

Color Difference Histogram Capsule Network (CDH-CapsNet) for Plant Disease Recognition

Steve Okyere-Gyamfi*

Kwame Nkrumah University of Science and Technology, University Post Office (PMB), Kumasi, Ghana

Catholic University of Ghana, P. O. Box 363, Sunyani, Ghana

E-mail: steve.og@cug.edu.gh

ORCID iD: <https://orcid.org/0000-0002-7295-0465>

*Corresponding author

Michael Asante

Kwame Nkrumah University of Science and Technology, University Post Office (PMB), Kumasi, Ghana

E-mail: masante.sci@knust.edu.gh

ORCID iD: <https://orcid.org/0000-0002-4623-6903>

Kwame Ofosuhene Peasah

Kwame Nkrumah University of Science and Technology, University Post Office (PMB), Kumasi, Ghana

E-mail: kopeasah.cos@knust.edu.gh

ORCID iD: <https://orcid.org/0000-0003-3863-4642>

Yaw Marfo Missah

Kwame Nkrumah University of Science and Technology, University Post Office (PMB), Kumasi, Ghana

E-mail: ymissah.cos@knust.edu.gh

ORCID iD: <https://orcid.org/0000-0002-2926-681X>

Vivian Akoto-Adjepong

University of Energy and Natural Resources P. O. Box 214, Sunyani, Ghana

E-mail: vivian.akoto-adjepoing@uenr.edu.gh

ORCID iD: <https://orcid.org/0000-0003-4608-0758>

Received: 04 January 2026; Revised: 22 March 2026; Accepted: 25 April 2026; Published: 08 June 2026

Abstract: Plant diseases adversely affect the quantity and quality of food production, contributing to food insecurity. Prompt identification, diagnosis, and intervention can significantly minimize economic and ecological losses. By reducing the use of agrochemicals through timely disease detection, the environmental impact can be mitigated. Traditional manual methods for recognizing plant diseases are prevalent but are often limited, time-consuming, costly, and ineffective. Convolutional Neural Network (CNN) architectures have demonstrated excellent capabilities in detecting plant diseases and other complex images, but they lack spatial or rotational invariance and require extensive data in various forms to be effective. This is typically achieved by applying data augmentation, as the datasets in the field of agriculture are often limited. Capsule Networks address CNN's limitations, but their encoder network is inefficient at feature extraction, hence it does not perform well on complex images. This study seeks to modify and improve CapsNet by combining a Color Difference Histogram (CDH) with a Capsule Network that includes two extra convolutional, three max pooling layers, three batch normalization layers, and reduces the primary capsule channels in the original CapsNet to 16 from 32 for efficient plant disease detection in apples, bananas, grapes, corn, mangoes, pepper, potatoes, rice, tomato, and on the CIFAR-10 dataset. This approach improved the original CapsNet in terms of validation accuracies by 5.83%, 14.82%, 5.9%, 4.42%, 20.87%, 40.12%, 4.41%, 0.76%, 9.49%, and 13.97% on apple, banana, grape, corn, mango, pepper, potato, rice, tomato, and CIFAR-10 datasets respectively. The CDH-CapsNet achieved better results in terms of accuracy, sensitivity, F1-Score, precision, specificity, Receiver Operating Characteristic (ROC), Precision-Recall (PR) values, parameter count, and disk size, surpassing the original CapsNet and CapsNet models presented in available research. The original CapsNet and CDH-CapsNet exhibited strong performance on datasets such as the Rice dataset, possibly because of high-quality images and low intra-class variance. The findings suggest that this efficient and computationally less demanding supportive tool can significantly enhance plant disease classification by offering a lightweight, scalable solution that can be adapted for field use in resource-constrained settings, contributing to efforts aligned with the SDG 2 goal. However, environmental factors such as inconsistent lighting and complex backgrounds encountered in practical

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

scenarios may affect the model's effectiveness. Subsequent studies will aim to overcome these issues and broaden the model's applicability.

Index Terms: Color Difference Histogram (CDH), Convolutional Neural Network (CNN), Capsule Neural Network, Dynamic Routing, Plant Disease Detection.

1. Introduction

Ensuring food safety faces significant challenges due to crop protection issues, with crop diseases being a primary concern. These diseases reduce crop yield quality, leading to economic losses and prompting the use of agrochemicals, which in turn adversely affect the environment. Ensuring plant health is crucial for achieving the goals of ending hunger, reducing poverty, and promoting economic and ecological development, as food security is a key objective in the 2030 Sustainable Development Goals (SDGs). Evaluating plant diseases manually is time-consuming and prone to mistakes, so automated plant disease detection is necessary for accurate diagnosis.

In various research projects, deep learning and machine learning methods have been employed for classifying diseases in plants. In agriculture, CNN models are predominantly employed for the diagnosis of plant diseases [1-10], etc. Nevertheless, they have notable limitations, such as disregarding the spatial arrangement of image components and being unstable to affine transformations like translation, rotation, and scaling. These shortcomings necessitate an extensive dataset encompassing all possible object orientations in images for effective training. Consequently, this results in higher computational costs in terms of resources and training time, as well as the time-consuming nature of implementing data augmentation techniques. The CNN pooling operation is recognized for losing spatial structure information, leading to CNNs becoming invariant instead of equivariant as they delve deeper.

To address this issue, Capsule Networks with dynamic routing between capsules were introduced. Capsule Networks effectively capture spatial information and distinguish between different poses, orientations, and textures, making them robust to uneven class distributions and reducing the need for extensive data. However, despite these benefits, Capsule Networks also have their drawbacks. They exhibit weak performance when handling complex images or those with varied backgrounds, and struggle with learning lower-level descriptions [11].

To overcome the limitations of Capsule Networks, some researchers expand their models by increasing depth and width, which often leads to longer training durations and a higher number of parameters. However, deeper architectures do not always guarantee improved performance. Therefore, selecting effective feature extraction methods for the encoder is crucial to enhance the overall effectiveness of the CapsNet model.

This research aims to introduce a shallow Capsule Network with a low parameter count, designed for effective plant disease classification by integrating Color Difference Histogram (CDH) features along a few layers, which can help improve the performance of CapsNet on complex images.

Experimental evaluations conducted on ten widely available datasets reveal that the proposed CDH-CapsNet model achieves performance comparable to leading state-of-the-art models in identifying plant diseases and handling other complex images. Moreover, visualizations of class capsule clustering, activation maps, and image reconstructions highlight the enhanced feature extraction capability of CDH-CapsNet over the original CapsNet and other CapsNet variants.

This study offers the following contributions;

- Suggest an efficient, low-parameter CDH-CapsNet, designed for effective feature extraction, to assist CapsNet in learning lower-level details of complex images.
- The proposed CDH-CapsNet model produces a minimal number of parameters, which reduces computational and memory requirements, enhancing its suitability for implementation on embedded devices or resource-constrained devices.
- Evaluate the performance of the CDH-CapsNet model with other variants of CapsNet
- The research enhanced explainable AI by offering a comprehensive visual explanation of layer outputs.

However, even though the various CapsNet models found in the literature, used for comparative analysis, utilized the same datasets as in this study, environmental factors such as inconsistent lighting and complex backgrounds encountered in practical scenarios may influence the model's effectiveness. Hence, subsequent studies will aim to overcome these issues and broaden the model's applicability by concentrating on testing the model in more varied and challenging environments, expanding the diversity of datasets.

The remainder of this study is organized as follows: Section 2 presents related work on Capsule Networks for plant disease detection and the CIFAR-10 dataset. Section 3 describes the proposed model architecture, including the integration of Color Difference Histograms and Capsule Network layers, as well as the experimental setup, datasets, and evaluation metrics. Section 4 discusses the results and performance analysis across all datasets used in the study. Finally, Section 5 concludes the study with key findings and future work.

2. Related Works

The constraints of human-based diagnosis prompted the integration of algorithms to forecast plant diseases. Here are some CapsNet models found in the literature;

Verma and friends developed an improved feature computation method using Squeeze and Excitation (SE) Networks, which are applied before the original Capsule networks (CapsNet) in the classification process to evaluate plant disease intensity. They integrated two SE networks, ResNet with CapsNet and AlexNet with CapsNet. Using leaf images from the PlantVillage dataset, they specifically focused on Late Blight disease in tomatoes, categorizing the images into late, middle, early, and healthy stages. The SE-Alex-CapsNet model achieved a maximum accuracy of 92.1% with 9 million parameters, while the SE-Res-CapsNet model reached a higher accuracy of 93.75% with 19 million parameters [12].

Mensah and his team suggested using Gabor filters and CapsNet to detect distorted and blurred images of citrus and tomato diseases using the dataset from Plant Village. According to their experimental results, the proposed model achieved a validation accuracy of 98.13% for tomato diseases and 93.33% for citrus diseases, with the tomato model comprising 12 million parameters [13].

Vasudevan and Karthick suggested a hybrid approach for identifying grape diseases by augmenting and segmenting images. Initially, from images of plants, they used a graph-based method to derive the leaf area. Subsequently, an E-GAN was employed to augment the size of the dataset. Finally, disease identification and staging were carried out using CapsNet. Despite the model's complex features leading to longer inference times, the evaluation utilized PlantVillage dataset grape leaf images, achieving a validation accuracy of 97.63% [14].

Mensah and collaborators again introduced substituting the SoftMax, dynamic routing, and CNN employed in the conventional CapsNet with a sigmoid function, k-means routing, and Local Binary Pattern, respectively. Their model attained a validation accuracy of 92.72% on fashion-MNIST, 99.68% on MNIST, 75.80% on CIFAR-10, 98.06% on tomato, 99.41% on citrus, and 96.79% on maize. It utilized 2.8 million parameters for MNIST and fashion MNIST, 5.2 million for CIFAR-10, and 8.4 million for tomato, citrus, and maize datasets [15].

Xu and co-authors developed a dilated convolution CapsNet aimed at detecting diseases in apple leaves. They extended the CapsNet framework by integrating two modules of dilated Inception using various dilation rates, thereby improving the receptive fields of the convolution to obtain multi-level features from input images. Their model attained a validation accuracy of 93.16% in detecting diseases in apples [16].

Also, Verma and his team used CapsNet to train PlantVillage's colored potato leaf image dataset. The CapsNet model matched the results of cutting-edge models, generating 9,856,768 parameters and observing 91.83% accuracy [17].

Mensah and Ayidzoe again suggested a CapsNet technique that uses a dual convolution input channel and sends its results to the layers of the conventional CapsNet. On the tomato and CIFAR-10 datasets, the model attained 93.03% and 76.58% validation accuracies with 6.04 million and 5.48 million parameters, respectively [18].

Oladejo and Ademola also developed an improved CapsNet architecture specifically for diagnosing diseases in banana leaves. They optimized the original CapsNet by reducing the fully connected layer to just two neurons (512 and 1024) to speed up training. They also integrated a momentum optimizer to enhance network processing speed. Their model successfully identified black sigatoka, healthy, and banana bacterial wilt leaves, achieving a remarkable test accuracy of 95% [19].

Once more, Mensah and co-authors used Gabor instead of CNN to capture texture and spatial arrangements, and max pooling immediately after Gabor to help extract only important characteristics and reduce the dimensions of vectors. On the tomato dataset from the PlantVillage dataset, the model generated 8,708,128 parameters and attained 97.98% validation accuracy [20].

Abouelmagd and the team also suggested a CapsNet technique by applying data augmentation and pre-processing to minimize overfitting. On the tomato dataset, which comprises ten classes, the model achieved 96.39% accuracy [21].

Zang and co-authors proposed a technique to detect diseases by replacing the convolution kernel of the residual network (ResNet) with a 3x3 kernel. They introduced an attention mechanism to encode pertinent characteristics. The optimized ResNet was then incorporated into CapsNet. The merged model SE-SK-CapResNet attained 95.08%, 98.58%, and 97.19% validation accuracies on AI Challenger 2018, PlantVillage, and Tomato Leaf Disease datasets, respectively [22].

Altan developed a CapsNet framework to examine different feature learning approaches and enhance the learning abilities of deep learning methods. They modified the original CapsNet architecture by adjusting the nodes within the fully connected layers (FC) to FC1-960, FC2-768, and FC3-4096. This adjustment yielded remarkable classification performance, with specificity, accuracy, and sensitivity rates of 97.49%, 95.76%, and 96.37%, respectively, for bell pepper plant disease classification [23].

All these CapsNet models demonstrated strong performance across different datasets. However, when it comes to detecting plant diseases using complex images, models must be both efficient and low in parameters so that they can be usable on resource-constrained devices. However, most of these models required extensive preprocessing before their results were achieved, which is not feasible in real-world scenarios, and generated a large number of parameters, making them computationally expensive.

3. Proposed Approach

This study aims to introduce a streamlined CapsNet, designed with fewer parameters, that strives to achieve superior performance while enhancing feature extraction capabilities.

3.1. Original/Traditional CapsNet Model

Figure 1 illustrates the Original/traditional CapsNet model [11], which is used as the baseline model.

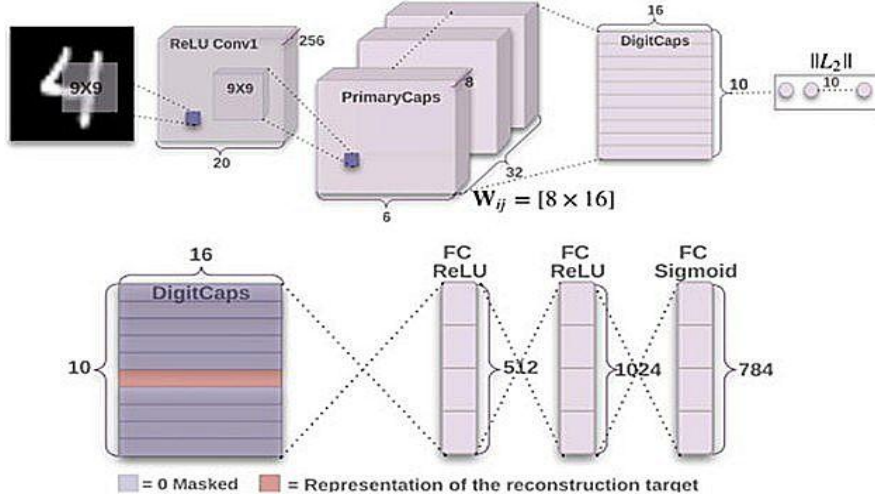


Fig. 1. Original CapsNet architecture [11]

3.2. Color Difference Histogram (CDH) Feature Map Detection

Here, edge orientation and color quantization are identified. The subsequent step involves detecting microstructures using both edge and color quantization, employing a fraction of one pixel and a 3x3 filter for CDH [24].

This process entails comparing the centre value of the filter with its eight neighboring values, resulting in the generation of color and edge maps. These maps are then utilized to calculate the disparity Δ in edge orientation and color intensity for each $L*a*b^*$ element, which are subsequently stored as edge and color features. The last step involves displaying these characteristics as a unified histogram.

In a quantized image $C(x, y)$, where the values are represented as $w \in 0, 1, \dots, W - 1$, locations of neighboring pixels (x, y) and $(x' y')$ with corresponding values for the color indexes $C(x, y) = w_1$ and $(x' y') = w_2$ are denoted. Similarly, in an edge orientation image $\theta(x, y)$, where the values are represented as $v \in 0, 1, \dots, V - 1$, the angles at (x, y) and $(x' y')$ are represented as $\theta(x, y) = v_1$ and $\theta(x' y')$ respectively. For adjacent pixels that are separated by a distance D , and have respective quantization values for edge and color orientations, denoted as V and W . The color difference histogram is specified by equations (1) and (2).

$$H_{color}(C(x, y)) = \begin{cases} \sum \sqrt{(\Delta L^2) + (\Delta a^2) + (\Delta b^2)} \\ \text{where } \theta(x, y) = \theta(x' y'); \max(|x - x'|), (|y - y'| = D) \end{cases} \quad (1)$$

$$H_{ori}(\theta(x, y)) = \begin{cases} \sum \sqrt{(\Delta L^2) + (\Delta a^2) + (\Delta b^2)} \\ \text{where } C(x, y) = C(x' y'); \max(|x - x'|), (|y - y'| = D) \end{cases} \quad (2)$$

Where ΔL , Δa , and Δb represent the variations in values in two color pixels, D is set to 1. If edge orientation is denoted by W and the color quantization by V , then the features of CDH are described in equation (3) as follows:

$$H_{CDH} = H_{color}(0), H_{color}(1) \dots H_{color}(W - 1), H_{ori}(0), H_{ori}(1) \dots H_{ori}(V - 1) \quad (3)$$

Where the color histogram is represented by H_{color} , while the edge orientation histogram is denoted as H_{ori} . For example, if we assign 72 dimensions to color quantization and 18 dimensions to edge orientation, the combined features for image retrieval would amount to 90 dimensions. This collective feature set is symbolized as H .

The color difference histogram maps the differences in color intensity and contrast across adjacent pixels, highlighting texture and color patterns that are crucial for image classification using deep learning.

3.3. Proposed CDH-CapsNet Model

Figure 2 illustrates the proposed CDH-CapsNet model, which comprises a CDH layer [24], convolutional layers, batch normalization layers, max-pooling layers, a primary capsule, a class capsule, and a fully connected layer. It is a modification of the original Capsule Network by Sabour and colleagues [11]. We incorporated additional layers such as the Color Difference Histogram (CDH) layer, two convolutional layers, and three max pooling layers, all to improve feature extraction from input images. Additionally, three batch normalization layers were added to ensure consistent data distribution, simplify learning, and accelerate training. We also reduced the primary capsule channels from 32 of the original CapsNet to 16, which also lowered the parameter count and computational cost, mitigated overfitting, sped up training, and improved suitability for small datasets and resource-constrained devices. The input image size was reduced to $32 \times 32 \times 3$, which helped lower the parameter count and computational cost.

Input images are resized to $32 \times 32 \times 3$ and passed through the CDH layer, which is integrated as a preprocessing feature extraction layer in the CDH-CapsNet architecture. Specifically, CDH captures critical low-level visual cues related to color variation and edge or texture orientation, which are particularly relevant in identifying plant disease symptoms such as discoloration, spotting, and lesions. The CDH layer efficiently extracts key features from the images without increasing computational overhead. The features obtained from the CDH stage are given to a convolutional layer (Conv1) consisting of 256 filters with a 3×3 kernel size, a stride of 1, and a ReLU activation function. This operation results in a $30 \times 30 \times 256$ feature map. The feature map obtained is passed through a batch normalization layer, followed by a max pooling layer (MP1), resulting in a $15 \times 15 \times 256$ feature map. The resulting feature map is passed through a second convolutional layer (Conv2), which consists of 128 filters with a 3×3 kernel, a stride of 1, and a ReLU activation function. This operation generates a $13 \times 13 \times 128$ feature map, which is subsequently sent through a batch normalizer, followed by a max pooling feature extraction stage (MP2), reducing the dimensions to $7 \times 7 \times 128$. The resulting feature map is further processed by a third convolutional layer (Conv3), which applies 32 filters of size 3×3 , a stride of 1, and utilizes the ReLU squashing function. This results in a feature map of dimensions $5 \times 5 \times 32$. This output is then passed through a batch normalizer, followed by a max pooling feature extraction stage (MP3), reducing it to a $3 \times 3 \times 32$ feature map. The features extracted are subsequently fed into the primary capsule, which is configured with 8 dimensions and 16 channels, reduced from 32 in the original CapsNet. This capsule uses a 3×3 kernel and a stride of 2. The resulting output from the primary capsules is then directed to the Plant_DiseaseCaps (class capsule) for classifying plant diseases. The Plant_DiseaseCaps module consists of 16-dimensional capsules and also considers the total number of classes. Its output is passed to a reconstruction layer, which attempts to reconstruct the features captured by the Plant_DiseaseCaps. These features are then forwarded to the decoder component in the capsule network. This decoder then interprets the entity's attributes. The decoder itself comprises three fully connected layers containing 512, 1024, and 3072 neurons, respectively.

Color distortions, lesion spots, and chlorosis, which are common symptoms in plant diseases, are better captured by CDH than by traditional convolutional filters alone. This hybrid approach ensures that CapsNet receives not only shape and texture information but also robust color difference cues, improving its ability to generalize across visually similar disease classes. The CDH, convolution, and max pooling layers all assisted in capturing key features from the input images to enhance feature extraction and classification performance of the proposed CDH-CapsNet. Figure 2 uses apple leaf diseases as an example for illustration.

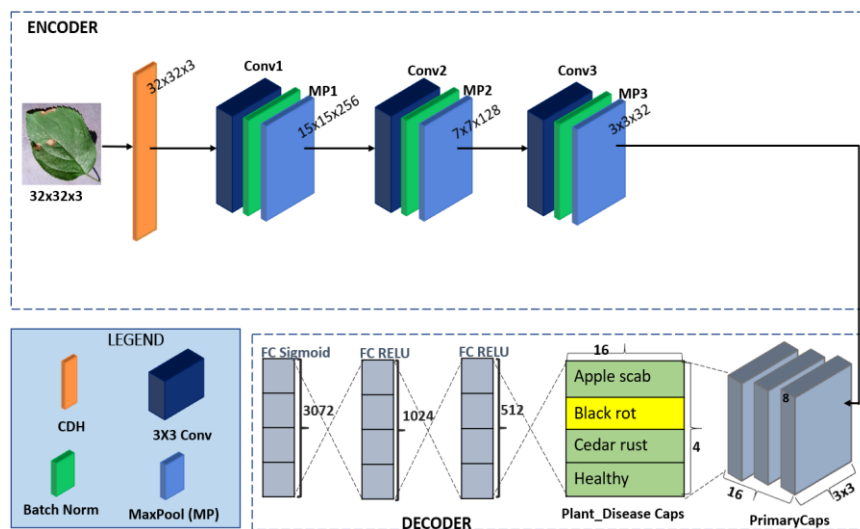


Fig. 2. Architecture of proposed CDH-CapsNet model

3.4. Data Description

The apple, grape, corn, pepper, potato, and tomato datasets, scaled at 256×256 , are a portion of the Plant Village [25]. <https://github.com/spMohanty/PlantVillage-Dataset>

Apple: It contains 3171 images grouped into four classes: 0: Apple_scab (630 images), 1: Black_rot,(621 images) 2: Cedar_apple_rust (275 images), and 3: healthy (1645 images).

Grape: It contains 4062 images grouped into four classes; 0: Black_rot(1180 images), 1: Esca_(Black_Measles) (1383 images), 2: Leaf_blight_(Isariopsis_Leaf_Spot) (1076 images), and 3: healthy (423 images).

Corn: It contains 3852 images grouped into four classes: 0: Cercospora_leaf_spot, Gray_leaf_spot (513 images), 1: Common_rust (1192 images), 2: Northern_Leaf_Blight (985 images), and 3: healthy (1162 images).

Pepper: It contains 2475 images grouped into two classes: 0: healthy (1478 images), and 1: Bacterial_spot (997 images).

Potato: It contains 2152 images grouped into three classes: 0: Healthy (152 images), 1: Early_blight (1000 images), and 2: Late_blight (1000 images).

Tomato: It contains 18160 images grouped into ten classes; 0: Bacterial_spot (2127 images), 1: Early_blight (1000 images), 2: Late_blight (1909 images), 3: Leaf_Mold (952 images), 4: Septoria_leaf_spot (1771 images), 5: Spider_mites Two-spotted_spider_mite (1676 images), 6: Target_Spot (1404 images), 7: Tomato_mosaic_virus (373 images), 8: Tomato_Yellow_Leaf_Curl_Virus (5357 images), and 9: Healthy (1591 images).

Banana: It contains 937 images grouped into four classes: 0:Cordana (162 images), 1:Pestalotiopsis (173 images), 2: Sigatoka (473 images), and 3: Healthy (129 images) [26]. <https://data.mendeley.com/datasets/9tb7k297ff/1>

Mango: It contains 4000 images grouped into eight classes: 0: Anthracnose (500 images), 1: Bacterial Canker (500 images), 2: Cutting Weevil (500 images), 3: Die Back (500 images), 4: Gall Midge (500 images), 5: Powdery Mildew (500 images), 6: Powdery Mildew (500 images), and 7: Healthy (500 images) [27]. <https://data.mendeley.com/datasets/hxsnvwt3r/1>

Rice: It contains 5932 images grouped into four classes: 0: Bacterial_blight (1584 images), 1: Blast (1440 images), 2: Brown_spot (1600 images), and 3: Tungro (1308 images) [28]. <https://data.mendeley.com/datasets/fwcj7stb8r/1>

CIFAR 10: consists of 60,000 32x32x3 images. Of these, 10,000 are for testing and 50,000 are for training. They have different backgrounds and are more complex than Fashion-MNIST [29]. <https://www.cs.toronto.edu/~kriz/cifar.html>

To streamline the varied dimensions of the plant datasets, the images were all resized to $32 \times 32 \times 3$ dimensions and then divided into training and testing sets using an 80:20 ratio. No data augmentation technique was applied to the dataset, as CapsNet can handle smaller datasets, and the CDH-CapsNet was designed to extract highly important features that aid in better classification. Most of the 10 datasets are imbalanced; hence, if the CDH-CapsNet performs well on them, it will reflect the model's robustness in handling real-world imbalanced datasets.

3.5. Experimental Setup

The experiments conducted in this study were performed on a Windows 64-bit system using a Super GPU, with an 8GB VRAM NVIDIA GeForce RTX 2080, and 32 GB of RAM. The model was implemented using TensorFlow as the backend, with coding done in Python via the Anaconda distribution and the Keras library. Training was carried out over 200 epochs using the original image dataset, with a learning rate of 0.001 and a batch size of 100. To ensure effective training, the model that delivered the highest performance during the process was saved. The loss was evaluated using margin loss, denoted as L_k [11], with specific details outlined below;

$$L_k = T_k \max(0, m^+ - ||v_k||)^2 + \lambda(1 - T_k) \max(0, ||v_k|| - m^-)^2 \quad (4)$$

Where T_k equals 1 if class k is engaged, and otherwise 0. The hyperparameters λ , m^- , and m^+ are determined during the learning phase.

All datasets were resized to 32X32X3. This experimental setup was used to run the original CapsNet, which served as the baseline model, and the proposed CDH-CapsNet model.

3.6. Performance Evaluation Measures

This study utilized the following metrics for classification purposes

- **Validation Accuracy:** Assesses the rate of correct classifications among all instances. The stated aggregate validation accuracy summarizes the findings from all experimental runs.
- **Loss:** Quantifies the discrepancy between predicted values and true labels. This evaluation uses margin loss when testing.
- **Confusion Matrix:** Helps thoroughly analyze the count of correctly and incorrectly classified images. It helps to

- assess the sensitivity (recall), accuracy, precision, specificity, and other metrics.
- *Precision (P)*: The percentage of positive predictions that are accurate
- *Recall (R) or Sensitivity*: the fraction of positive cases that the model correctly recognizes compared to all actual positives in the dataset.
- *Specificity*: the true negative rate calculated over all actual negatives in the dataset.
- *F1-Score*: An average that harmoniously blends precision and recall.
- *Area under the curve (AUC)*: To evaluate the model's effectiveness on datasets with uneven class distributions, Receiver Operating Characteristic (ROC) and precision-recall (PR) curves are generated [30,31]. A higher AUC signifies better model performance.
- *Clustering*: We employ t-distributed stochastic neighbour embedding to obtain and analyze the clusters formed in the class capsule layer of the model.
- *Model complexity*: evaluates the model by considering the number of parameters and the disk space it occupies.

4. Results and Discussion

This section presents the results and discussion of the proposed CDH-CapsNet model on various plant disease image datasets as well as other complex datasets, such as CIFAR-10. The performance of the CDH-CapsNet is compared with the original CapsNet model [11], using the same datasets and experimental setup. Additionally, comparisons are made between the CDH-CapsNet, the original CapsNet, and other state-of-the-art CapsNet models reported in the literature.

We employed evaluation metrics such as sensitivity (SEN), precision (PRE), F1-Score (FS), specificity (SPE), size on disk, parameter count, AUC (ROC and PR) curves, accuracy (ACC), and loss. An ablation study was also conducted to assess the model's flexibility and robustness, as well as to determine the contribution of each component to its performance. Again, prediction, reconstruction, and visualization of clusters and activation maps from various layers were performed to help elucidate the inner workings of the CapsNet model, thereby opening the black box and enabling users to have confidence in it, while also contributing to the field of XAI [32-34].

4.1. Performance Evaluation

Figure 3 displays the loss and accuracy graphs for four datasets (apple, potato, rice, and CIFAR-10) for the traditional (original) and CDH-CapsNet models. Looking at the Rice dataset's loss and accuracy graphs, we observe that the original CapsNet model experienced notable fluctuations, as opposed to the CDH-CapsNet system, which showed early convergence and was steady to the end. Also, it can be observed that the CDH-CapsNet attained higher validation accuracies of 98.74%, 95.77%, 99.51%, 97.01%, 99.25%, 99.80%, 99.77%, 100%, 98.08%, and 77.55% with early convergence on apple, banana, grape, corn, mango, pepper, potato, rice, tomato, and CIFAR-10 datasets respectively as compared to 92.91%, 80.95%, 93.49%, 92.59%, 78.38%, 59.68%, 95.36%, 99.24%, 88.59%, and 63.58% attained by the traditional or original CapsNet model on the same datasets. This demonstrates the superior generalization ability of the CDH-CapsNet on unseen data compared to the original CapsNet.

Figure 4 displays the confusion matrices for the original and CDH-CapsNet models on the apple and banana datasets, depicting dataset images with correct and incorrect predictions. The blue-highlighted diagonal cells indicate the model's accurate predictions. Misclassifications, displayed in white, appear below and above these correct predictions. The rates of true positives, true negatives, false positives, and false negatives, along with per-class accuracy, sensitivity, specificity, F1-score, and precision, were determined using these values. It can be observed from these various confusion matrices that the CDH-CapsNet exhibits minimal misclassification compared to the traditional CapsNet model for all classes across the different datasets.

From Table 1, considering the accuracy per class for the classes 0, 1, 2, and 3 in the Apple dataset using the traditional or original CapsNet model, 4.25%, 3.62%, 1.42%, and 4.88% of images from these classes were misclassified, unlike with the CDH-CapsNet model that reduced the misclassification to only 1.1%, 0.47%, 0.16%, and 0.79% of the images that belong to these classes respectively. Also, from Table 2 considering the per class accuracy for the classes 0, 1, 2, and 3 in the Banana dataset using the traditional or original CapsNet model, 4.76%, 14.81%, 9.52%, and 8.99 % of images from these classes were misclassified, unlike with the CDH-CapsNet model that misclassified only 0.53%, 3.7%, 1.06%, and 3.17% of the images that belong to these classes respectively. This demonstrates that the CDH-CapsNet model outperforms the original or traditional model in classifying images across different categories for the various datasets.

For precision using the Apple dataset, the original CapsNet misclassified 6.96% of images in class 0, 10.16% in class 1, 8.93% in class 2, and 5.65% in class 3. In contrast, the CDH-CapsNet model reduced these misclassifications to 1.63%, 1.59%, 0%, and 1.20% for the same classes. Also, when the Banana dataset was used, the original CapsNet misclassified 12.50% of images in class 0, 41.46% in class 1, 7.69% in class 2, and 32% in class 3. In contrast, the CDH-CapsNet reduced these misclassifications to 2.9%, 0%, 1.05%, and 18.75% for the same classes. This indicates that the CDH-CapsNet model achieved greater precision, with fewer images being incorrectly assigned to the wrong classes in comparison to the original CapsNet model.

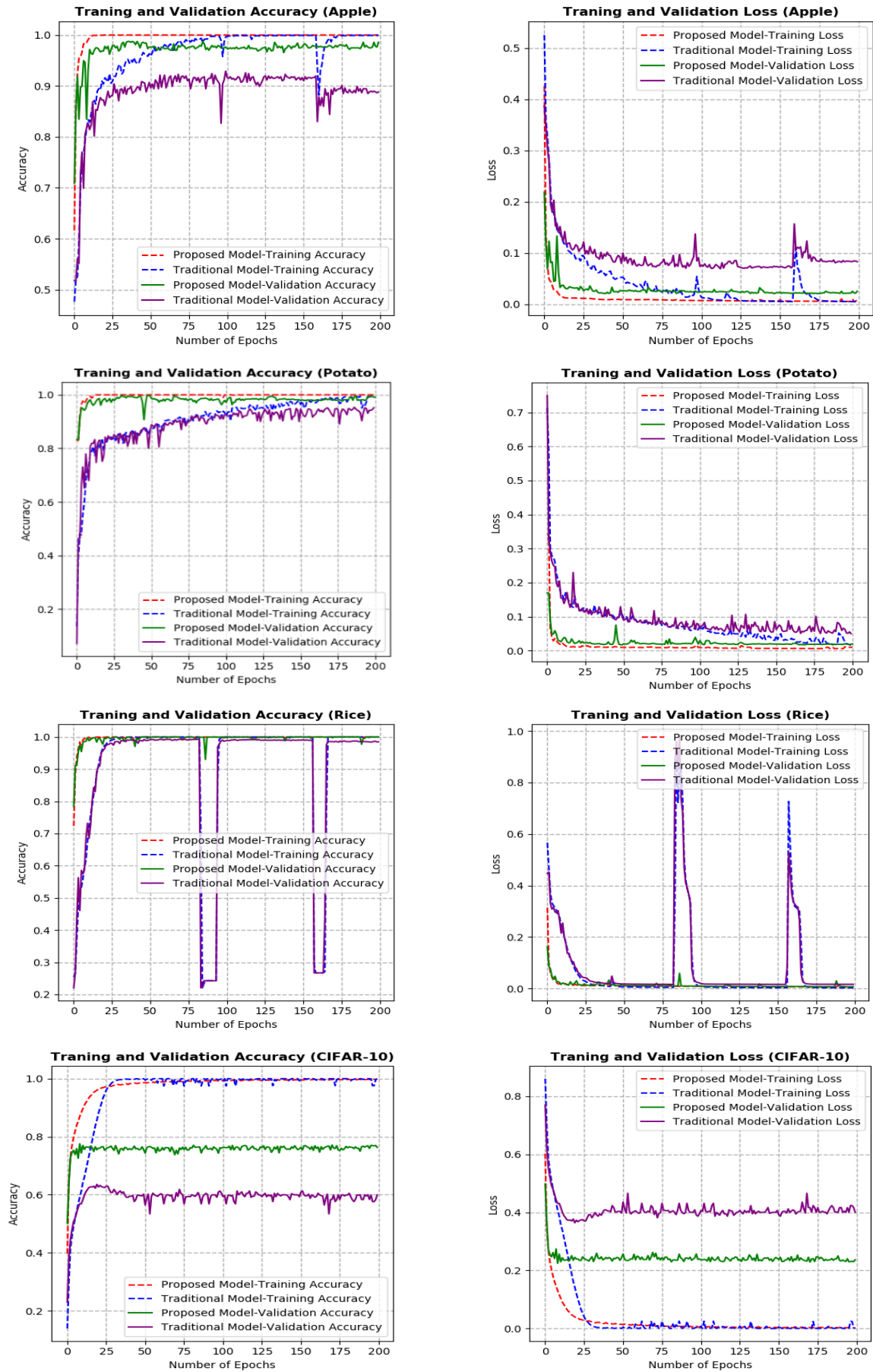


Fig. 3. Accuracy and loss graphs for the apple, potato, rice, and CIFAR-10 datasets for the proposed CDH-CapsNet and original CapsNet models

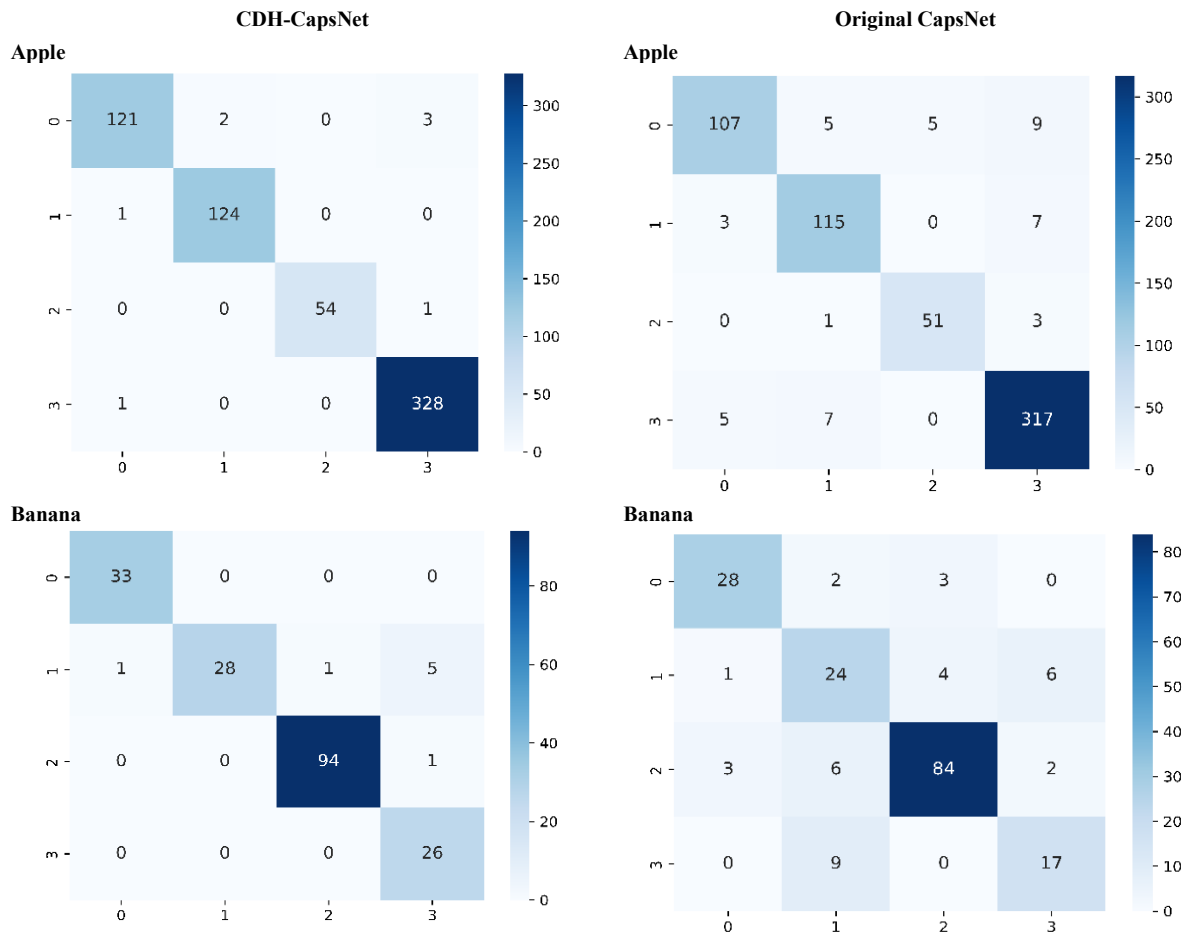


Fig. 4. Confusion matrices for CDH-CapsNet and the Original CapsNet model using the Apple and Banana datasets

Table 1. Performance metrics for CDH-CapsNet and original CapsNet using the apple dataset

Model (Dataset)	Class	TP	FP	FN	TN	ACC (%)	PRE (%)	SEN (%)	SPE (%)	FS (%)	Data Size
Original/ Traditional CapsNet (Apple)	0	107	8	19	501	95.75	93.04	84.92	98.62	88.80	126
	1	115	13	10	497	96.38	89.84	92.00	97.45	90.91	125
	2	51	5	4	575	98.58	91.07	92.73	99.14	91.89	55
	3	317	19	12	287	95.12	94.35	96.35	93.79	95.34	329
CDH-CapsNet (Apple)	0	121	2	5	507	98.90	98.37	96.03	99.61	97.19	126
	1	124	2	1	508	99.53	98.41	99.20	99.61	98.80	125
	2	54	0	1	580	99.84	100	98.18	100	99.08	55
	3	328	4	1	302	99.21	98.80	99.70	98.69	99.25	329

Table 2. Performance metrics for CDH-CapsNet and original CapsNet using the banana dataset

Model (Dataset)	Class	TP	FP	FN	TN	ACC (%)	PRE (%)	SEN (%)	SPE (%)	FS (%)	Data Size
Original/ Traditional CapsNet (Banana)	0	28	4	5	152	95.24	87.50	84.45	97.44	85.95	33
	1	24	17	11	137	85.19	58.54	68.86	88.96	63.28	35
	2	84	7	11	87	90.48	92.31	88.42	92.55	90.32	95
	3	17	8	9	155	91.01	68.00	65.39	95.09	66.67	26
CDH-CapsNet (Banana)	0	33	1	0	155	99.47	97.10	100	99.36	98.53	33
	1	28	0	7	154	96.30	100	80.00	100	88.89	35
	2	94	1	1	93	98.94	98.95	98.90	98.94	98.93	95
	3	26	6	0	157	96.83	81.25	100	96.32	89.66	26

For sensitivity analysis using the Apple dataset (Table 1), the original CapsNet model failed to detect 15.08% of class 0 images, 8.00% of class 1 images, 7.27% of class 2 images, and 3.65% of class 3 images, indicating high false negative assignments. In contrast, the CDH-CapsNet model missed only 3.97%, 0.8%, 1.82%, and 0.3% of images for the same classes, indicating fewer false negatives. Also, when the Banana dataset (in Table 2) was considered, the original CapsNet model failed to detect 15.55% of images in class 0, 31.14% in class 1, 11.58% in class 2, and 34.61% in class 3, resulting in high false-negative assignments. In contrast, the CDH-CapsNet model missed only 0%, 20%, 1.1%, and 0% of images for the same classes, indicating fewer false negatives. This suggests that the CDH-CapsNet model significantly reduced the number of missed images across different classes, demonstrating a more accurate classification than the original model.

For specificity, utilizing the Apple dataset (in Table 1), the original CapsNet model had higher false positives, misclassifying 1.38% of class 0 images, 2.55% of class 1 images, 0.86% of class 2 images, and 6.21% of class 3 images, which are incorrect assignments to these categories. In contrast, the CDH-CapsNet model had lower false positives, misclassifying only 0.39%, 0.39%, 0%, and 1.31% of images belonging to the same classes. Additionally, when the Banana dataset was used (in Table 2), the original CapsNet model yielded higher false positives, misclassifying 2.56% of class 0 images, 11.04% of class 1 images, 7.45% of class 2 images, and 4.91% of class 3 images, which are incorrect assignments to these categories. In contrast, the CDH-CapsNet model had lower false positives, misclassifying only 0.64%, 0%, 1.06%, and 3.68% of images belonging to the same classes. This indicates that the CDH-CapsNet was more effective in correctly excluding images outside a defined class than the original CapsNet model.

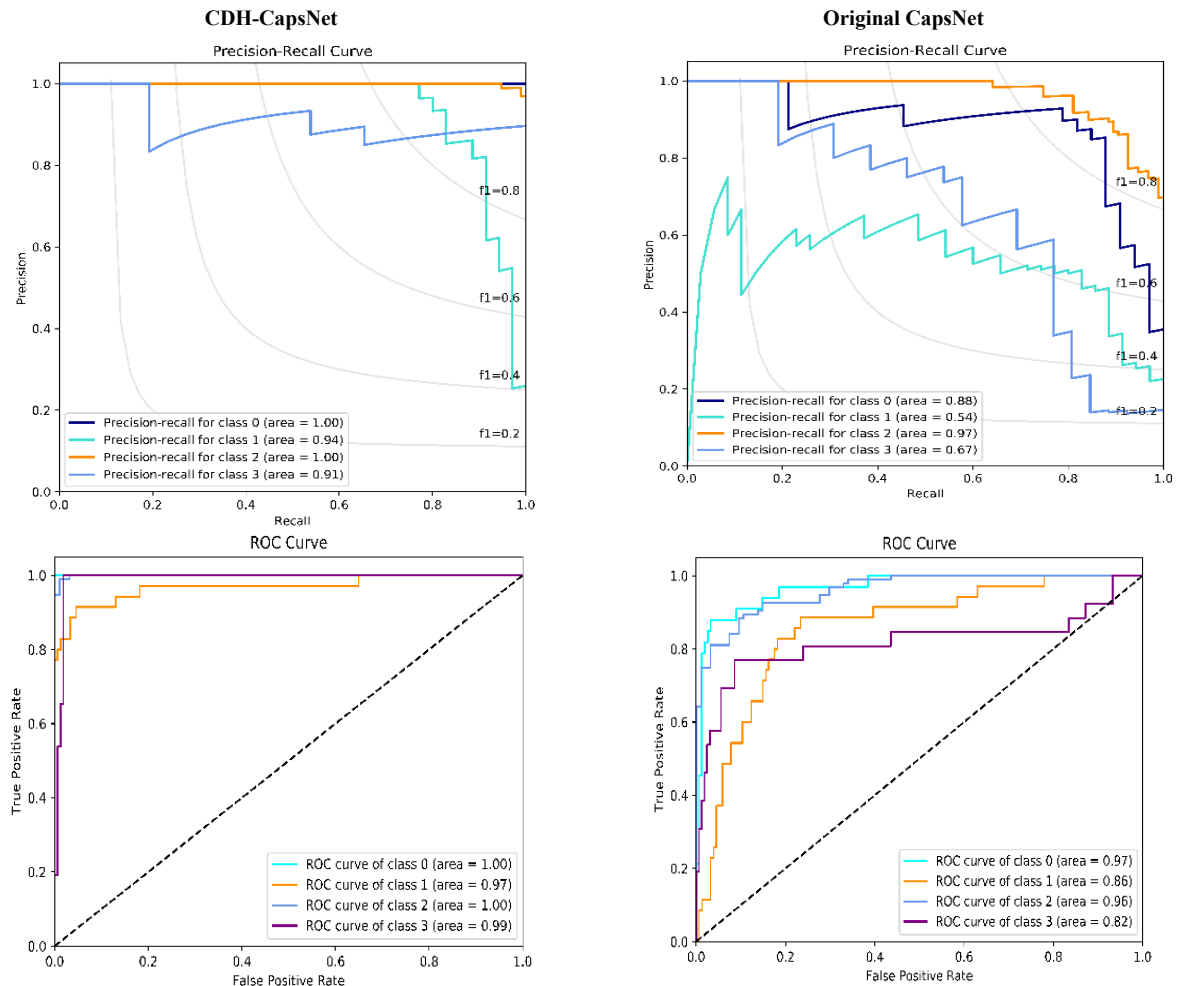


Fig. 5. ROC and PR curves for CapsNet and the Original CapsNet using the banana dataset

For F1-Score, which indicates a balance of precision and sensitivity or recall, the original model struggled 11.20% in class 0, 9.09% in class 1, 8.11% in class 2, and 4.66% in class 3 to be more precise and sensitive to images than the CDH-CapsNet, which only struggled 2.81%, 1.20%, 0.92%, and 0.75% as a fair balance for the classes 0, 1, 2, and 3 for the apple dataset as seen in Table 1. Also, the original model struggled 14.05% in class 0, 36.72% in class 1, 9.68% in class 2, and 33.33% in class 3 to be more precise and sensitive to images than the CDH-CapsNet, which struggled 1.47%, 11.11%, 1.07%, and 10.34% as a fair balance for the classes 0, 1, 2, and 3 when the Banana dataset was considered (as seen in Table 2).

The CDH-CapsNet achieved higher values for precision, specificity, sensitivity, F1-score, and accuracy per class on all datasets than the traditional CapsNet model.

Figure 5 depicts the results of both ROC and PR curves for the banana dataset. Analysis of the ROC curves reveals that the CDH-CapsNet model attained better representation and higher values compared to the original CapsNet model. The CDH-CapsNet achieved overall ROC and PR values of 99.00% and 96.25%, respectively, when considering the banana dataset, compared to the lower ROC and PR values of 90.25% and 76.50%, respectively, attained by the original CapsNet for the same dataset. This indicates that the CDH-CapsNet model is more robust and can differentiate between the various classes at different thresholds than the original CapsNet model. The proposed model achieved high PR values across all datasets, indicating strong performance on imbalanced data. This highlights its robustness, particularly in highly imbalanced datasets such as those for apples, bananas, corn (maize), grapes, peppers, potatoes, and tomatoes.

Overall, the CDH-CapsNet model demonstrated highly promising performance, considering reliability, accuracy, and the probability of negative predictions and false positives on all the datasets used, outperforming the original CapsNet and performing comparably well compared to other cutting-edge CapsNet models. This represents a significant improvement for plant disease diagnosis. The better results attained by the CDH-CapsNet are due to CDH's selective extraction of essential features, the addition of two extra layers of convolutions, and the max pooling layers combined to make the encoder network very efficient in feature extraction from the input images.

4.2. Ablation Study

Further experimentation was conducted to determine the model's components, which significantly impacted its performance [35]. This demonstrates the model's flexibility and resilience, as it performs even when some layers are removed. Considering the Apple and Potato datasets, it is evident that the CDH-CapsNet achieves 95.43% and 98.14% validation accuracies, respectively, without the CDH layer, and achieves appreciable validation accuracies when the CDH layer is implemented with some of the other layers in place. Table 3 shows that the model is impacted more by the CDH layer than by any other layer.

Table 3. Results from the ablation study

Layers	Validation accuracy (%)	
	Apple	Potato
-CDH	95.43	98.14
-Conv1	97.80	98.61
-MP1	98.11	99.54
-Conv2	97.95	99.07
-MP2	98.27	99.67
-Conv3	97.80	99.07
-MP3	98.43	99.64
+All Layers	98.43	99.77

4.3. Size on Disk and Parameter Count

Dealing with a high volume of parameters is a significant challenge for deep learning models. This results in increased computational demands and processing time, leading to high memory usage and limiting their suitability for devices with limited resources. The CDH-CapsNet model produced fewer parameters and is smaller on disk than the original or traditional CapsNet and other cutting-edge architectures in the literature. For apple, banana, grape, corn, mango, pepper, potato, rice, tomato, and CIFAR-10 datasets, a reduction in Millions (M) of 6.04M, 6.04M, 6.04M, 6.04M, 7.08M, 5.52M, 5.78M, 6.04M, 7.59M, and 7.59M respectively in parameter count is realized comparing the CDH-CapsNet to the original or traditional CapsNet, and 23MB, 23MB, 23MB, 23MB, 26.9MB, 21MB, 22MB, 23MB, 29MB, and 29MB respectively of size on disk in Megabytes (MB) reduction is also realized for the same dataset comparing the CDH-CapsNet and the traditional CapsNet as seen in Table 4, indicating the lower computational complexity of the CDH-CapsNet compared with that of the original or traditional CapsNet, making it appropriate for embedded or mobile device implementation. This can be attributed to the non-trainable parameters generated by the CDH, as well as the reduction in the number of channels in the primary capsules and the kernel size. It can also be inferred from Table 4 that the original CapsNet model is more than 40% computationally more expensive than the CDH-CapsNet. The Comparison of parameter count in millions (M) and size on disk in Megabytes (MB) for the CDH-CapsNet, traditional/Original, and other CapsNet models found in the literature is shown in Table 4.

Table 4. Parameter count (P)(Millions (M)) and size on disk (S)(MB) comparison

CapsNet Models/Reference		P(M) and S (MB)									
		Apple	Banana	Grape	Maize	Mango	Pepper	Potato	Rice	Tomato	CIFAR-10
Gabor CapsNet [13]	P	-	-	-	-	-	-	-	-	12.00	-
CapsNet [15]	P	-	-	-	8.40	-	-	-	-	8.40	5.20
CapsNet[17]	P	-	-	-	-	-	-	9.86	-	-	-
Dual-Input CapsNet[18]	P	-	-	-	-	-	-	-	-	6.04	5.48
K-Means CapsNet [36]	P	-	-	-	-	-	-	-	-	5.12	-
Gabor-Maxpooled CapsNet [37]	P	-	-	-	-	-	-	-	-	8.71	-
Shallow/Multi-Input CapsNet [38]	P	-	-	-	-	-	-	-	-	4.10/ 4.00	4.60/ 4.30
Original CapsNet [11]	P	10.13	10.13	10.13	10.13	11.21	9.59	9.86	10.13	11.75	11.75
	S	38.6	38.6	38.6	38.6	42.7	36.5	37.6	38.6	44.8	44.8
CDH-CapsNet Proposed	P	4.09	4.09	4.09	4.09	4.13	4.07	4.08	4.09	4.16	4.16
	S	15.6	15.6	15.6	15.6	15.8	15.5	15.6	15.6	15.8	15.8

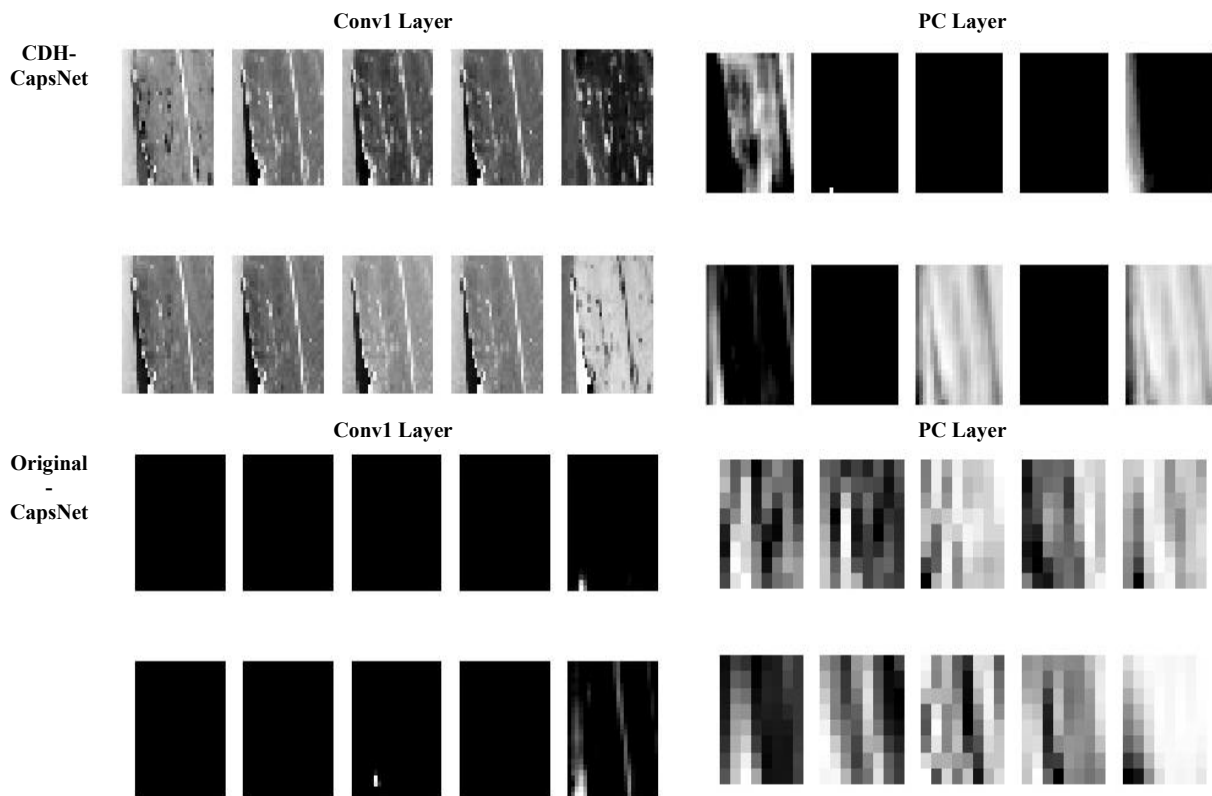


Fig. 6. Activation maps for CDH-CapsNet and Original CapsNet models using the corn dataset

4.4. Visualization of Activation Maps and Clusters

Figure 6 illustrates the visualized feature maps from the CDH and traditional CapsNet models using the corn dataset. The features extracted by the initial convolution layers for CDH and traditional CapsNet are analyzed and compared. A feature map from the CDH is extracted by the first convolution layer (conv1) and compared with the feature map from the first convolution layer of the traditional/original CapsNet. It can be seen that the activations from the proposed CDH-CapsNet model are much more detailed than those of the traditional CapsNet. This suggests that the convolution layer alone, as found in traditional CapsNets, is insufficient to extract highly relevant features from the input image. This limitation in feature extraction by the original CapsNet impacts the Primary Capsule layer, as it fails to receive the essential features needed to distinguish between different classes in the dataset. It can also be observed that the CDH layer extracts pertinent features, which are then passed to the first convolutional layer (conv1) for the CDH-CapsNet and then to the primary capsule layer (PC), thereby distinguishing between the different classes in the dataset. This is evident when comparing the activation maps of the traditional CapsNet model in Figure 6 to those of the CDH-CapsNet model in the same Figure. Moreover, the t-distributed stochastic neighbour embedding (t-SNE) technique was employed to visualize the separability of clusters formed at the Class Capsule layer of the models. Figure 7 displays the clusters formed at the

class capsule layers for the CIFAR-10, tomato, and mango datasets of the CDH-CapsNet (first column) and original CapsNets (second column), respectively. The formation of distinct clusters signifies the model's strong ability to categorize each image into its appropriate class accurately. It is evident that the CDH-CapsNet model can discriminate or distinguish the various classes of the datasets well, with fewer and closer outliers than the original CapsNet model. This enhances the clarity of CapsNet's internal processes and supports the development of explainable artificial intelligence (XAI).

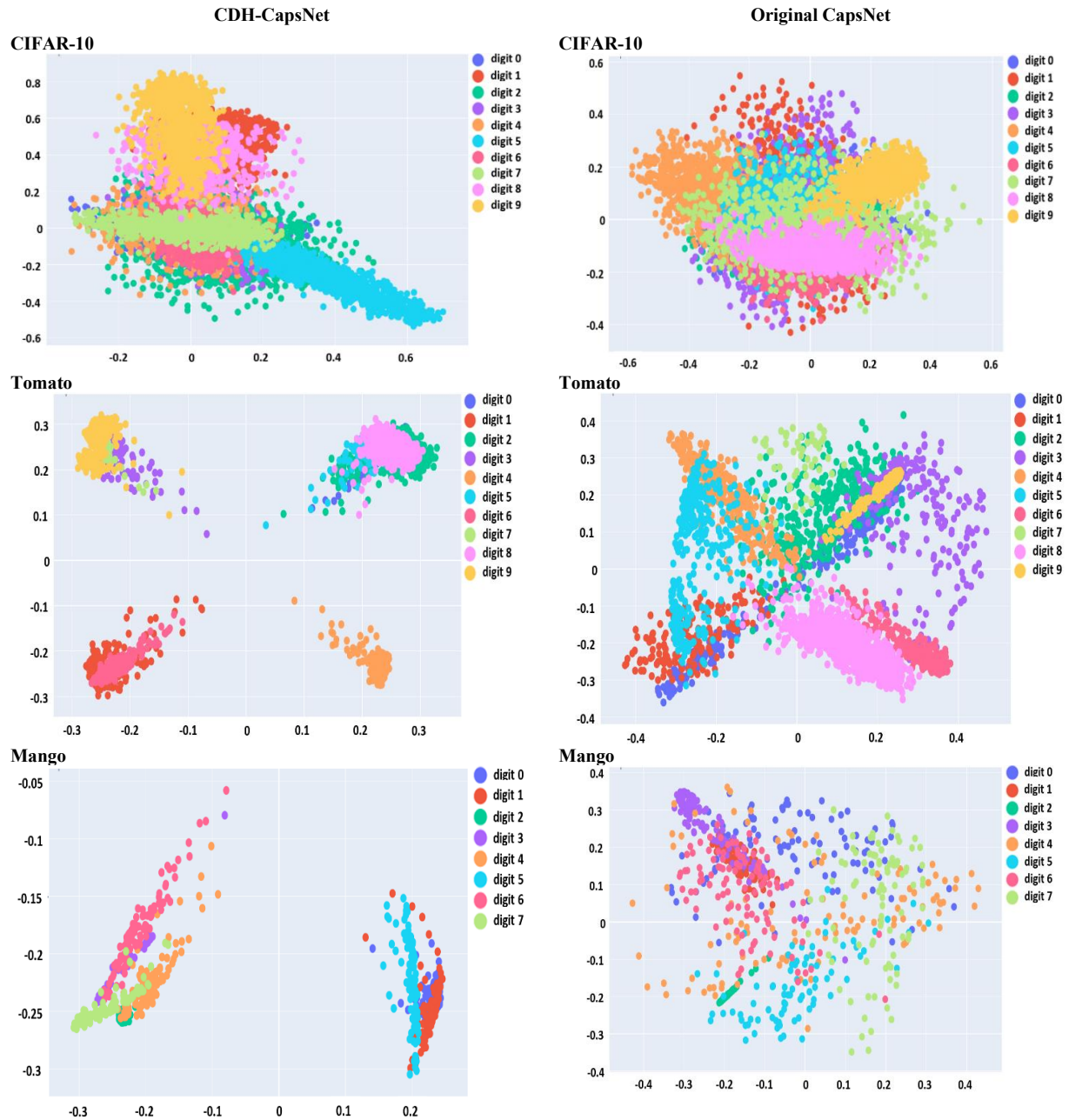


Fig. 7. Clusters formed at the class capsule layers (CDH-CapsNet and original models) for CIFAR-10, tomato, and mango datasets

4.5. Prediction and Reconstruction

The reconstruction method helps to know the probable class of an input image, and the decoder network is used to display the reconstructed images of plant disease images. Considering the pepper and rice datasets, using the CDH-CapsNet and Original CapsNet models, the images generated by the CDH-CapsNet model show better quality visuals and higher-class predictions than those generated by the original CapsNet model, as shown in Figure 8. The first columns in this figure show the class labels along with the predicted images that require reconstruction.

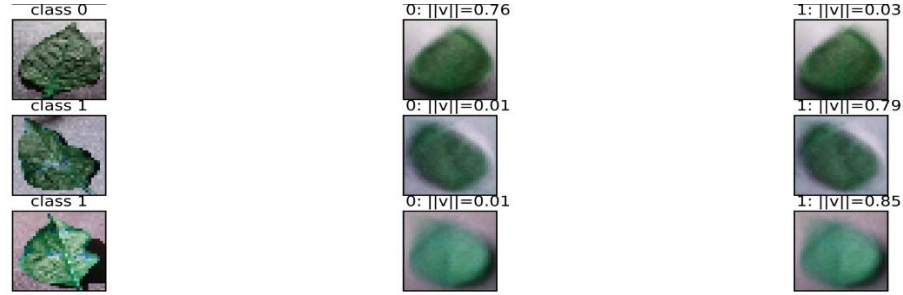
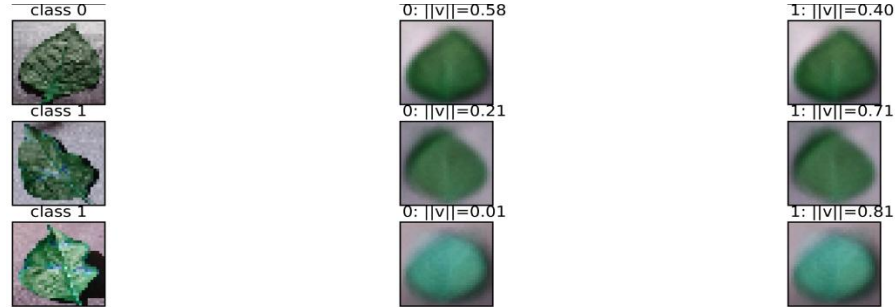
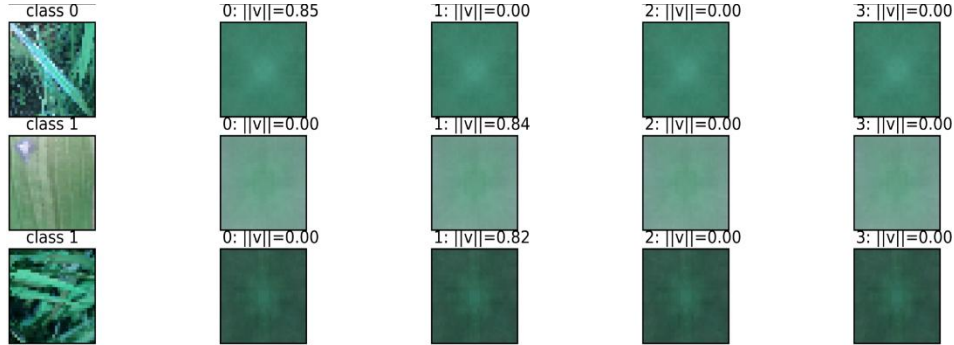
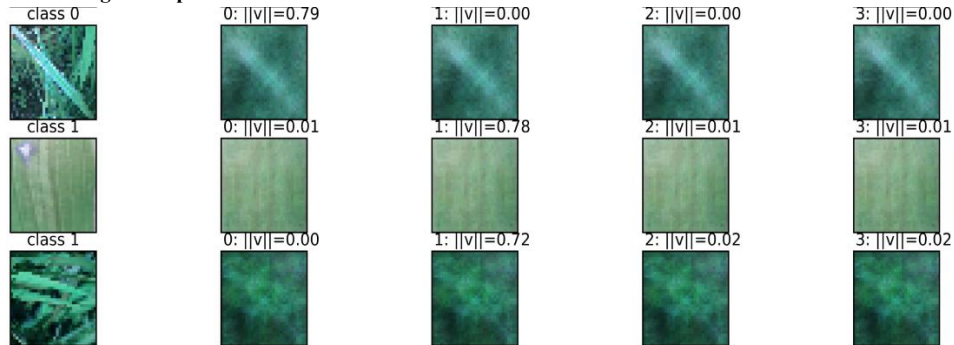
Pepper – CDH-CapsNet**Pepper – Original CapsNet****Rice – CDH-CapsNet****Rice – Original CapsNet**

Fig. 8. Reconstructed images for CDH-CapsNet and Original CapsNet using the pepper and rice datasets

4.6. Comparison of Results

We compared the proposed CDH-CapsNet model with the original CapsNet and several other leading models reported in the literature. Although our model emphasizes dynamic routing with structural enhancements, we included various routing methods in the comparison. As shown in Table 5, CDH-CapsNet outperformed the original CapsNet with accuracy improvements of 5.83%, 14.82%, 5.9%, 4.42%, 20.87%, 40.12%, 4.41%, 0.76%, 9.49%, and 13.97% on the apple, banana, grape, corn, mango, pepper, potato, rice, tomato, and CIFAR-10 datasets, respectively. It also achieved competitive results against other top-tier models. Again, on the datasets mentioned, the CDH-CapsNet model achieves parameter reductions of 6.04M, 6.04M, 6.04M, 6.04M, 7.08M, 5.52M, 5.78M, 6.04M, 7.59M, and 7.59M, respectively, compared to the original CapsNet. Corresponding reductions in disk size are 23 MB, 23 MB, 23 MB, 23 MB, 26.9 MB, 21 MB, 22 MB, 23 MB, 29 MB, and 29 MB. These reductions demonstrate the lower computational complexity of CDH-CapsNet, making it more fitting for deployment on resource-constrained devices. This notable performance is primarily attributed to the inclusion of CDH, along with two convolutional, three max pooling, and three batch normalization layers

in the encoder network, which significantly enhances feature extraction and makes a meaningful contribution to the field of computer vision.

CapsNet variants and hybrid CapsNet models, which used the same plant disease datasets and CIFAR-10, are also compared, as shown in Table 5. For hybrid models like Gabor CapsNet [13], E-GAN CapsNet [14], K-Means CapsNet [36], Gabor-Maxpooled CapsNet [37], Shallow/Multi-Input CapsNet [38], SE-SK CapsNet[39], Multi-Channel CapsNet[40], etc., the CDH-CapsNet performed comparably well in terms of accuracy, number of parameters, and size on disk as seen in Tables 4 and 5. Only CapsNet models were considered, as the study aims to improve the performance of CapsNet.

An essential aspect of this study was to evaluate not only the performance of the proposed CDH-CapsNet model on plant-specific datasets but also its ability to generalize beyond the narrow domain of plant pathology. To this end, the CIFAR-10 dataset was deliberately included not to assess the model's efficacy in general-purpose object recognition per se, but to serve as a rigorous benchmark for testing its generalization capability. While plant disease datasets typically exhibit fine-grained features and high inter-class similarity, CIFAR-10 presents an entirely different challenge: it contains diverse, low-resolution images of unrelated object classes, with cluttered backgrounds and significant intra-class variability. These characteristics make it a valuable testbed for understanding how well a model can handle more complex, unstructured visual data. The CDH-CapsNet model achieved 77.55% accuracy on CIFAR-10, which, although lower than its performance on plant datasets, is comparable to and slightly exceeds the results of the prior original CapsNet on the same dataset that achieved 63.58%, and other CapsNet-based models on the same dataset such as those reported by Mensah et al. [15,18, 38], who achieved 75.80%, 76.58%, and 75.7/63.95 respectively. This consistency suggests that our model possesses general representational ability and is not merely overfitting to the structure and regularity of the plant disease images. It also confirms the robustness of the model's feature extraction and classification pipeline, especially in handling diverse and noisy input data. This helps to demonstrate that CDH-CapsNet is an improvement over CapsNet.

The CDH-CapsNet model delivered outstanding performance across all evaluated datasets, outperforming the original CapsNet on several occasions. However, this strong performance is likely influenced by the high quality and low intra-class variability of the datasets used, conditions that may not accurately reflect the complexity of real-world scenarios. Although the models consistently achieve high accuracy on standardized benchmark plant datasets, their effectiveness in uncontrolled environments characterized by varying lighting conditions, background distractions, and image noise remains to be established. This underscores the need for comprehensive validation under real-world field conditions to assess the model's practical robustness. It must be noted that the original CapsNets and other state-of-the-art models compared in Tables 4 and 5 used the same datasets, and the main aim of this study was to help improve CapsNets, which was achieved. Furthermore, while the model performs exceptionally well on clean, well-structured datasets, its varying performance across more diverse datasets suggests potential sensitivity to data complexity and class imbalance. These insights underscore the need for cautious interpretation of benchmark results, as high accuracy in controlled settings does not always translate to similar success in real-world applications.

Table 5. Comparison of the proposed CDH-capsnet model and previous works

CapsNet Models/Reference	Validation Accuracy (%)									
	Apple	Banana	Grape	Corn	Mango	Pepper	Potato	Rice	Tomato	CIFAR-10
Gabor CapsNet [13]	-	-	-	-	-	-	-	-	98.13	-
E-GAN CapsNet [14].	-	-	97.63	-	-	-	-	-	-	-
CapsNet [15]	-	-	-	96.79	-	-	-	-	98.06	75.80
Dilated CapsNet[16]	93.16	-	-	-	-	-	-	-	-	-
CapsNet[17]	-	-	-	-	-	-	91.83	-	-	-
Dual-Input CapsNet[18]	-	-	-	-	-	-	-	-	93.03	76.58
CapsNet[19]	-	95.00	-	-	-	-	-	-	-	-
CapsNet [21]	-	-	-	-	-	-	-	-	96.39	-
CapsNet[23]	-	-	-	-	-	95.76	-	-	-	-
K-Means CapsNet [36]	-	-	-	97.99	-	-	-	-	98.80	-
Gabor-Maxpooled CapsNet [37]	-	-	-	-	-	-	-	-	97.98	-
Shallow/Multi-Input CapsNet [38]	-	-	-	-	-	-	-	-	97.33/ 94.04	75.7/ 63.95
SE-SK CapsNet[39]	-	-	-	-	-	-	-	-	97.19	-
Multi-Channel CapsNet[40]	-	-	-	-	-	-	-	-	98.15	-
ConvCapsNet[41]	-	-	99.12	-	-	-	-	-	-	-
Original CapsNet [11]	92.91	80.95	93.49	92.59	78.38	59.68	95.36	99.24	88.59	63.58
CDH-CapsNet Proposed	98.74	95.77	99.51	97.01	99.25	99.80	99.77	100	98.08	77.55

5. Conclusions

An efficient CDH-CapsNet model was proposed for classifying nine publicly available plant disease datasets and the CIFAR-10 dataset. To enhance the original CapsNet, we incorporated a Color Difference Histogram (CDH) layer, two additional convolutional layers, and three max pooling layers, all to improve feature extraction from input images. Additionally, three batch normalization layers were added to ensure consistent data distribution, simplify learning, and accelerate training. We reduced the primary capsule channels from 32 to 16, which also lowered the parameter count and computational cost, mitigated overfitting, sped up training, and improved suitability for small datasets and resource-constrained devices. The input image size was also reduced to 32x32x3, which helped lower the parameter count and computational cost.

The CDH layer proved to be an effective feature extractor, enabling the proposed CDH-CapsNet model to achieve strong performance across multiple evaluation metrics, including specificity, accuracy, precision, sensitivity, F1-score, and PR and ROC curve values, comparable to state-of-the-art models. Furthermore, the model attained fewer parameters and a smaller disk size, making it feasible for deployment on devices with limited resources. Visualization of class clusters, activation maps, and reconstructed images showed improved quality compared to the original CapsNet. An ablation study further contributed to the model's interpretability and the broader field of explainable AI (XAI).

In summary, this study successfully enhanced the performance of the original Capsule Network on complex images, such as plant disease images and CIFAR-10, through the development of the CDH-CapsNet model. The proposed architecture is efficient and computationally lightweight, consistently outperforming the original CapsNet and several other advanced CapsNet variants when evaluated on benchmark datasets. This positions CDH-CapsNet as a promising tool for plant disease classification, offering a scalable solution suitable for field use in resource-constrained settings and contributing to efforts aligned with Sustainable Development Goal 2 (SDG 2). Beyond its immediate technical merits, this work has significant implications for the field of computer vision and agricultural technology. It demonstrates that, with thoughtful architectural innovations, Capsule Networks' performance on complex images can be improved to achieve even deployability on edge devices. This opens pathways for applying CapsNet-based systems in practical, real-world scenarios where lightweight and interpretable models are crucial.

Moreover, the insights gained from reducing primary capsule channels and integrating CDH features provide valuable guidance for designing efficient neural networks tailored to small datasets and resource-limited environments. These findings highlight the potential of hybrid approaches that combine specific feature extraction capabilities with advanced neural architectures, contributing to ongoing discussions on how to bridge the gap between theoretical model capabilities and practical deployment.

While these results are promising, we acknowledge that the high performance (especially on the rice dataset) may partly reflect the relatively homogeneous nature of the dataset used. Real-world conditions often involve diverse lighting, complex backgrounds, and environmental noise, which could affect the model's performance. Future work will focus on improving generalizability and practical relevance by testing the model in more diverse and challenging environments, expanding the range of datasets, and exploring real-time deployment on edge and mobile devices for agricultural applications. Additionally, we aim to enhance the encoder network to further boost model performance and achieve greater optimization through reduced parameter counts.

All the Declarations and Statements

Author Contributions Statement

Steve Okyere-Gyamfi: Conceptualization, Data Curation, Methodology, Model Training, Validation, Visualization, Performance Evaluation, and Writing: Proposed the research concept, designed the overall framework, managed data collection, conducted dataset preprocessing, built and implemented the research model, carried out model training and validation of results, compared performance with existing models, and prepared the initial manuscript draft.

Michael Asante: Supervision, Project Management, and Writing – Review and Editing: Supervised project implementation, assisted in organizing key milestones and timelines, critically reviewed and refined the manuscript, and ensured overall clarity and logical flow.

Kwame Ofosuhen Peasah: Supervision, Project Management, and Writing – Review and Editing: Supervised project implementation, assisted in organizing key milestones and timelines, critically reviewed and refined the manuscript, and ensured overall clarity and logical flow.

Yaw Marfo Missah: Supervision, Project Management, and Writing – Review and Editing: Supervised project implementation, assisted in organizing key milestones and timelines, critically reviewed and refined the manuscript, and ensured overall clarity and logical flow.

Vivian Akoto-Adjepong: Data Curation, Validation, Performance Evaluation, and Writing– contributed to the literature survey: Participated in data curation, assisted in validating results, evaluated performance in comparison with existing models, and supported the literature review process.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors state that they have no conflicts of interest.

Funding Declaration

None

Data Availability Statement

This study analyzed publicly available datasets. The datasets can be found here:

Apple, grape, corn, pepper, potato, and tomato datasets: <https://github.com/spMohanty/PlantVillage-Dataset> : accessed on 26th September 2024.

Banana dataset: <https://data.mendeley.com/datasets/9tb7k297ff/1> : accessed on 10th October 2024.

Mango dataset: <https://data.mendeley.com/datasets/hxsnvwt3r/1> : accessed on 7th August 2024.

Rice dataset: <https://data.mendeley.com/datasets/fwcj7stb8r/1> : accessed on 10th August 2024.

CIFAR-10 dataset: <https://www.cs.toronto.edu/~kriz/cifar.html>: accessed on 12th October 2024.

Ethical Declarations

None

Acknowledgments

None

Declaration of Generative AI in Scholarly Writing

None

Abbreviations

The following abbreviations are used in this manuscript:

CNN - Convolutional Neural Network

CapsNet - Capsule Neural Network

CDH - Color Difference Histogram

ROC - Receiver Operating Characteristic

PR - Precision-Recall

AUC - Area under the curve

SDG - Sustainable Development Goals

GPU - Graphics Processing Unit

VRAM - Video Random Access Memory

RTX - Ray Tracing Texel eXtreme

CIFAR-10 - Canadian Institute for Advanced Research 10-class dataset

XAI - Explainable Artificial Intelligence

t-SNE - t-distributed stochastic neighbour embedding

PC - primary capsule layer

CC - Class Capsule layer

APPENDIX

None

References

- [1] V. K. Vishnoi, K. Kumar, B. Kumar, S. Mohan, and A. A. Khan, "Detection of Apple Plant Diseases Using Leaf Images Through Convolutional Neural Network," *IEEE Access*, vol. 11, no. November 2022, pp. 6594–6609, 2023, doi: 10.1109/ACCESS.2022.3232917.
- [2] H. Liu et al., "WSRD-Net: A Convolutional Neural Network-Based Arbitrary-Oriented Wheat Stripe Rust Detection Method," *Front. Plant Sci.*, vol. 13, no. May, pp. 1–15, 2022, doi: 10.3389/fpls.2022.876069.
- [3] S. Jasrotia, J. Yadav, N. Rajpal, M. Arora, and J. Chaudhary, "Convolutional Neural Network Based Maize Plant Disease Identification," *Procedia Computer Science*, vol. 218, pp. 1712–1721, 2023, doi: 10.1016/j.procs.2023.01.149.
- [4] J. Lin et al., "GrapeNet: A Lightweight Convolutional Neural Network Model for Identification of Grape Leaf Diseases," *Agric. MDPI*, 2022.
- [5] V. Rajpoot, R. Dubey, P. K. Mannepalli, P. Kalyani, S. Maheshwari, A. Dixit, and A. Saxena, "Mango Plant Disease Detection System Using Hybrid BBHE and CNN Approach," *Traitement du Signal*, vol. 39, no. 3, pp. 1071–1078, 2022, doi: 10.18280/ts.390334.

- [6] Y. A. Bezabih, A. O. Salau, B. M. Abuhayi, A. A. Mussa, and A. M. Ayalew, "CPD-CCNN: classification of pepper disease using a concatenation of convolutional neural network models," *Sci. Rep.*, vol. 13, no. 1, pp. 1–13, 2023, doi: 10.1038/s41598-023-42843-2.
- [7] M. H. Al-Adhaileh, A. Verma, T. H. H. Aldhyani, and D. Koundal, "Potato Blight Detection Using Fine-Tuned CNN Architecture," *Mathematics*, vol. 11, no. 6, 2023, doi: 10.3390/math11061516.
- [8] S. Verma, P. Kumar, and J. P. Singh, "A Unified Lightweight CNN-based Model for Disease Detection and Identification in Corn, Rice, and Wheat," *IETE J. Res.*, 2023, doi: 10.1080/03772063.2023.2181229.
- [9] D. L. Shanthi, K. Vinutha, N. Ashwini, and S. Vashistha, "Tomato Leaf Disease Detection Using CNN," *Procedia Comput. Sci.*, vol. 235, no. 2023, pp. 2975–2984, 2024, doi: 10.1016/j.procs.2024.04.281.
- [10] M. Ashraf, M. Abrar, N. Qadeer, A. A. Alshdadi, T. Sabbah, and M. A. Khan, "A Convolutional Neural Network Model for Wheat Crop Disease Prediction," *Comput. Mater. Contin.*, vol. 75, no. 2, pp. 3867–3882, 2023, doi: 10.32604/cmc.2023.035498.
- [11] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017, pp. 3856–3866.
- [12] S. Verma, A. Chug, R. P. Singh, A. P. Singh, and D. Singh, "SE-CapsNet: Automated evaluation of plant disease severity based on feature extraction through Squeeze and Excitation (SE) networks and Capsule networks," *Kuwait Journal of Science*, vol. 49, no. 1, pp. 1–31, 2022, doi: 10.48129/kjs.v49i1.10586.
- [13] P. K. Mensah, B. A. Weyori, and M. A. Ayidzoe, "Gabor Capsule Network for Plant Disease Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 10, pp. 388–395, 2020.
- [14] N. Vasudevan and T. Karthick, "A Hybrid Approach for Plant Disease Detection Using E-GAN and CapsNet," *Comput. Syst. Sci. Eng.*, vol. 46, no. 1, pp. 337–356, 2023, doi: 10.32604/csse.2023.034242.
- [15] P. Mensah, B. Asubam, and A. Abra, "Exploring the performance of LBP-capsule networks with K-Means routing on complex images," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2574–2588, 2022, doi: 10.1016/j.jksuci.2020.10.006.
- [16] C. Xu, X. Wang, and S. Zhang, "Dilated convolution capsule network for apple leaf disease identification," *Front. Plant Sci.*, vol. 13, no. November, pp. 1–13, 2022, doi: 10.3389/fpls.2022.1002312.
- [17] S. Verma, A. Chug, and A. P. Singh, "Exploring capsule networks for disease classification in plants," *J. Stat. Manag. Syst.*, vol. 23, no. 2, pp. 307–315, 2020, doi: 10.1080/09720510.2020.1724628.
- [18] P. K. Mensah and M. A. Ayidzoe, "Overview of CapsNet Performance Evaluation Methods for Image Classification using a Dual Input Capsule Network as a Case Study," *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 29–43, 2022, doi: 10.12785/ijcds/120104.
- [19] B. F. Oladejo and O. O. Ademola, "Automated Classification of Banana Leaf Diseases using an Optimized Capsule Network Model," pp. 119–130, 2020, doi: 10.5121/csit.2020.100910.
- [20] P. K. Mensah, B. A. Weyori, and A. A. Mighty, "Max-pooled fast learning gabor capsule network," 2020 Int. Conf. Artif. Intell. Big Data, *Comput. Data Commun. Syst. icABCD 2020 - Proc.*, 2020, doi: 10.1109/icABCD49160.2020.9183823.
- [21] L. M. Abouelmagd, M. Y. Shams, H. S. Marie, and A. E. Hassanien, "An optimized capsule neural networks for tomato leaf disease classification," *Eurasip J. Image Video Process.*, vol. 2024, no. 1, 2024, doi: 10.1186/s13640-023-00618-9.
- [22] X. Zhang, Y. Mao, Q. Yang, and X. Zhang, "A Plant Leaf Disease Image Classification Method Integrating Capsule Network and Residual Network," *IEEE Access*, vol. 12, no. February, pp. 44573–44585, 2024, doi: 10.1109/ACCESS.2024.3377230.
- [23] G. ALTAN, "Performance Evaluation of Capsule Networks for Classification of Plant Leaf Diseases," *Int. J. Appl. Math. Electron. Comput.*, vol. 8, no. 3, pp. 57–63, Sep. 2020, doi: 10.18100/ijamec.797392.
- [24] G. H. Liu and J. Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognit.*, vol. 46, no. 1, pp. 188–198, 2013, doi: 10.1016/j.patcog.2012.06.001.
- [25] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," 2015, [Online]. Available: <http://arxiv.org/abs/1511.08060>
- [26] S. E. Arman, M. A. B. Bhuiyan, H. M. Abdullah, S. Islam, T. T. Chowdhury, and M. A. Hossain, "BananaLSD: A banana leaf images dataset for classification of banana leaf diseases using machine learning," *Data Br.*, vol. 50, p. 109608, 2023, doi: 10.1016/j.dib.2023.109608.
- [27] S. I. Ahmed et al., "MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves," *Data Br.*, vol. 47, p. 108941, 2023, doi: 10.1016/j.dib.2023.108941.
- [28] P. K. Sethy, N. K. Barpanda, A. K. Rath, & S. K. Behera, "Deep feature-based rice leaf disease identification using support vector machine," *Computers and Electronics in Agriculture*, 105527, 175 May, 2020, <https://doi.org/10.1016/j.compag.2020.105527>
- [29] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," Technical Report, University of Toronto, Toronto, ON, Canada, 2009.
- [30] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv preprint arXiv:1602.07360*, 2016.
- [31] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, 2021, doi: 10.1145/3446776.
- [32] W. Samek and K.-R. Müller, "Towards Explainable Artificial Intelligence," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11700 LNCS, 2019, pp. 5–22. doi: 10.1007/978-3-030-28954-6_1.
- [33] A. Shahroudnejad, P. Afshar, K. N. Plataniotis, and A. Mohammadi, "IMPROVED EXPLAINABILITY OF CAPSULE NETWORKS: RELEVANCE PATH BY AGREEMENT," in 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov. 2018, pp. 549–553. doi: 10.1109/GlobalSIP.2018.8646474.
- [34] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, Dec. 2019, doi: 10.1126/scirobotics.aay7120.
- [35] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation Studies in Artificial Neural Networks," *arXiv preprint arXiv:1901.08644*, 2019.
- [36] P. K. Mensah, B. A. Weyori, and M. A. Ayidzoe, "Capsule network with K-Means routing for plant disease recognition," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 1, pp. 1025–1036, 2021, doi: 10.3233/JIFS-201226.
- [37] P. K. Mensah, B. A. Weyori, & A. A. Mighty, "Max-pooled fast learning gabor capsule network." 2020 International Conference

- on Artificial Intelligence, Big Data, Computing and Data Communication Systems, IcABCD-Proceedings, 2020, <https://doi.org/10.1109/icABCD49160.2020.9183823>
- [38] P. K. Mensah, B. A. Weyori, and M. A. Ayidzoe, "Evaluating shallow capsule networks on complex images," *Int. J. Inf. Technol.*, vol. 13, no. 3, pp. 1047–1057, 2021, doi: 10.1007/s41870-021-00694-y.
- [39] X. Zhang, Y. Mao, Q. Yang, & X. Zhang, "A Plant Leaf Disease Image Classification Method Integrating Capsule Network and Residual Network", *IEEE Access*, 44573–44585, 12(February). 2024, <https://doi.org/10.1109/ACCESS.2024.3377230>
- [40] M. Peker, "Multi-channel capsule network ensemble for plant disease detection," *SN Appl. Sci.*, vol. 3, no. 7, 2021, doi: 10.1007/s42452-021-04694-2.
- [41] A. D. Andrushia, T. M. Neebha, A. T. Patricia, S. Umadevi, N. Anand, and A. Varshney, "Image-based disease classification in grape leaves using convolutional capsule network," *Soft Computing*, vol. 27, no. 3, pp. 1457-1470, 2023, doi:10.1007/s00500-022-07446-5.

Authors' Profiles



Mr. Steve Okyere-Gyamfi is a lecturer at the Catholic University of Ghana in the Department of Computer Science. He obtained his Master's in Computer Science from Kwame Nkrumah University of Science and Technology in 2018. His research areas are Machine Learning, Deep Learning, Data Analytics, and Hash Functions.



Prof. Michael Asante obtained his PhD in Systems Engineering at the University of Reading, United Kingdom. He is currently a professor of computer science at the Department of Computer Science at the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. His research interests are in machine learning, networking, network security, hash functions, etc.



Dr. Kwame Ofosuhen Peasah is a lecturer with the Department of Computer Science at the Kwame Nkrumah University of Science and Technology. He holds first and doctoral degrees in Computer Science. His research interests include machine learning, digital forensics, and cybersecurity.



Prof. Yaw Marfo Missah holds a Doctor of Computer Science degree and is an associate professor at Kwame Nkrumah University of Science and Technology. His research areas include Computer Vision in Medical Plant Identification, Data Mining and Visualisation, Machine Learning Algorithms, and Image Steganography.



Dr. Vivian Akoto-Adjepong obtained a Ph.D. in Computer Science from the University of Energy and Natural Resources, Ghana, in 2024. She is a lecturer in the Department of Computer Science and Informatics at the same University. Her research areas include artificial intelligence, machine learning, and deep learning.

How to cite this paper

Steve Okyere-Gyamfi, Michael Asante, Kwame Ofosuhen Peasah, Yaw Marfo Missah, Vivian Akoto-Adjepong, "Color Difference Histogram Capsule Network (CDH-CapsNet) for Plant Disease Recognition", *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.18, No.3, pp.45-63, 2026. DOI:10.5815/ijisa.2026.03.03