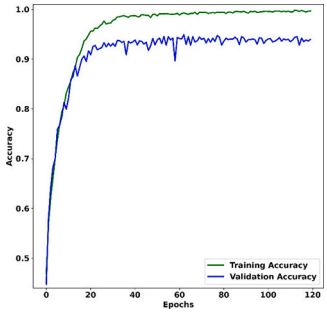
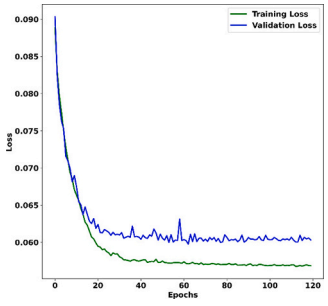
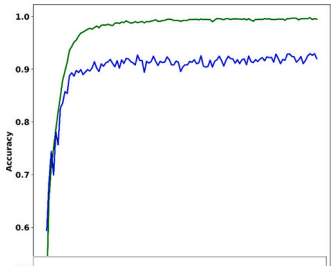
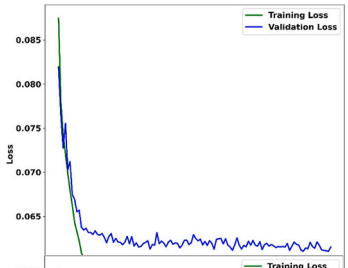
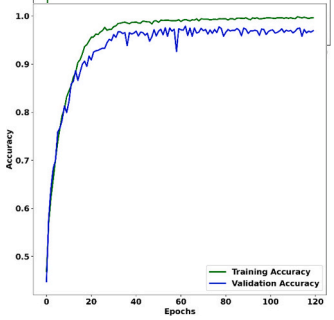
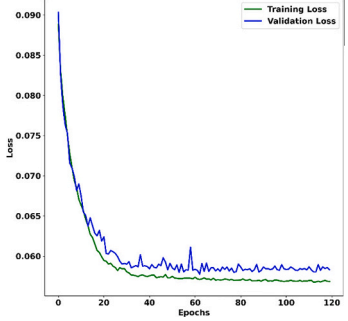
To demonstrate the individual contributions of each module within the proposed system architecture, a series of experiments and their corresponding results are presented in Table 4. The individual contributions of the Global Context, Multi-Head Attention, and Inception modules to the overall performance and computation cost of the proposed architecture were evaluated by integrating them separately with the baseline seven-layered CNN architecture. The outcomes of these experiments provide evidence of the importance and rationale for the overall testing performance of the proposed architecture.

Table 4 presents a summary of the ablation studies that evaluate the potency of the different modules employed in the proposed network. The assessment of the modules' performance is based on several metrics, including Accuracy, F1-score, Precision, Recall, and Specificity scores. The following subsections elaborate on the experimental setup and training process used in the study.

4.6.1. Analysis of the 7-layer linear CNN

This experiment analyses the performance of the baseline 7-layer CNN network. The linear layered CNN structure, devoid of additional modules, was trained for 120 epochs. The proposed network architecture consists of a sequence of convolutional blocks succeeded by batch normalization and the ReLU activation function. Within the seven

Table 3
Analysis of ablation experiments and their variations in accuracy and loss during network training.

Model	Accuracy vs Epochs graph	Loss vs Epochs graph
Baseline CNN		
Baseline CNN + Inception		
Baseline CNN + MHA Block		
Baseline CNN + Inception + GC Block		

convolutional blocks, three are further followed by max pooling and dropout layers to improve the network's generalization ability. Table 3 presents the analysis of the baseline CNN, showing the variations in accuracies and loss metrics over 120 epochs of network training. After training, the model was tested using an unseen collection of coffee leaf images, and the resultant testing accuracy was 85.71%. Additionally, other testing metrics evaluated from the model, namely Precision, Recall, F1, and Specificity, are presented in Table 4.

4.6.2. Effectiveness of the inception block

In this section, we have added the inception mechanism to the baseline 7-layer CNN network. Two inception blocks were added to the baseline network and trained for 120 epochs. The optimizer and loss function parameters chosen for this experiment are the same as those set in the 7-layer CNN experiment. Table 3 presents the analysis of the baseline CNN with Inception blocks, showcasing the variations in accuracies and loss metrics over 120 epochs of network training. The training accuracy increased at a fast pace initially and later stabilized. Overall, there was an improvement in the performance of the baseline CNN network when the inception mechanism was incorporated into the network. After training, the model was tested using an unseen collection of coffee leaf images, and the resultant testing accuracy was 92.28%. We can observe that the addition of inception to the baseline model increased the testing accuracy by 6.57%, and the increments in other metrics are presented in Table 4.

4.6.3. Effectiveness of the MHA block

In this section, we have added the multi-head attention module to the baseline 7-layer CNN network. The multi-head attention module computes attention scores for each section of the feature map and uses these scores to generate a weighted sum of the feature map values. This weighted sum is then used as an input to the fully connected layer, allowing the model to focus on the most relevant part of the image. It provides a more robust and effective feature representation, leading to improved accuracy and performance compared to traditional CNN models. The epoch counts, optimizer, loss function, weight decay, and other parameters used during experimentation were the same as those used in earlier experiments. Table 3 presents the changes in training and validation accuracy and loss values throughout the training, including AUC-ROC and the confusion matrix. An improved testing accuracy of 91.66% was obtained when compared to the baseline CNN experiment, indicating the effectiveness of the MHA block in extracting complex relationships between different subsets of the feature map that actively contribute to the diagnosis of coffee leaf disease. In addition, Table 4 presents other testing measures provided by the model.

4.6.4. Effectiveness of inception and GC blocks

This section presents the training of the 7-layered baseline CNN network integrated with Inception and GC modules. Through the utilization of the GC block, spatial correlation is observed between the targeted pixel and the entirety of the image, while the inception modules extract features across multiple abstraction levels from the image to produce highly representative feature maps. The neural network comprises two inception modules and a single GC module, each positioned independently amid the convolutional layers. The epoch counts, model hyperparameters and loss function implemented in this experiment were the same as those used in the above experiments. Table 3 presents the variations in training and validation accuracy and loss values during training. The baseline CNN network employed with the Inception and Global Context module obtained a testing accuracy of 96%.

4.6.5. Analysis of the proposed network

The proposed architecture is a combination of the modules integrated and experimented with earlier, i.e., Inception, Global Context, and Multi-Head attention to our baseline CNN network to carry out leaf disease classification. Each module provides a unique contribution to the enhancement of feature learning and training and has ultimately improved the performance of the overall architecture. Inception modules are responsible for maintaining computation costs while performing convolutions at different filter levels in parallel. Global context provides an understanding of a particular feature with respect to the whole image target, highlighting important features that play a vital role in identifying the disease class. Additionally, multi-head attention executes vanilla attention to equally divided features in parallel and produces an attention map that contains the relational information between features.

The proposed network was trained with the augmented BRACOL dataset, and an accuracy of 98.57% was obtained from the test set. The modules implementing global context and attention mainly contributed to performance improvement. Table 4 contains other metrics evaluated from the proposed architecture and their improvements with respect to the experiments conducted. Fig. 5 presents the variations in accuracy and loss metrics during network training, along with the AUC-ROC graph and confusion matrix obtained during network testing. Table 5 represents the GradCAM visualization at the last convolutional layer of the proposed network for a given image of a particular class, indicating the regions that contributed to disease detection. In conclusion, the proposed network achieved the best performance with reduced parameters.

4.7. Performance analysis

Table 6 presents a comparison between the performance of the proposed network and the existing methods for coffee leaf disease classification. As noted in the literature review, a significant number of research works have used pre-trained networks. Due to the pre-trained weights and decent degree of optimization, the pre-trained models perform effectively on a modest amount of data. However, it was observed that the models showed biases toward certain classes or attributes and lacked the ability to concentrate on the more important features; thus, they were not suitable for all classification tasks. This is addressed in our research work by using customized deep learning architectures embedded with inception, global context, and attention modules.

We have presented related research works that have performed multi-class coffee disease classification based on coffee leaf images that have used the BRACOL/similar coffee leaf datasets. Related works placed in comparison with our proposed work have performed model training using images containing a single healthy/diseased coffee leaf in each image. The comparative results presented in Table 6 indicate that the proposed network achieved better performance in terms of overall

Table 4
Summary of the ablation studies performed in the proposed work.

Model	Accuracy	Precision	Recall	F1-Score	Specificity	Cohen Kappa score	Number of parameters	Training Time (H:M:S)
Baseline CNN	0.8571	0.8569	0.8576	0.8572	0.9518	0.8219	6,997,477	11:59:47
Baseline CNN + Inception	0.9228	0.9228	0.9235	0.9133	0.9807	0.9040	7,168,053	12:04:12
Baseline CNN + MHA	0.9166	0.9154	0.9147	0.9147	0.9841	0.8957	7,014,117	12:09:35
Baseline CNN + Inception + GC	0.9600	0.9598	0.9602	0.9599	0.9906	0.9510	7,456,550	12:23:03
Proposed Network	**0.9857**	**0.9857**	**0.9854**	**0.9855**	**0.9964**	**0.9820**	**7,598,454**	**12:49:41**

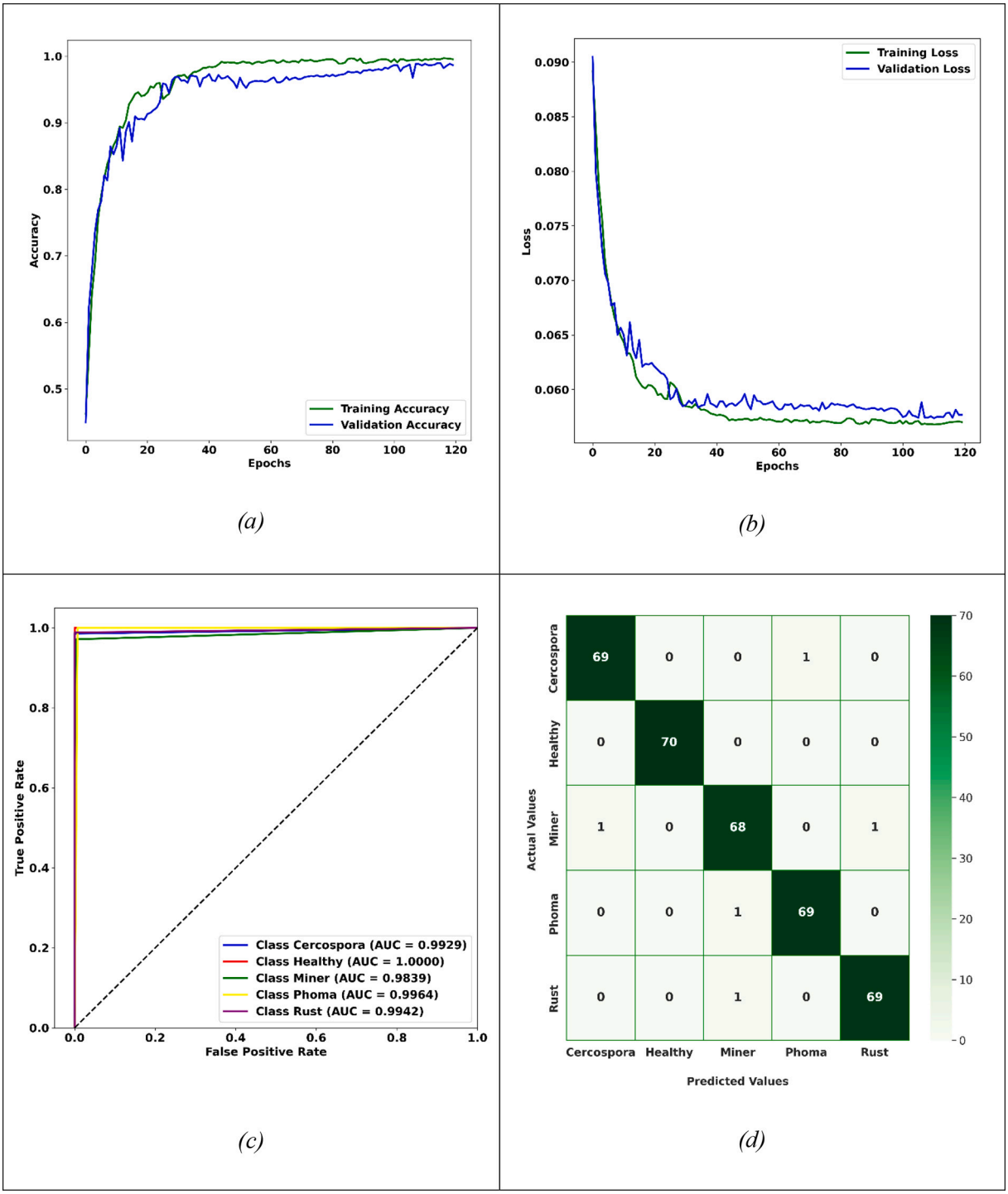
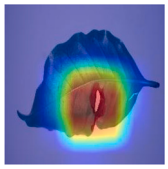
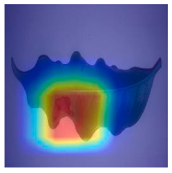
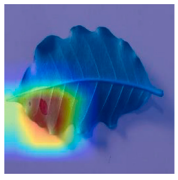
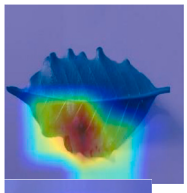
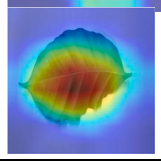


Fig. 5. – Analysis of the proposed network (a) Accuracy (b) Loss (c) ROC curve (d) Confusion matrix.

accuracy compared to existing works. Unlike other existing works, the proposed network utilizes attention and context modules, exhibiting the ability to identify information in an input most relevant to completing a classification task. Additionally, the inclusion of the inception module helped the network learn complex features easily, without insisting on an increase in network structure complexity and computation.

In addition, works performed with custom-made datasets have had disease classes similar to the classes set in our classification task and have been placed in comparison. Montalbo merged images of three different coffee species leaves and grouped them into seven classes – the five classes seen in the BRACOL dataset along with Sooty molds and Red Spider Mite (Montalbo 2022). Nevertheless, it should be noted that the three-staged pre-trained model approach employed in the research did not incorporate any additional feature learning modules; hence, it was not able to achieve the level of performance demonstrated by the proposed approach. Tassis & Krohling (Tassis and Krohling 2022) and Kumar et al. (Kumar et al. 2020) used transfer learning with pre-trained architectures and achieved an accuracy of 95.24 and 97.61, respectively.

Table 5
GradCAM visualizations for images of different classes.

S. No.	Class	Input image	GradCAM output
1.	Phoma		
2.	Miner		
3.	Cercospora		
4.	Rust		
5.	Healthy		

While transfer learning can be a powerful tool for achieving high accuracy rates, it is important to note that pre-trained architectures are not always well-suited for every task.

Moreover, the network proposed in this study demonstrated superior performance compared to recent research experiments conducted on the same dataset. De Vita et al. developed a custom convolutional neural network (CNN) consisting of convolutional, max pooling, and dropout layers, achieving an accuracy of 96.24% on the test set (De Vita et al. 2020). Novtahaning et al. proposed an ensemble network combining EfficientNet-B0, VGG16, and ResNet-152, achieving an accuracy of 97.31% (Novtahaning et al. 2022). Similarly, the performances evaluated in the research conducted by Kumar et al., Montalbo, and Tassis et al. were also outperformed by the proposed network, which achieved a remarkable testing accuracy of 98.57% on the BRACOL dataset (Kumar et al. 2020; Montalbo 2022; Tassis and Krohling 2022).

In contrast, the proposed custom CNN network was more effective in detecting coffee leaf disease compared to previous studies that did not use attention or context mechanisms. The proposed system uses the relationships between features and contextual information, which helps it learn features more precisely and achieve higher accuracy rates. Moreover, a comparison of the number of trainable parameters between the proposed network and state-of-the-art architectures is presented in Table 7. The proposed network contained 7,598,454 parameters and acquired a testing accuracy of 98.57%, while VGG16, with 134,281,029 parameters, exhibited the lowest testing accuracy of 90.67%. The analysis indicates that the proposed network outperforms state-of-the-art architectures by demonstrating enhanced performance, with a total number of parameters equal to 5.65% of the parameters present in VGG16. The number of parameters and the corresponding testing accuracy of other state-of-the-art architectures, such as AlexNet, ResNet50, InceptionV3, Xception, EfficientNet-B3, and DenseNet-121, are presented in Table 7. In contrast, DenseNet-121 had the lowest number of parameters compared to the proposed network, inferring that the proposed network is more computationally efficient than the other state-of-the-art architectures, except for DenseNet-121. Nevertheless, the performance of DenseNet-121 in classifying coffee leaf diseases was not significant compared to that of the proposed network.

Table 6
Performance analysis of the proposed work with recent research on coffee leaf disease detection.

S. No	Source	Methodology	Dataset	Accuracy (%)
1	Syahputra et al. (Syahputra et al. 2020)	Naïve Bayes	Custom-made	92.00
2	Abrham et al. (Abrham et al. 2018)	Back-propagation ANN & Decision Tree	Custom-made	94.50
3	Marcos et al. (Marcos et al. 2019)	Customized CNN	Custom-made	95.00
4	Tassis & Krohling (Tassis and Krohling 2022)	Transfer learning with MobileNetV2	BRACOL	95.24
5	Montalbo (Montalbo 2022)	Transfer learning using EfficientNet-B0, DenseNet-121 & VGG16	BRACOL + LiCoLe + RoCoLe	95.98
6	Essoh et al. (Essoh et al. 2022)	Multi-class SVM & KNN	Custom-made	96.00
7	De Vita et al. (De Vita et al. 2020)	Customized CNN	BRACOL	96.24
8	Novtahaning et al. (Novtahaning et al. 2022)	Ensemble architecture using EfficientNet-B0, VGG16 & ResNet-152	BRACOL	97.31
9	Kumar et al. (Kumar et al. 2020)	Transfer learning using InceptionV3	BRACOL	97.61
10	**Proposed Work**	**Customized CNN with Inception, GC, and MHA**	**BRACOL**	**98.57**

Table 7
Comparison with state-of-the-art architectures.

Model	Number of parameters	Accuracy	F1-Score	Specificity	Cohen Kappa Score
VGG16	134,281,029	0.9067	0.9061	0.9437	0.8832
AlexNet	57,024,325	0.9588	0.9579	0.9781	0.9484
Resnet50	23,518,277	0.9657	0.9646	0.9867	0.9571
InceptionV3	23,834,568	0.9543	0.9532	0.9872	0.9429
Xception	20,817,197	0.9439	0.9423	0.9835	0.9298
EfficientNet-B3	10,703,917	0.9608	0.9597	0.9768	0.9509
DenseNet-121	8,062,504	0.9081	0.9041	0.9603	0.8849
Proposed Network	**7,598,454**	**0.9857**	**0.9855**	**0.9964**	**0.9820**

5. Conclusion

Coffee leaf diseases are a major threat to the global coffee industry. The widespread consumption of coffee necessitates the effective management of these diseases, which can significantly reduce crop yield and quality. Existing machine learning and deep learning systems for coffee leaf disease detection have limitations, such as low accuracy and poor inter-class metrics. To address these limitations, a custom-layered convolutional neural network that utilizes inception modules, a global context module, and a multi-head attention module is proposed to obtain precise classification. Inception modules capture significant feature maps at different levels of abstraction, while the global context module incorporates global contextual information to enhance diagnostic accuracy. The multi-head attention module captures dependencies among features from different subsets of the feature map, resulting in a richer representation of the input.

The proposed network achieved an accuracy of 98.57%, which outperforms existing methods on coffee leaf disease detection. The network is also generalizable to other plant species, such as rice and maize. Furthermore, the network size can be optimized through quantization or pruning techniques while maintaining optimal performance. The proposed network has the potential to be extended to identify additional coffee leaf diseases, such as coffee wilt, red spider mite, and sooty molds. However, it is important to note that the network was trained exclusively on a dataset obtained from laboratory settings. To address this constraint, it is recommended to incorporate a more diverse spectrum of data from various sources, including field-based data, as an avenue for future work. Finally, to facilitate practical implementation, the network can be integrated into a web or mobile application for real-time deployment.

Data availability

Data will be made available on request.

References

Abrham Debasu Mengistu, Seffi Gebeyehu Mengistu, Dagnachew Melesew, 2018. An Automatic Coffee Plant Diseases Identification Using Hybrid Approaches of Image Processing and Decision Tree. Indon. J. Elect. Eng. Comp. Sci. (IJEECS) 9 (3), 806–811.

Afifi, A., Alhumam, A., Abdelwahab, A., 2020. Convolutional neural network for automatic identification of plant diseases with limited data. Plants 10 (1), 28.

Andrade, C.C.L., de Resende, M.L.V., Moreira, S.I., Mathioni, S.M., Botelho, D.M.S., Costa, J.R., Andrade, A.C.M., Alves, E., 2021. Infection process and defense response of two distinct symptoms of Cercospora leaf spot in coffee leaves. In: Phytoparasitica, Vol. 49, Issue 4. Springer Science and Business Media LLC, pp. 727–737. https://doi.org/10.1007/s12600-021-00902-2.

Annual Review. International Coffee Organization, Year 2021-2022. https://www.ico.org/documents/cy2022-23/annual-review-2021-2022-e.pdf (Accessed 18 Feb 2023).

Barbedo, J.G.A., 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput. Electron. Agric. 153, 46–53.

Binney, Enoch, Ren, Dongxiao, 2022. Coffee leaf diseases classification and the effect of fine-tuning on deep convolutional neural networks. In: International Journal For Multidisciplinary Research (Vol. 4, Issue 5). International Journal for Multidisciplinary Research (IJFMR). https://doi.org/10.36948/ijfmr.2022.v04i05.861.

Bordin Yamashita, J.V.Y., Leite, J.P.R.R., 2023. Coffee disease classification at the edge using deep learning. In: Smart Agricultural Technology, vol. 4. Elsevier BV, p. 100183. https://doi.org/10.1016/j.atech.2023.100183.

Cao, Y., Xu, J., Lin, S., Wei, F., Hu, H., 2019. Gcnet: non-local networks meet squeeze-excitation networks and beyond. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 0-0).

Chowdhury, M., Burhan, U., 2021. Coffee leaf disease recognition using gist feature. Intern. J. Inform. Eng. Electron. Bus. 13 (2), 55–61.

Cohen, J., 1960. A coefficient of agreement for nominal scales. In: Educational and Psychological Measurement, Vol. 20, Issue 1. SAGE Publications, pp. 37–46. https://doi.org/10.1177/001316446002000104.

Dantas, J., Motta, I.O., Vidal, L.A., Nascimento, E.F.M.B., Bilio, J., Pupe, J.M., Veiga, A., Carvalho, C., Lopes, R.B., Rocha, T.L., Silva, L.P., Pujol-Luz, J.R., Albuquerque, É.V. S., 2021. A comprehensive review of the coffee leaf miner Leucoptera coffeella (Lepidoptera: Lyonetiidae)—A Major Pest for the coffee crop in Brazil and others Neotropical countries. In: Insects (Vol. 12, Issue 12, p. 1130). MDPI AG. https://doi.org/10.3390/insects12121130.

De Vita, F., Nocera, G., Bruneo, D., Tomaselli, V., Giacalone, D., Das, S.K., 2020, September. Quantitative analysis of deep leaf: A plant disease detector on the smart edge. In: 2020 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE, pp. 49–56.

Divyashri, P., Pinto, L.A., Mary, L., Manasa, P., Dass, S., 2021. The real-time mobile application for identification of diseases in coffee leaves using the CNN model. In: 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC). 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE. https://doi.org/10.1109/icesc51422.2021.9532662.

Esgario, J.G., de Castro, P.B., Tassis, L.M., Krohling, R.A., 2022. An app to assist farmers in the identification of diseases and pests of coffee leaves using deep learning. Inform. Proc. Agricult. 9 (1), 38–47.

Esgario, J.G., Krohling, R.A., Ventura, J.A., 2020. Deep learning for classification and severity estimation of coffee leaf biotic stress. Comput. Electron. Agric. 169, 105162.

Essoh, S.L.E., Kenfack, H.M.T., Ebele, B.A.M., Mbietieu, A.M., Etoua, O.V.E., 2022. Detection and classification of coffee plant diseases by image processing and machine learning. In: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer International Publishing, pp. 137–149. https://doi.org/10.1007/978-3-030-93314-2_9.

Faisal, M., Leu, J., Avian, C., Prakosa, S.W., Köppen, M., 2023a. DFNet: Dense fusion convolution neural network for plant leaf disease classification. In: Agronomy Journal. Wiley. https://doi.org/10.1002/agj2.21341.

Faisal, M., Leu, J.-S., Darmawan, J.T., 2023b. Model selection of hybrid feature fusion for coffee leaf disease classification. In: IEEE Access, vol. 11. Institute of Electrical and Electronics Engineers (IEEE), pp. 62281–62291. https://doi.org/10.1109/access.2023.3286935.

Fan, X., Luo, P., Mu, Y., Zhou, R., Tjahjadi, T., Ren, Y., 2022. Leaf image based plant disease identification using transfer learning and feature fusion. In: Computers and Electronics in Agriculture, vol. 196. Elsevier BV, p. 106892. https://doi.org/10.1016/j.compag.2022.106892.

Ganguly, S., Bhowal, P., Oliva, D., Sarkar, R., 2022. BLeafNet: A Bonferroni mean operator based fusion of CNN models for plant identification using leaf image classification. In: Ecological Informatics, vol. 69. Elsevier BV, p. 101585. https://doi.org/10.1016/j.ecoinf.2022.101585.

Hailu, B.T., Siljander, M., Maeda, E.E., Pellikka, P., 2017. Assessing spatial distribution of Coffea arabica L. in Ethiopia's highlands using species distribution models and geospatial analysis methods. In: Ecological Informatics, vol. 42. Elsevier BV, pp. 79–89. https://doi.org/10.1016/j.ecoinf.2017.10.001.

Hasan, Reem Ibrahim, et al., 2023. Automatic clustering and classification of coffee leaf diseases based on an extended kernel density estimation approach. Plants 12 (8), 1603.

Javierto, D.P.P., Martin, J.D.Z., Villaverde, J.F., 2021. Robusta Coffee Leaf Detection based on YOLOv3- MobileNetv2 model. In: 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM). IEEE. https://doi.org/10.1109/hnicem54116.2021.9731899.

Krishnan, S., 2017. Sustainable coffee production. In: Oxford Research Encyclopedia of Environmental Science.

Krohling, R.A., Esgario, J., Ventura, J.A., 2019. BRACOL–a Brazilian Arabica coffee leaf images dataset to identification and quantification of coffee diseases and pests. Mend. Data 1.

Kumar, M., Gupta, P., Madhav, P., 2020, June. Disease detection in coffee plants using convolutional neural network. In: In 2020 5th International Conference on Communication and Electronics Systems (ICCES). IEEE, pp. 755–760.

Laércio Zambolim and Eveline Teixeira Caixeta. "An overview of physiological specialization of coffee leaf rust – new designation of pathotypes", Intern. J. Curr. Res., 13, (01), 15564–15575.

Madhukar, R.K., Chaurasiya, A., Chaturvedi, P., 2022. A systematized chronicity based disease classification in coffee leaves using deep learning. In: 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). IEEE. https://doi.org/10.1109/icosec54921.2022.9951915.

Marcos, A.P., Silva Rodovalho, N.L., Backes, A.R., 2019. Coffee Leaf Rust Detection Using Convolutional Neural Network. In: 2019 XV Workshop de Visão Computacional (WVC). 2019 XV Workshop de Visão Computacional (WVC). IEEE. https://doi.org/10.1109/wvc.2019.8876931.

Marin, D.B., Ferraz, G.A. e S., Santana, L.S., Barbosa, B.D.S., Barata, R.A.P., Osco, L.P., Ramos, A.P.M., Guimarães, P.H.S., 2021. Computers and Electronics in Agriculture, vol. 190. Elsevier BV, p. 106476. https://doi.org/10.1016/j.compag.2021.106476.

Martinez, F., Montiel, H., Martinez, F., 2022. A machine learning model for the diagnosis of coffee diseases. Int. J. Adv. Comput. Sci. Appl. 13 (4).

Miranda, J. da R., Alves, M. de C., Pozza, E.A., Santos Neto, H., 2020. Detection of coffee berry necrosis by digital image processing of landsat 8 oli satellite imagery. In: International Journal of Applied Earth Observation and Geoinformation (Vol. 85, p. 101983). Elsevier BV. https://doi.org/10.1016/j.jag.2019.101983.

Montalbo, F.J.P., 2022. Automated diagnosis of diverse coffee leaf images through a stage-wise aggregated triple deep convolutional neural network. Mach. Vis. Appl. 33 (1), 19.

Montalbo, F.J.P., Hernandez, A.A., 2020 February. An optimized classification model for Coffea Liberica disease using deep convolutional neural networks. In: 2020 16th

IEEE International Colloquium on Signal Processing & its Applications (CSPA). IEEE, pp. 213–218.
Montalbo, F.J.P., Hernandez, A.A., 2020b. Classifying Barako coffee leaf diseases using deep convolutional models. Intern. J. Adv. Intell. Inform. 6 (2), 197–209.
Novtahaning, D., Shah, H.A., Kang, J.-M., 2022. Deep learning ensemble-based automated and high-performing recognition of coffee leaf disease. In: Agriculture (Vol. 12, Issue 11, p. 1909). MDPI AG. https://doi.org/10.3390/agriculture12111909.
Oliveira, A.J., Assis, G.A., Guizilini, V., Faria, E.R., Souza, J.R., 2019. Segmenting and detecting nematode in coffee crops using aerial images. In: Lecture Notes in Computer Science. Springer International Publishing, pp. 274–283. https://doi.org/10.1007/978-3-030-34995-0_25.
Pandey, A., Jain, K., 2022. A robust deep attention dense convolutional neural network for plant leaf disease identification and classification from smart phone captured real world images. In: Ecological Informatics, vol. 70. Elsevier BV, p. 101725. https://doi.org/10.1016/j.ecoinf.2022.101725.
Ramamurthy, K., Thekkath, R.D., Batra, S., Chattopadhyay, S., 2023. A novel deep learning architecture for disease classification in Arabica coffee plants. In: Concurrency and Computation: Practice and Experience. Wiley. https://doi.org/10.1002/cpe.7625.
Sharma, V., Tripathi, A.K., Mittal, H., 2023. DLMC-net: Deeper lightweight multi-class classification model for plant leaf disease detection. In: Ecological Informatics, vol. 75. Elsevier BV, p. 102025. https://doi.org/10.1016/j.ecoinf.2023.102025.
Shubhashini pal., 2021. Precision-agriculture: an image net – based multilayer convolution neural network for leaf disease detection in coffee Plant in Early-Stage System. J. Res. Proc. 1 (2), 351–363. Retrieved from. https://www.i-jrp.com/index.php/jrp/article/view/72. Retrieved from.
Silva, M. do C., Guerra-Guimarães, L., Diniz, I., Loureiro, A., Azinheira, H., Pereira, A.P., Tavares, S., Batista, D., Várzea, V., 2022. An overview of the mechanisms involved in coffee-Hemileia vastatrix interactions: Plant and pathogen perspectives. In: Agronomy, vol. 12, Issue 2. MDPI AG, p. 326. https://doi.org/10.3390/agronomy12020326.
Singh, R., Krishnan, P., Bharadwaj, C., Das, B., 2023. Improving prediction of chickpea wilt severity using machine learning coupled with model combination techniques under field conditions. In: Ecological Informatics, vol. 73. Elsevier BV, p. 101933. https://doi.org/10.1016/j.ecoinf.2022.101933.
Sorte, L.X.B., Ferraz, C.T., Fambrini, F., dos Reis Goulart, R., Saito, J.H., 2019. Coffee leaf disease recognition based on deep learning and texture attributes. Proc. Comp. Sci. 159, 135–144.
Suparyanto, T., Firmansyah, E., Cenggoro, T.W., Sudigyo, D., Pardamean, B., 2022, February. Detecting Hemileia vastatrix using vision AI as supporting to food security for smallholder coffee commodities. In: IOP Conference Series: Earth and Environmental Science, vol. 998, No. 1. IOP Publishing, p. 012044.
Syahputra, R., Triayudi, A., Sholihati, I.D., 2020. Application of expert system to diagnose pests and diseases in coffee plant using web-based Naïve Bayes: application of expert system to diagnose pests and diseases in coffee plant using web-based Naïve Bayes. J. Mant. 3 (4), 383–392. Available at: http://iocscience.org/ejournal/index.php/mantik/article/view/57 (Accessed 19 Feb 2023).
Szegedy, C., Liu, Wei, Jia, Yangqing, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. https://doi.org/10.1109/cvpr.2015.7298594.
Tassis, L.M., de Souza, J.E.T., Krohling, R.A., 2021. 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, 106191.
Tassis, L.M., Krohling, R.A., 2022. Few-shot learning for biotic stress classification of coffee leaves. Artif. Intell. Agricult. 6, 55–67.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Proces. Syst. 30.
Ventura, José, Costa, Hélcio, Lima, Inorbert, 2019. Chapter 18 Conilon Coffee Diseases Management 2019.
Volsi, B., Telles, T.S., Caldarelli, C.E., da Camara, M.R.G., 2019. The dynamics of coffee production in Brazil. In: Aldrich, S.P. (Ed.), PLOS ONE (Vol. 14, Issue 7, p. e0219742). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pone.0219742.
Waldamichael, F.G., Debelee, T.G., Ayano, Y.M., 2021. Coffee disease detection using a robust HSV color-based segmentation and transfer learning for use on smartphones. In: International Journal of Intelligent Systems (Vol. 37, Issue 8, pp. 4967–4993). Hindawi Limited. https://doi.org/10.1002/int.22747.
Wallelign, S., 2020. An Intelligent System for Coffee Grading and Disease Identification (Doctoral dissertation). École Nationale d'Ingénieurs de Brest.
Yebasse, M., Shimelis, B., Warku, H., Ko, J., Cheoi, K.J., 2021. Coffee disease visualization and classification. Plants 10 (6), 1257.