agronomy

MDPI

Article

Deep Learning-Based Phenotypic Analysis of Soybean Diseases and Assessment of Phylogenetic Signal

My Abdelmajid Kassem [1,*], Dounya Knizia [2] and Khalid Meksem [2]

[1] Plant Genomics and Bioinformatics Laboratory, Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC 28301, USA
[2] School of Agricultural Sciences, Southern Illinois University, Carbondale, IL 62901, USA; dounya.knizia@siu.edu (D.K.); meksem@siu.edu (K.M.)
* Correspondence: mkassem@uncfsu.edu; Tel.: +1-910-672-1955

Abstract

Soybean diseases caused by fungal, bacterial, and viral pathogens represent a major constraint to global agricultural productivity. Although molecular phylogenetic analyses have advanced the understanding of pathogen evolution, the extent to which disease phenotypes reflect evolutionary relationships remains poorly understood. In this study, we developed an integrative framework combining deep learning-based phenotypic analysis with phylogenetic inference to investigate the relationship between soybean disease symptoms and pathogen evolution. An EfficientNet-B0 convolutional neural network (CNN) was trained to classify 10 soybean disease classes comprising 703 leaf images and achieved a mean cross-validation accuracy of 98.72 ± 1.17%, a weighted F1-score of 98.74 ± 1.16%, and a macro F1-score of 98.47 ± 1.74%. Evaluation on a held-out test set generated through image-level partitioning yielded an accuracy of 96.19%, a weighted F1-score of 96.28%, and a macro F1-score of 95.86%. Latent feature embeddings revealed a structured phenotypic space with clear separation among most disease classes and enabled quantitative analyses of phenotypic similarity. To provide biological context, taxonomy-derived distance matrices and sequence-based phylogenetic analyses of the fungal subset using 28S rRNA sequences were compared with CNN-derived phenotypic representations. A Mantel test identified a moderate and statistically significant association between phenotypic and phylogenetic distances (Spearman r = 0.3393, $p = 0.0050$), indicating that pathogen evolutionary history contributes to disease phenotype while explaining only part of the observed phenotypic variation. Overall, the results demonstrate that deep learning effectively captures biologically meaningful phenotypic information while highlighting that disease symptoms arise from the combined influence of pathogen evolution, host responses, and environmental conditions. This study provides an integrative framework for combining image-based phenotyping with phylogenetic analysis to support biologically informed interpretation of plant disease phenotypes.

Keywords: soybean disease; deep learning; plant phenotyping; convolutional neural network; phylogenetic analysis; Mantel test; plant pathology; image analysis

Academic Editor: Thomas Bournaris

Received: 16 June 2026
Revised: 27 July 2026
Accepted: 30 July 2026
Published: 3 August 2026

1. Introduction

Soybean (*Glycine max* L.) is one of the most economically important leguminous crops worldwide, providing a major source of plant-based protein and oil for human consumption, livestock feed, and industrial applications [1]. Global soybean production has

increased substantially over the past decades; however, yield stability remains threatened by a wide range of diseases caused by fungal, bacterial, and viral pathogens [2]. These diseases can lead to severe yield losses, estimated to exceed billions of dollars annually, and significantly impact crop quality and food security [3,4]. Therefore, improving disease detection, understanding disease mechanisms, and developing effective management strategies are critical priorities in modern agriculture.

Plant diseases manifest through a variety of visible symptoms, including leaf lesions, chlorosis, necrosis, wilting, and morphological deformities. These phenotypic expressions arise from complex interactions between host plants and invading pathogens, governed by pathogen virulence factors and host immune responses [5]. The plant immune system, involving pattern-triggered immunity (PTI) and effector-triggered immunity (ETI), plays a crucial role in shaping disease outcomes [6]. Despite advances in molecular plant pathology, disease diagnosis in practice still relies heavily on visual inspection of symptoms by trained experts. Such approaches are inherently subjective, labor-intensive, and prone to misclassification, particularly when symptoms overlap across diseases or vary due to environmental conditions [7].

Recent advances in imaging technologies and artificial intelligence have enabled the development of automated plant disease detection systems. In particular, deep learning methods, especially convolutional neural networks (CNNs), have demonstrated remarkable success in image-based classification tasks [8]. Early work by Mohanty et al. (2016) [9] showed that deep CNNs can achieve high accuracy in identifying plant diseases across multiple crop species. Subsequent studies have further improved performance through optimized architectures, transfer learning, and data augmentation techniques [10–12]. For soybean specifically, deep learning approaches have been successfully applied to detect diseases such as sudden death syndrome, frogeye leaf spot, seed classification, and soybean mosaic virus [13,14].

Recent advances in deep learning, transfer learning, and explainable artificial intelligence (XAI) have substantially improved the accuracy and interpretability of automated plant disease detection systems [15,16]. Modern convolutional neural network (CNN) architectures and transformer-based approaches have demonstrated strong performance across diverse crop species and disease types, particularly when combined with transfer learning and augmentation strategies [17,18].

While these studies highlight the effectiveness of deep learning for disease classification, most focus primarily on predictive accuracy, with limited emphasis on the biological interpretation of learned representations. Understanding what features the model learns and how these features relate to underlying biological processes remains an open challenge [19]. Explainable artificial intelligence (XAI) techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), provide a means to visualize salient regions in input images, offering insights into model decision-making [20]. However, beyond interpretability, an important yet underexplored question is whether learned visual features capture biologically meaningful relationships among diseases. Recent studies have further emphasized the importance of explainable AI techniques, such as Grad-CAM and saliency-based visualization, for improving the interpretability and reliability of deep learning models in plant pathology applications [21,22].

In parallel, advances in genomics and phylogenomics have significantly enhanced our understanding of pathogen evolution. Molecular markers, including the internal transcribed spacer (ITS) and ribosomal RNA genes such as 28S rRNA, are widely used to infer phylogenetic relationships among fungal pathogens [23,24]. These analyses reveal evolutionary patterns such as divergence, speciation, and host adaptation [25]. For instance, phylogenetic studies have clarified relationships among key soybean pathogens, including

Fusarium, Cercospora, and Phakopsora species [26,27]. Despite these advances, linking phylogenetic relationships to observable disease phenotypes remains challenging, as similar symptoms can arise from unrelated pathogens due to convergent evolution or shared host responses [28].

The emerging field of phenomics provides a promising framework to bridge this gap. Phenomics involves the large-scale quantification and analysis of phenotypic traits, often using high-throughput imaging and computational methods [29]. By extracting quantitative representations of disease symptoms from images, it becomes possible to compare phenotypes systematically across diseases. Embedding-based approaches, where images are mapped to high-dimensional feature spaces, allow for the computation of similarity relationships between diseases [30]. These phenotypic similarity measures can then be analyzed using network-based methods to uncover patterns and clusters [31–33]. High-throughput phenotyping and image-based machine learning approaches are increasingly being applied in soybean research for disease detection, stress assessment, and yield prediction, highlighting the growing importance of AI-driven phenomics in crop improvement and precision agriculture [34–36].

Network science and systems biology have demonstrated the value of integrating heterogeneous data types to understand complex biological systems [37]. In plant pathology, combining phenotypic data with phylogenetic information could provide novel insights into the relationship between disease appearance and pathogen evolution. For example, visually similar diseases may cluster together in feature space, potentially reflecting shared infection mechanisms or host responses, even when caused by phylogenetically distant pathogens [28]. Conversely, closely related pathogens may produce distinct symptoms due to differences in virulence strategies or environmental interactions.

Despite its potential, the integration of deep phenotyping and phylogenomics remains largely unexplored in crop disease research. Previous studies have either focused on image-based classification or on molecular phylogenetics independently, without systematically comparing phenotypic and genetic relationships [38]. Furthermore, statistical approaches such as the Mantel test provide a quantitative framework to assess correlations between distance matrices, enabling rigorous comparison between phenotypic similarity and genetic divergence [39].

In this study, we present an exploratory integrative framework to investigate the relationship between soybean disease phenotypes and pathogen evolutionary relationships. Unlike previous studies that have primarily focused on improving disease classification accuracy, our objective is not only to develop an accurate deep learning classifier but also to examine whether the latent feature representations learned by a convolutional neural network contain biologically meaningful information. Specifically, we combine deep learning-based disease classification, Grad-CAM visualization, deep feature embedding analysis, taxonomic and sequence-based phylogenetic analyses, and Mantel correlation analysis within a unified analytical framework.

Rather than attempting direct genotype–phenotype inference, this study evaluates whether image-derived phenotypic embeddings exhibit detectable phylogenetic structure when compared with independently derived evolutionary relationships. By considering both fungal and non-fungal soybean diseases, we provide a broader assessment of phenotype–phylogeny relationships across diverse pathogen groups while recognizing that sequence-based phylogenetic analyses are applicable only to the fungal subset.

This work addresses an important gap in the current literature by moving beyond conventional image classification to investigate the biological relevance of deep learning-derived phenotypic representations. In doing so, the proposed framework not only provides accurate soybean disease recognition but also offers an interpretable approach for

exploring relationships between disease phenotypes and pathogen evolution. More broadly, this study demonstrates how explainable artificial intelligence, deep feature embeddings, and comparative phylogenetic analyses can be integrated into a reproducible framework for investigating biologically meaningful patterns in plant disease phenotypes, thereby providing a foundation for future multimodal studies that combine phenotypic, genomic, and evolutionary information.

2. Materials and Methods

2.1. Dataset and Preprocessing

A curated dataset comprising 703 soybean leaf images representing 10 disease classes was used in this study (Figure 1; Table 1). The dataset was assembled primarily from publicly available images obtained through Kaggle (Kaggle.com), supplemented with images of Sudden death syndrome (SDS) collected in our previous study on soybean SDS resistance. The dataset includes diseases caused by fungal, bacterial, and viral pathogens, providing a diverse representation of soybean disease phenotypes.

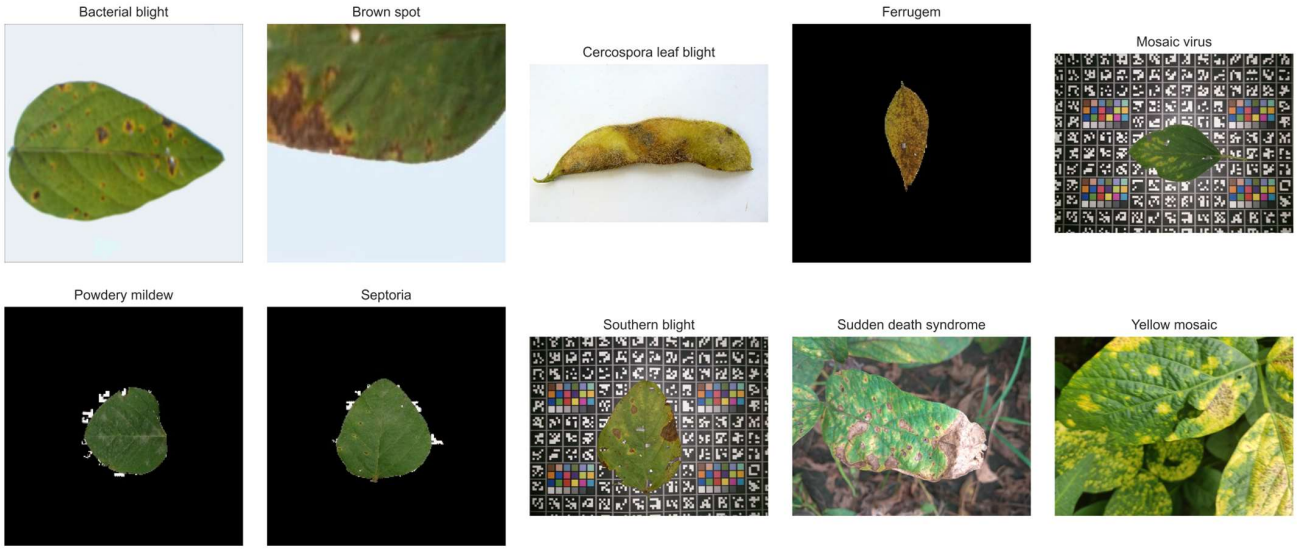

Figure 1. Representative soybean leaf images used for deep learning model development. Representative images from the ten soybean disease classes included in the dataset: bacterial blight, brown spot, Cercospora leaf blight, ferrugem (Asian soybean rust), mosaic virus, powdery mildew, Septoria leaf spot, southern blight, sudden death syndrome, and yellow mosaic. Images illustrate the diversity of disease symptoms, lesion morphology, leaf shape, imaging conditions, and background characteristics encountered during model training, validation, and independent testing.

The partitioning boundary was defined at the individual-image level. The assembled dataset did not provide consistent identifiers linking all images to specific plants, leaves, plots, fields, geographic locations, or imaging sessions. Consequently, group-wise partitioning at the plant, leaf, or field level could not be implemented across the complete dataset. The split ensured that no identical image occurred in more than one subset, but it cannot establish that visually related images were obtained from different biological specimens or collection settings. This limitation is explicitly considered when interpreting the reported classification performance and the generalizability of the proposed framework.

The final dataset consisted of the following disease classes: bacterial blight (n = 88), brown spot (n = 81), Cercospora leaf blight (n = 5), Ferrugem (n = 65), mosaic virus (n = 22), powdery mildew (n = 137), Septoria (n = 21), southern blight (n = 62), Sudden death syndrome (n = 112), and yellow mosaic (n = 110), for a total of 703 images. The dataset was highly imbalanced, with class sizes ranging from 5 to 137 images (largest-to-smallest class

ratio = 27.4:1). Powdery mildew represented the largest class, whereas Cercospora leaf blight, mosaic virus, and Septoria were substantially underrepresented. This imbalance reflects realistic agricultural conditions in which fungal diseases are generally more prevalent than bacterial and viral diseases (Figure 2; Table 1).

Table 1. Class-wise distribution of the soybean disease dataset showing the total number of images and their allocation into the training, validation, and held-out test sets using the deterministic class-aware splitting procedure.

Disease Class	Total	Train	Validation	Test
Bacterial blight	88	62	13	13
Brown spot	81	57	12	12
Cercospora leaf blight	5	3	1	1
Ferrugem	65	45	10	10
Mosaic virus	22	16	3	3
Powdery mildew	137	95	21	21
Septoria	21	15	3	3
Southern blight	62	44	9	9
Sudden death syndrome	112	78	17	17
Yellow mosaic	110	78	16	16

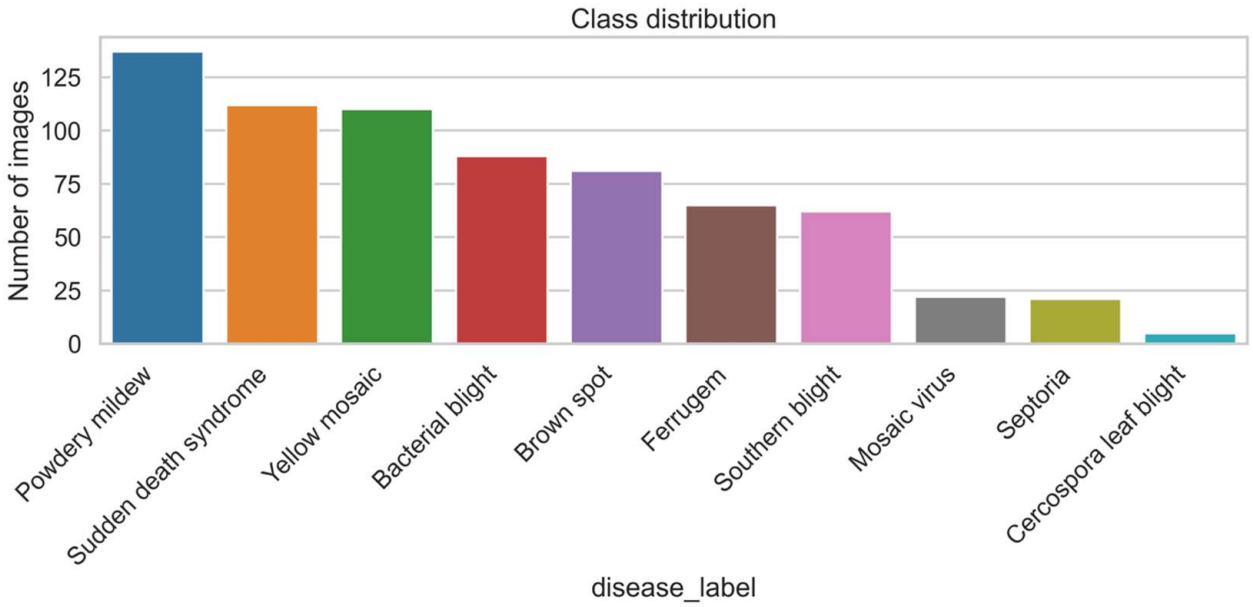

Figure 2. Distribution of soybean disease classes in the dataset. The dataset is highly imbalanced, with class sizes ranging from 5 images for Cercospora leaf blight to 137 images for Powdery mildew, corresponding to a largest-to-smallest class ratio of 27.4:1.

At the pathogen level, fungal diseases accounted for the majority of images, followed by viral and bacterial diseases (Figure 3). Consequently, the dataset captures biologically meaningful variation while presenting a realistic class imbalance challenge for deep learning-based disease recognition.

All images underwent standardized preprocessing before model training. Images were resized to a fixed resolution of 224 × 224 pixels and normalized using the ImageNet-1K V1 mean and standard deviation associated with the pretrained EfficientNet-B0 weights provided by torvision v0.26.0 (PyTorch v2.11.0). Data augmentation was applied exclusively to the training subset and included random rotations, horizontal and vertical flips, random resized cropping, color jitter, and affine transformations. These augmentations simulate

natural variability in field conditions, including differences in illumination, viewing angle, leaf orientation, and scale, thereby improving model robustness and reducing overfitting.

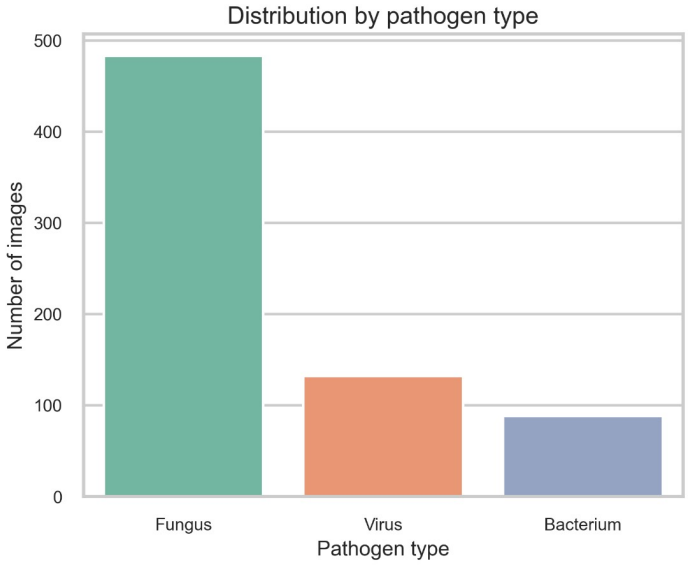

Figure 3. Distribution of images grouped by pathogen type. Fungal diseases dominate the dataset, followed by viral and bacterial classes.

To ensure reproducibility while maintaining representation of all disease classes, images were partitioned using a deterministic class-aware allocation with a fixed random seed (42). The final partition consisted of 493 training images, 105 validation images, and 105 independent test images. For each disease class containing at least three images, the allocation procedure ensured that the training, validation, and test subsets each contained at least one representative image. Consequently, the rarest class, Cercospora leaf blight (n = 5), was allocated three training images, one validation image, and one test image. No image appeared in more than one subset.

To reduce image-level leakage, the training, validation, and test subsets were mutually exclusive, and filenames and available metadata were inspected for exact or apparent duplicate images. Nevertheless, because specimen- and field-level identifiers were unavailable for the dataset, biological-source independence among the three partitions could not be verified.

2.2. Model Architecture and Training

An EfficientNet-B0 convolutional neural network (CNN) was employed for soybean disease classification because of its favorable balance between predictive performance and computational efficiency. The network was initialized using ImageNet-pretrained weights and adapted for the soybean disease dataset by replacing the original classification layer with a fully connected layer containing 10 output neurons, corresponding to the ten disease classes. The EfficientNet-B0 architecture combines depthwise separable convolutions with compound scaling of network depth, width, and input resolution, enabling efficient hierarchical feature extraction while maintaining a relatively small model size [1,2].

Input images were resized to 224 × 224 pixels and normalized using the ImageNet mean and standard deviation. During training, only the training subset was augmented using random resized cropping, horizontal and vertical flipping, random rotations, affine transformations, and color jitter to improve robustness to variations in illumination, viewing angle, leaf orientation, and scale. The validation and independent

test sets were evaluated without data augmentation using identical preprocessing and normalization procedures.

Model training employed a weighted categorical cross-entropy loss to reduce the influence of class imbalance by assigning greater penalty to errors involving underrepresented disease classes. Training was performed using the AdamW optimizer implemented in PyTorch v2.11.0 (PyTorch Foundation, San Francisco, CA, USA), with an initial learning rate of 1×10^{-4}, a weight decay of 1×10^{-4}, and a batch size of 16. The classification head retained the default EfficientNet-B0 dropout probability of 0.2. Learning-rate adaptation was performed using the ReduceLROnPlateau learning-rate scheduler implemented in PyTorch v2.11.0 (PyTorch Foundation, San Francisco, CA, USA), operating on the validation weighted F1-score with a reduction factor of 0.5 and a patience of two epochs. Training proceeded for a maximum of 20 epochs, with early stopping triggered after five consecutive epochs without improvement in the validation-weighted F1-score. All experiments were performed using a fixed random seed (42) to ensure reproducibility.

To mitigate the effects of the highly imbalanced class distribution, two complementary strategies were implemented during training. First, class-specific weights were incorporated into the cross-entropy loss function. Second, a weighted random sampler increased the probability that minority-class images were selected during mini-batch construction. These strategies improved the contribution of underrepresented disease classes during optimization while preserving the original class distribution of the dataset.

Training dynamics were monitored using the training and validation loss, classification accuracy, weighted F1-score, and macro F1-score. The model achieving the highest validation weighted F1-score was retained for subsequent evaluation on the independent test set and for all downstream analyses, including Gradient-weighted Class Activation Mapping (Grad-CAM) visualization implemented in PyTorch v2.11.0 (PyTorch Foundation, San Francisco, CA, USA), feature embedding extraction, phenotypic similarity analysis, and phylogenetic comparisons.

To further assess model performance within the available dataset, stratified five-fold cross-validation was performed independently of the fixed train–validation–test partition. For each fold, the EfficientNet-B0 model was reinitialized with ImageNet-pretrained weights and trained using the same preprocessing, augmentation, weighted loss, weighted sampling, optimization strategy, learning-rate scheduler, and early-stopping criteria described above. Classification performance was evaluated using accuracy, weighted F1-score, and macro F1-score, and the results are reported as the mean $\pm$ standard deviation across the five validation folds. The fixed train–validation–test model was used exclusively for the independent test evaluation and all downstream biological analyses.

2.3. Model Evaluation

Model performance was evaluated using an image-level held-out test set that was not used during model training or hyperparameter optimization. Overall classification performance was quantified using accuracy, precision, recall, weighted F1-score, and macro F1-score. Accuracy provides an overall measure of correct classifications, whereas the weighted F1-score accounts for class frequency by weighting each class according to its support. The macro F1-score assigns equal importance to every disease class and therefore provides a complementary measure of performance under highly imbalanced class distributions.

To assess class-specific performance, precision, recall, F1-score, and support were calculated for each disease class using a classification report generated from the independent test predictions. A confusion matrix was constructed to visualize correct classifications and

misclassification patterns among disease classes, thereby identifying diseases that were more difficult to distinguish.

Model robustness was additionally assessed using stratified five-fold cross-validation. Classification performance for each validation fold was quantified using accuracy, weighted F1-score, and macro F1-score, and the results are reported as the mean ± standard deviation across all five folds. The cross-validation analysis was used to estimate performance within the available dataset, whereas the independently held-out test set provided the final unbiased evaluation of the trained model.

To ensure complete consistency between numerical and graphical performance summaries, the classification report and confusion matrix were generated from the same prediction set, and the class support values reported in the classification metrics were verified to match the corresponding row totals of the confusion matrix.

2.4. Model Interpretability

To improve the interpretability of the deep learning model, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to visualize the image regions that contributed most strongly to individual disease predictions. Grad-CAM generates class-specific activation heatmaps by computing the gradients of the predicted class with respect to the final convolutional feature maps, thereby identifying the spatial regions that most influenced the model's decision.

Grad-CAM analyses were performed using the best-performing EfficientNet-B0 model selected according to the highest validation weighted F1-score. Representative images from each soybean disease class were analyzed, and the resulting heatmaps were superimposed onto the original leaf images to facilitate biological interpretation. This approach enabled qualitative assessment of whether the model focused on diagnostically relevant disease symptoms, including lesions, chlorosis, necrosis, discoloration, and other characteristic pathological features, rather than unrelated background regions.

The Grad-CAM visualizations served as an additional validation of model behavior by providing insight into the decision-making process of the CNN. Consistent localization of disease-associated symptoms supports the biological plausibility of the learned feature representations and complements the quantitative evaluation metrics reported for the independent test set.

2.5. Feature Representation and Embedding Visualization

To investigate the visual feature representations learned by the CNN, high-dimensional feature embeddings were extracted from the penultimate layer of the best-performing EfficientNet-B0 model. These embeddings provide compact numerical representations of soybean leaf phenotypes and capture the discriminative visual characteristics learned during disease classification.

To visualize the organization of the learned feature space, t-distributed stochastic neighbor embedding (t-SNE) was applied to project the high-dimensional embeddings into two dimensions. The resulting visualization was used to qualitatively assess disease-class separability, clustering patterns, and relationships among diseases caused by fungal, bacterial, and viral pathogens. Well-separated clusters indicate that the model learned distinct phenotypic representations, whereas overlapping clusters suggest similarities in visual symptomatology.

Beyond visualization, the extracted feature embeddings served as the basis for all subsequent phenotypic analyses. Pairwise cosine similarity and Euclidean distance matrices were computed to quantify visual similarity among disease classes, and these matrices were subsequently used for hierarchical clustering, network analysis, and comparisons

with pathogen phylogenetic relationships. This embedding-based approach enabled quantitative characterization of phenotypic similarity without requiring manually annotated lesion features.

Because manually annotated lesion masks were unavailable, interpretability and embedding analyses were used to evaluate the learned phenotypic structure rather than to perform pixel-level lesion localization. Consequently, the analyses focused on the biological relationships captured by the learned feature representations rather than on segmentation-based measurements of disease symptoms.

2.6. Phenotypic Similarity Analysis

To quantify visual relationships among soybean diseases, feature embeddings extracted from the penultimate layer of the trained EfficientNet-B0 model were used to compute pairwise phenotypic similarity. For each disease class, embeddings from all representative images were aggregated to obtain a class-level feature representation. Pairwise cosine similarity was then calculated between all disease-class representations, producing a phenotypic similarity matrix in which larger values indicate greater similarity in the learned visual feature space.

To provide a complementary measure of phenotypic divergence, pairwise Euclidean distances were also calculated from the class-level embeddings. Whereas cosine similarity emphasizes similarity in feature orientation, Euclidean distance reflects absolute differences in feature representations. Together, these complementary metrics provide a more comprehensive characterization of phenotypic relationships among soybean diseases.

The resulting similarity and distance matrices were visualized as clustered heatmaps to identify groups of diseases exhibiting similar visual characteristics. Hierarchical clustering was performed using agglomerative clustering with average linkage, allowing diseases with similar learned phenotypes to be grouped objectively. These analyses enabled qualitative assessment of phenotypic organization while providing quantitative measures of visual similarity among disease classes.

To further examine the phenotypic structure, a network representation was constructed in which nodes corresponded to disease classes and edges connected disease pairs exceeding a predefined cosine similarity threshold. The threshold was selected empirically to emphasize stronger phenotypic relationships while avoiding excessive network density that could obscure overall structure. Edge weights were proportional to cosine similarity values, allowing stronger visual relationships to be represented by stronger network connections. The resulting network was intended as an exploratory visualization of learned phenotypic relationships rather than a formal graph-theoretical analysis.

2.7. Taxonomic and Sequence-Based Analysis

To provide biological context for the learned phenotypic relationships, evolutionary relationships among soybean disease pathogens were characterized using both taxonomy-based and sequence-based approaches. Taxonomic distance matrices were constructed from the hierarchical biological classification of each pathogen, with pairwise distances reflecting differences across successive taxonomic ranks. Greater distances therefore indicate greater evolutionary divergence between pathogens.

To enable direct comparison between learned phenotypic representations and pathogen evolution, pairwise genetic distance matrices derived from the phylogenetic analysis were compared with the CNN-derived phenotypic distance matrices. In addition, taxonomy-derived distance matrices were generated for all pathogen classes to provide a broader biological reference for phenotypic comparisons (Figures S1 and S2).

Sequence-based phylogenetic analyses were restricted to the fungal subset of the dataset because fungi represented the only pathogen group with multiple evolutionarily distinct disease classes, enabling meaningful comparative analyses. Representative nucleotide sequences were obtained for each fungal pathogen and aligned using a multiple sequence alignment procedure prior to phylogenetic reconstruction. Evolutionary relationships were then inferred using the Neighbor-Joining method to generate a phylogenetic tree representing the genetic relationships among the fungal pathogens (Figure S3). The bacterial and viral disease classes were not included in these sequence-based comparisons because they were insufficiently represented within the dataset for informative within-group phylogenetic analysis.

2.8. Comparative Clustering and Correlation Analysis

To compare the relationships captured by the learned phenotypic representations with pathogen evolutionary relationships, hierarchical clustering was performed independently on the CNN-derived phenotypic distance matrix and the sequence-based genetic distance matrix. Agglomerative hierarchical clustering with average linkage was used to generate dendrograms, allowing qualitative comparison of disease groupings inferred from visual phenotypes and pathogen phylogeny (Figures S4 and S5).

For fungal pathogens, quantitative agreement between phenotypic and genetic distances was evaluated using the Mantel test, which measures the correlation between two distance matrices while accounting for their non-independence. Phenotypic distance was calculated as $1 -$ cosine similarity between the class-level embedding representations, whereas genetic distance was derived from the fungal phylogenetic analysis. Statistical significance was assessed using permutation testing. The Mantel test was selected because it is specifically designed for comparing pairwise distance matrices and is widely used in ecological and evolutionary studies.

To further characterize the relationship between learned phenotypic similarity and pathogen evolution, Pearson and Spearman correlation coefficients were also calculated using the corresponding pairwise phenotypic and genetic distances. Pearson correlation quantified linear associations between the two distance measures, whereas Spearman correlation assessed monotonic relationships that were robust to non-normality and nonlinear trends. Scatter plots with fitted regression lines were generated to visualize these relationships (Figure S6).

Together, the hierarchical clustering, Mantel analysis, and correlation analyses provided complementary qualitative and quantitative assessments of the extent to which CNN-derived phenotypic similarity reflects the evolutionary relationships among fungal soybean pathogens.

2.9. Computational Environment and Software

All analyses were performed using the Python programming language (v3.11) [40] within a Jupyter Notebook environment (v7.5.5) [41]. Image processing, numerical computations, and data management were conducted using the scientific Python ecosystem, including NumPy (v1.26.4) [42], pandas (v2.3.3) [43], matplotlib (v3.10.8) [44], and seaborn (v0.13.2) [45]. Deep learning model development, training, and inference were implemented using PyTorch (v2.11.0) [46] and the TorchVision library (v0.26.0) [47] for pretrained model architectures and image transformations. Statistical analyses, distance calculations, clustering, and dimensionality reduction were performed using SciPy (v1.15.2) [48] and scikit-learn (1.7.2) [49], while phylogenetic and sequence analyses were conducted using Biopython (v1.86) [50].

The complete analytical workflow—including image preprocessing, model training, feature embedding extraction, Grad-CAM visualization, phenotypic similarity analysis, clustering, network construction, phylogenetic analysis, and statistical comparisons—was implemented entirely in Python to ensure methodological consistency and reproducibility. Model training, evaluation, and downstream analyses were executed within a unified Jupyter Notebook workflow, enabling complete regeneration of all figures, tables, and quantitative results from the original dataset.

To facilitate reproducibility, the computational workflow employed fixed random seeds, deterministic dataset partitioning, and automated recording of key software versions and execution environment information. These measures ensured that all reported results can be reproduced consistently using the same analytical pipeline.

3. Results

3.1. Model Training Dynamics and Performance

The EfficientNet-B0 model demonstrated stable and consistent convergence during training (Figure 4). Both the training and validation cross-entropy losses decreased rapidly during the initial epochs and progressively stabilized at low values, indicating effective optimization without evidence of substantial overfitting (Figure 4a). Similarly, classification accuracy, weighted F1-score, and macro F1-score increased steadily throughout training before reaching stable plateaus after approximately 10 epochs (Figure 4b,c). The close agreement between the training and validation curves suggests that the selected regularization, data augmentation, weighted loss, and early stopping strategies promoted good generalization despite the highly imbalanced dataset.

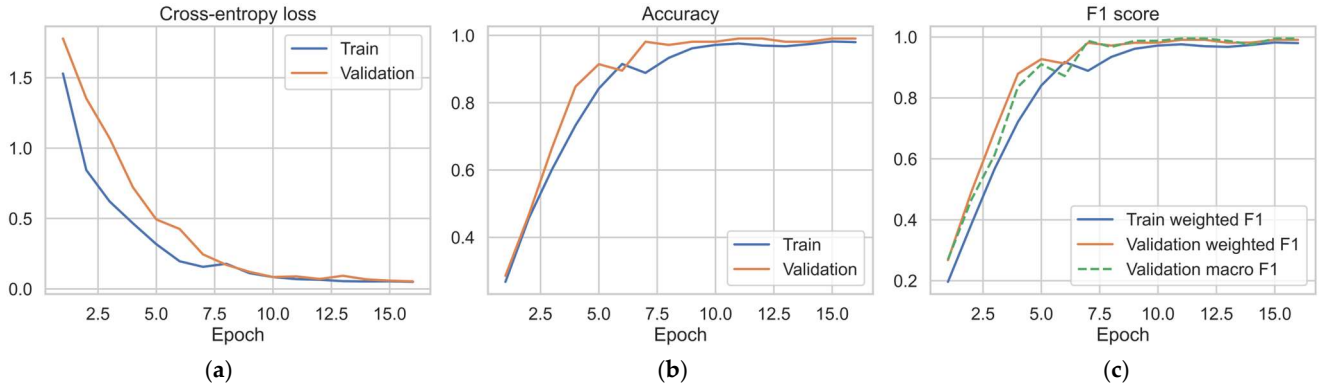

Figure 4. Training and validation performance over epochs. (**a**) Cross-entropy loss decreases steadily for both training and validation sets, indicating stable convergence. (**b**) Classification accuracy improves and plateaus near convergence. (**c**) Weighted F1-score follows a similar trend, demonstrating balanced performance across classes.

To further assess model robustness, stratified five-fold cross-validation was performed using identical preprocessing, augmentation, and optimization procedures for each fold. The model achieved a mean classification accuracy of 98.72% ± 1.17%, a mean weighted F1-score of 98.74% ± 1.16%, and a mean macro F1-score of 98.47% ± 1.74% across the five validation folds. The relatively small standard deviations indicate consistent performance across different training–validation partitions, demonstrating that the learned feature representations generalized well despite substantial differences in disease class frequencies.

The final model selected according to the highest validation weighted F1-score was subsequently evaluated on the image-level held-out test set, which was not used during model training or hyperparameter optimization. The model achieved a test accuracy of 96.19%, a weighted F1-score of 96.28%, a macro F1-score of 95.86%, and a cross-entropy loss

of 0.1392. As expected, performance on the independent test set was modestly lower than the mean cross-validation performance, reflecting the more stringent evaluation provided by a completely unseen dataset. Nevertheless, the consistently high classification metrics indicate that the model learned discriminative visual features associated with soybean diseases and achieved strong internal test performance within the available image dataset.

3.2. Classification Accuracy and Error Analysis

Evaluation of the independently held-out test set demonstrated strong classification performance across the ten soybean disease classes (Figure 5; Table 2). The confusion matrix is dominated by diagonal entries, indicating that the majority of test images were assigned to their correct disease classes. Seven of the ten disease classes—bacterial blight, brown spot, Cercospora leaf blight, Ferrugem, mosaic virus, southern blight, and Sudden death syndrome—were classified without error on the independent test set.

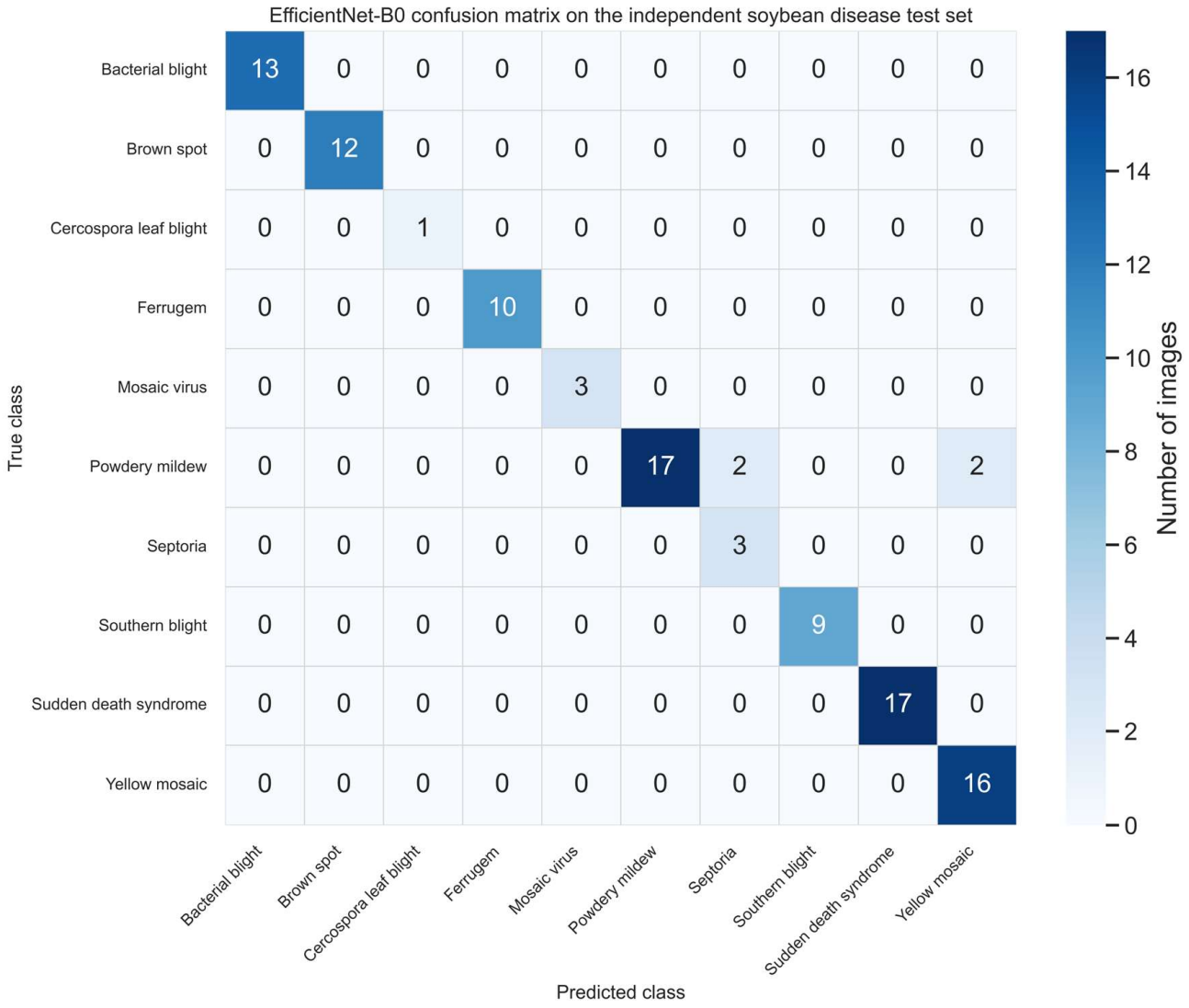

Figure 5. Raw confusion matrix for the image-level held-out test set. Most errors originated from Powdery mildew, with two images classified as Septoria and two classified as Yellow mosaic. All matrix values represent image counts, and the row totals correspond to the class-support values reported in Table 2.

Table 2. Class-specific precision, recall, F1-score, and support obtained from predictions on the image-level held-out test set. Values are reported separately for each soybean disease class to permit transparent evaluation under the highly imbalanced class distribution.

Disease Class	Precision	Recall	F1-Score	Support
Bacterial blight	1.000	1.000	1.000	13
Brown spot	1.000	1.000	1.000	12
Cercospora leaf blight	1.000	1.000	1.000	1
Ferrugem	1.000	1.000	1.000	10
Mosaic virus	1.000	1.000	1.000	3
Powdery mildew	1.000	0.810	0.895	21
Septoria	0.600	1.000	0.750	3
Southern blight	1.000	1.000	1.000	9
Sudden death syndrome	1.000	1.000	1.000	17
Yellow mosaic	0.889	1.000	0.941	16

Most misclassifications involved Powdery mildew, for which 17 of 21 test images were correctly classified. Two Powdery mildew images were incorrectly predicted as Septoria, and two were predicted as Yellow mosaic, resulting in a recall of 80.95% for this class. Although Septoria achieved perfect recall (3/3 correctly identified), one additional image from another class was incorrectly assigned to Septoria, reducing its precision to 60.0%. Similarly, Yellow mosaic correctly classified all 16 test images (100% recall), but two Powdery mildew images were incorrectly assigned to this class, resulting in a precision of 88.9%. These results indicate that the few remaining classification errors primarily occurred among diseases exhibiting partially overlapping visual symptom characteristics.

Performance estimates for the rarest disease classes should nevertheless be interpreted cautiously. Although the single independent test image of Cercospora leaf blight was classified correctly, this result is based on only one sample and therefore does not provide a reliable estimate of class-specific generalization performance. Likewise, Mosaic virus and Septoria were represented by only three test images each. Consequently, cross-validation provides a complementary estimate of model stability across image-level partitions, but it does not eliminate the uncertainty associated with classes represented by very few images. The held-out test set provides an additional internal evaluation based on images excluded from model training and hyperparameter selection.

Overall, the model achieved an independent test accuracy of 96.19%, with weighted and macro F1-scores of 96.28% and 95.86%, respectively. These findings indicate strong overall discrimination within the available dataset; however, the aggregate performance metrics should be interpreted together with the class-specific results and the very limited support available for several minority classes.

To avoid masking minority-class behavior through aggregate metrics, unweighted precision, recall, and F1-score are reported separately for every disease class in Table 2, together with the corresponding raw sample counts. These class-specific values were generated from the same held-out predictions used to construct the confusion matrix in Figure 5.

3.3. Model Interpretability

Grad-CAM visualization provided qualitative insight into the image regions that contributed most strongly to the CNN predictions (Figure 6). Across the representative examples from all ten soybean disease classes, the activation maps consistently emphasized symptomatic leaf tissue rather than background regions, indicating that the model primarily based its predictions on biologically relevant visual features. Regions receiving the strongest

attention included necrotic lesions, chlorotic tissue, discoloration, and other characteristic disease symptoms.

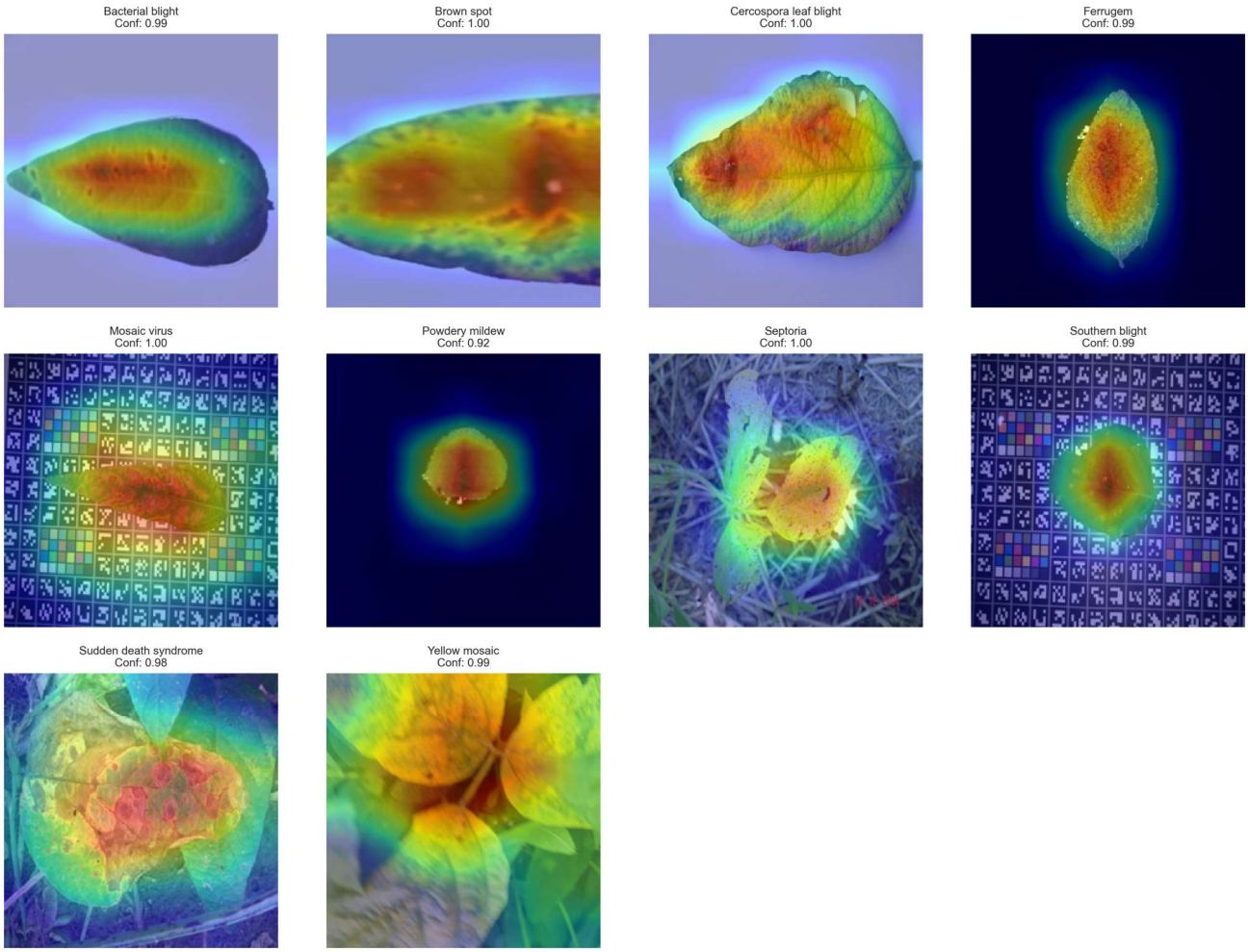

Figure 6. Grad-CAM visualizations highlighting salient regions contributing to model predictions. The model focuses on disease-specific lesion patterns and leaf regions, supporting biologically meaningful feature learning.

The spatial distribution of model attention varied among disease classes in accordance with their visual symptomatology. For diseases characterized by localized lesions, such as Ferrugem and Cercospora leaf blight, the activation maps were concentrated around the affected regions. In contrast, diseases exhibiting more diffuse symptom patterns, including Sudden death syndrome, Yellow mosaic, and Powdery mildew, produced broader activation patterns spanning larger portions of the leaf surface. These differences indicate that the CNN adapted its attention to disease-specific symptom morphology rather than relying on a fixed spatial pattern.

The Grad-CAM visualizations also demonstrated that model confidence was generally associated with biologically plausible regions of the images. Most representative examples were classified with high confidence, and the highlighted regions corresponded closely with visible disease symptoms. These observations support the interpretation that the learned feature representations capture diagnostically meaningful phenotypic characteristics that contribute to accurate disease classification.

Because manually annotated lesion masks were not available, Grad-CAM was used as a qualitative interpretability tool rather than a quantitative assessment of lesion localization accuracy. Consequently, the visualizations provide evidence that the model attended

to disease-associated features but should not be interpreted as precise delineations of lesion boundaries.

3.4. Feature Representation and Embedding Structure

The high-dimensional feature embeddings extracted from the penultimate layer of the EfficientNet-B0 model were visualized using t-distributed stochastic neighbor embedding (t-SNE) (Figure 7). The resulting two-dimensional projection revealed well-defined clusters corresponding to the ten soybean disease classes, indicating that the CNN learned highly discriminative feature representations capable of separating diseases with distinct visual characteristics.

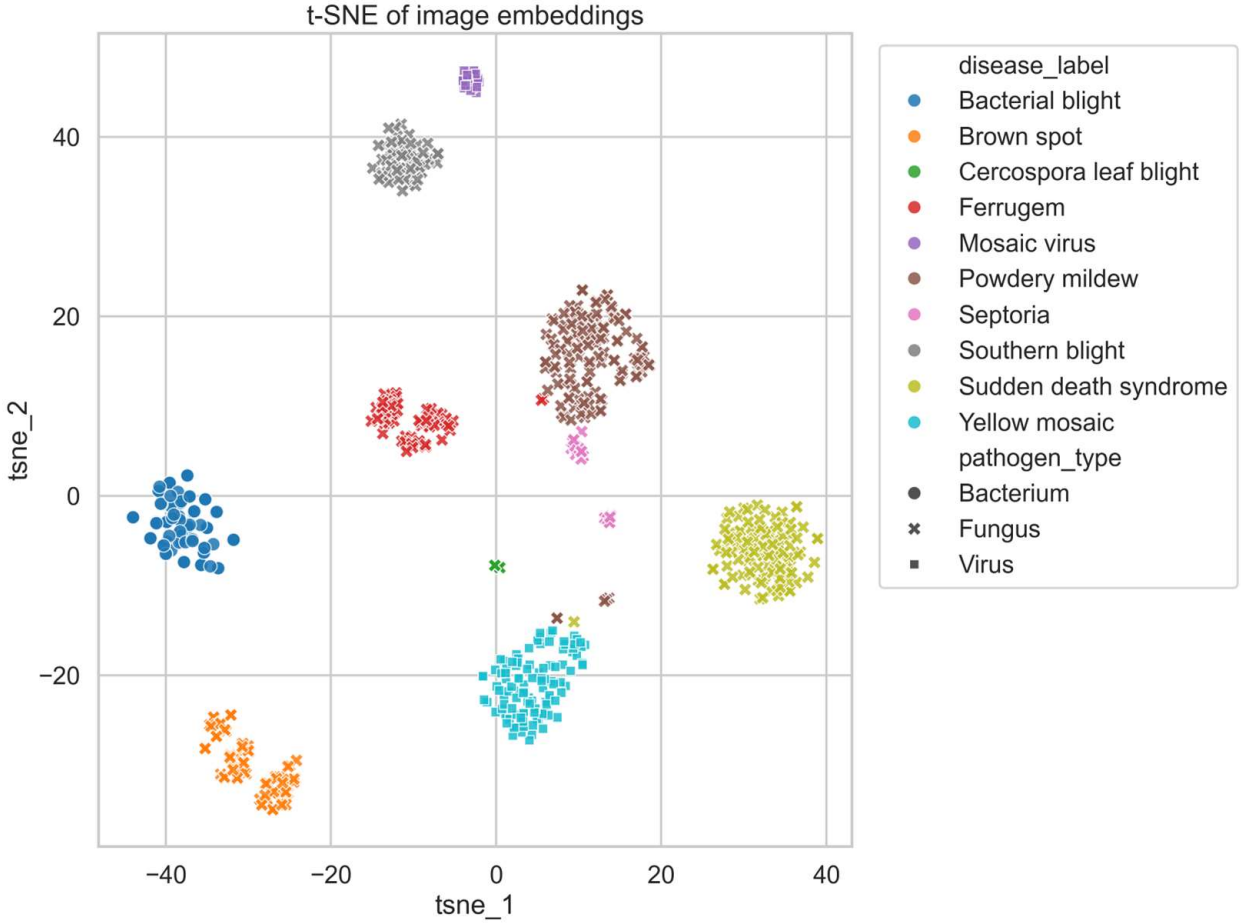

Figure 7. t-SNE projection of learned image embeddings. Distinct clustering of disease classes indicates strong separability in feature space, with partial grouping by pathogen type (fungal, bacterial, viral).

Most disease classes formed compact clusters with minimal overlap, demonstrating that images belonging to the same disease were consistently mapped to similar regions of the learned feature space. The bacterial blight, brown spot, Sudden death syndrome, and Yellow mosaic classes exhibited particularly well-separated clusters, whereas limited proximity was observed among some fungal diseases, reflecting similarities in certain visual symptoms rather than widespread overlap between classes.

Although pathogen type was represented by different marker symbols in the visualization, the embedding space was organized primarily according to disease phenotype rather than broad pathogen categories. Fungal diseases occupied multiple regions of the feature space, indicating substantial phenotypic diversity among fungal pathogens, while bacterial and viral diseases formed distinct disease-specific clusters. These findings suggest that

the CNN primarily learned discriminative visual characteristics associated with disease symptoms instead of grouping images solely according to pathogen taxonomy.

Overall, the embedding structure demonstrates that the learned feature representations effectively capture disease-specific phenotypic information while preserving biologically meaningful relationships among visually similar diseases. These embeddings formed the basis for the subsequent phenotypic similarity, clustering, and comparative phylogenetic analyses.

3.5. Phenotypic Similarity and Network Structure

Pairwise phenotypic similarity was quantified using cosine similarity between class-level feature embeddings extracted from the trained EfficientNet-B0 model (Figure 8; Supplementary Tables S1 and S2). The resulting similarity matrix demonstrates that most disease pairs exhibit low cosine similarity values, indicating that the CNN learned distinct visual representations for the majority of soybean diseases. Nevertheless, several disease pairs displayed moderate similarity, suggesting partial overlap in their visual symptom characteristics. The strongest relationships were observed between Mosaic virus and Southern blight (cosine similarity = 0.47), Bacterial blight and Brown spot (0.32), and Ferrugem and Septoria (0.29).

Figure 8. Pairwise phenotypic similarity between disease classes computed using cosine similarity of image embeddings. Higher values indicate visually similar disease manifestations.

The phenotypic similarity network provides an intuitive visualization of these relationships (Figure 9). Using an empirically selected similarity threshold, the resulting network formed a sparse graph in which only disease pairs with relatively strong visual similarity were connected. Most diseases exhibited only a limited number of connections, reflecting the overall distinctiveness of the learned phenotypic representations. Several fungal diseases occupied central positions within the network because they shared moderate similarity with multiple disease classes, whereas bacterial and viral diseases generally formed fewer connections.

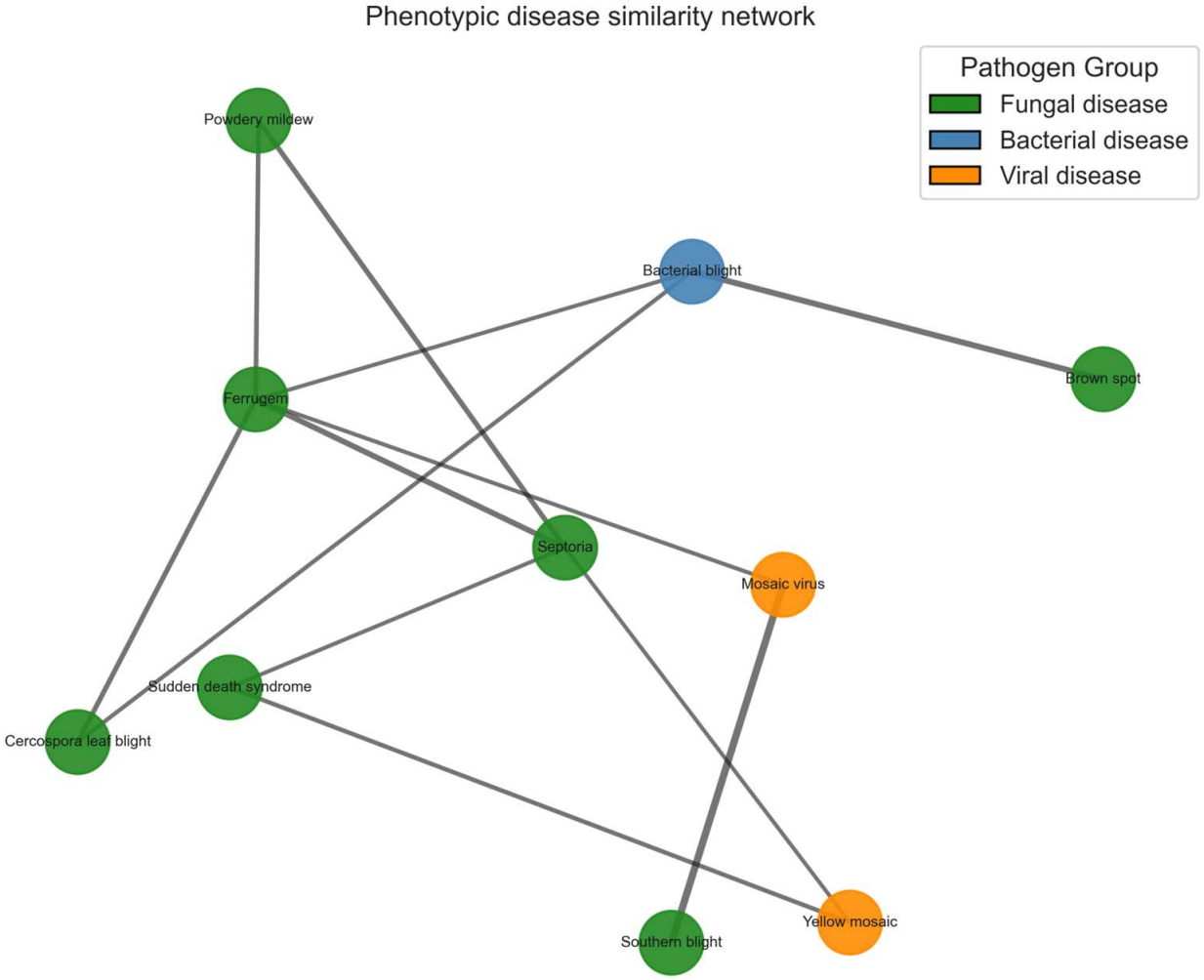

Figure 9. Network representation of disease similarity based on phenotypic features. Nodes represent diseases and edges denote high similarity relationships, revealing clusters of visually related conditions.

Aggregation of embeddings at the pathogen level revealed a similar overall pattern (Figure S7). Although moderate similarities were observed among a small number of pathogen pairs, most pathogen-level relationships remained weak, indicating substantial phenotypic diversity even among related pathogens. These observations suggest that visual similarity is influenced not only by pathogen taxonomy but also by the specific manifestation of disease symptoms on soybean leaves.

Overall, the phenotypic similarity analyses demonstrate that the CNN learned biologically meaningful feature representations that distinguish most soybean diseases while preserving relationships among diseases with partially overlapping visual characteristics.

These learned phenotypic relationships provide the basis for subsequent comparisons with pathogen taxonomy and evolutionary history.

3.6. Sequence-Based Phylogenetic Relationships

Sequence-based genetic distances among the fungal pathogens, inferred from 28S rRNA sequences, revealed clear patterns of evolutionary divergence (Figure S1). Closely related taxa, including *Septoria glycines* and *Cercospora kikuchii*, exhibited relatively small genetic distances, whereas more distantly related pathogens, such as *Sclerotium rolfsii*, showed substantially greater divergence from the remaining fungal species. Overall, the genetic distance matrix was consistent with established taxonomic relationships among the fungal pathogens included in this study.

Mapping these sequence-derived distances to their corresponding soybean diseases produced a disease-level genetic distance matrix that preserved the underlying evolutionary relationships (Figure S2). Diseases caused by closely related fungal pathogens, including Septoria and Cercospora leaf blight, exhibited relatively small genetic distances, whereas diseases associated with more evolutionarily distant pathogens were clearly separated. This disease-level representation enabled direct comparison between pathogen phylogeny and the CNN-derived phenotypic relationships.

The Neighbor-Joining phylogenetic tree further illustrated the evolutionary relationships among the fungal pathogens (Figure S3). *Septoria glycines* and *Cercospora kikuchii* formed a closely related lineage, whereas *Phakopsora pachyrhizi* occupied a distinct branch. *Fusarium virguliforme* and *Sclerotium rolfsii* were positioned on more divergent lineages, reflecting their greater evolutionary separation from the remaining fungal pathogens. The phylogenetic tree therefore provided the biological framework for subsequent comparisons between genetic relatedness and CNN-derived phenotypic similarity.

3.7. Comparative Clustering of Phenotypic and Genetic Relationships

Hierarchical clustering of the sequence-derived genetic distance matrix revealed distinct evolutionary groupings among the fungal pathogens (Figure S4). Closely related species, including *Septoria glycines* and *Cercospora kikuchii*, clustered together, whereas *Fusarium virguliforme* and *Sclerotium rolfsii* occupied more divergent branches, consistent with the phylogenetic tree and genetic distance analyses.

Hierarchical clustering of the CNN-derived phenotypic distance matrix produced a different organizational pattern (Figure S5). Although several disease pairs exhibited similar relationships in both dendrograms, the overall clustering structure differed from that inferred from sequence data. Diseases with comparable visual symptoms were frequently grouped together despite being evolutionarily distant, whereas some closely related fungal pathogens were separated because of differences in their observable disease phenotypes.

Comparison of the phenotypic and genetic dendrograms therefore revealed only partial concordance between visual similarity and pathogen evolutionary relationships. These findings suggest that the CNN-derived feature embeddings capture biologically meaningful aspects of disease symptom expression that are related to, but not fully determined by, pathogen phylogeny. Consequently, phenotypic similarity reflects both shared evolutionary history and convergent or divergent symptom development among soybean diseases. The degree of agreement between the phenotypic and genetic distance matrices was evaluated quantitatively using matrix correlation analyses, the results of which are presented in the following section.

3.8. Pathogen-Level Phenotypic Similarity

To facilitate comparison between pathogen taxonomy and CNN-derived phenotypic relationships, pathogen-level similarity matrices were generated from both biological

classification (Figure 10) and the learned feature embeddings (Figure S7). The taxonomy-based matrix reflects the expected evolutionary separation among fungal, bacterial, and viral pathogens, with closely related taxa exhibiting smaller taxonomic distances and more distantly related pathogens occupying distinct groups.

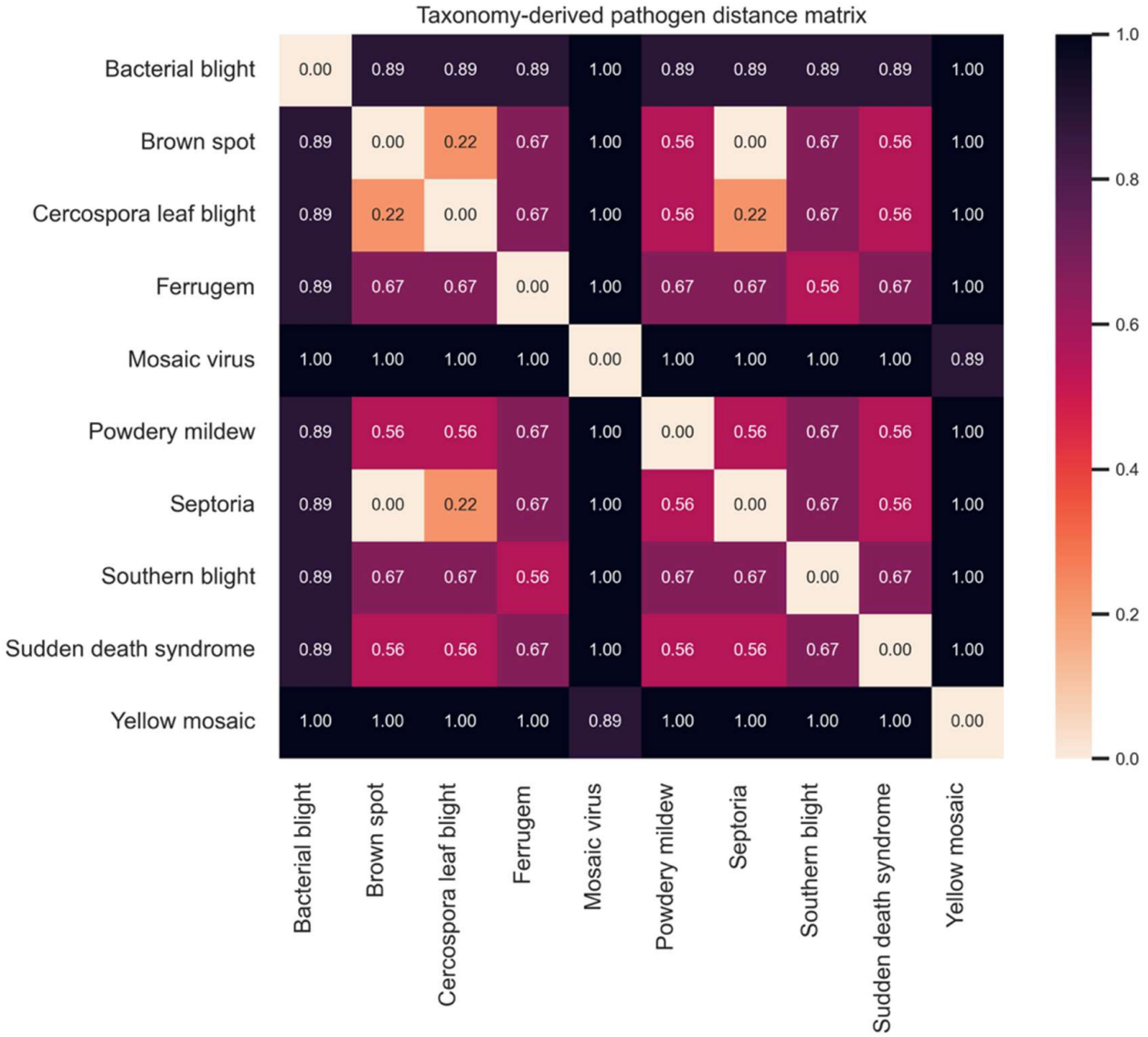

Figure 10. Taxonomy-based distance matrix between pathogens. Distances reflect biological relatedness derived from taxonomic hierarchy, with higher values indicating greater evolutionary divergence.

In contrast, the pathogen-level phenotypic similarity matrix derived from the CNN embeddings revealed a more heterogeneous pattern (Figure S7). Most pathogen pairs exhibited low to moderate cosine similarity, indicating that the learned visual representations remained largely distinct across diseases. Nevertheless, several pathogen pairs displayed moderate similarity, reflecting shared visual symptom characteristics despite differences in taxonomic relationships.

Comparison of the taxonomy-based and phenotype-based matrices indicates that pathogen classification alone does not fully explain the observed phenotypic relationships. While some visually similar diseases were caused by related pathogens, others exhibited comparable symptom patterns despite greater evolutionary separation. These findings are consistent with the disease-level analyses and demonstrate that the CNN primar-

ily learned relationships based on observable disease phenotypes rather than pathogen taxonomy alone.

Overall, the pathogen-level analyses complement the disease-level similarity results by demonstrating that learned phenotypic relationships are only partially aligned with biological classification, providing additional evidence that symptom expression captures information beyond evolutionary relatedness.

3.9. Phenotype–Phylogeny Association

To quantitatively assess the relationship between learned phenotypic similarity and pathogen evolutionary relationships, a Mantel test was performed using the fungal pathogen distance matrices (Figure S6). The analysis revealed a significant positive association between the CNN-derived phenotypic distances and the sequence-based genetic distances (Mantel Spearman r = 0.3393, $p = 0.0050$), indicating that visually similar fungal diseases tended to be associated with more closely related pathogens. Although statistically significant, the magnitude of the correlation was moderate, suggesting that pathogen phylogeny explains only part of the observed phenotypic variation.

The relationship between phenotypic and genetic distances is illustrated in Figure S6. Although considerable variability remained among individual disease pairs, the overall trend indicates increasing phenotypic similarity among more closely related fungal pathogens. These findings demonstrate that the feature representations learned by the CNN capture a measurable phylogenetic signal while also reflecting additional sources of phenotypic variation.

The moderate strength of the association likely reflects the multifactorial nature of disease symptom development. In addition to pathogen evolutionary history, visible disease phenotypes are influenced by host responses, infection stage, environmental conditions, and pathogen-specific mechanisms of symptom expression. Furthermore, the phenotype–phylogeny comparison was intentionally performed on the fungal subset of the dataset because it comprised multiple evolutionarily distinct pathogens, enabling meaningful comparative analysis. The bacterial and viral subsets contained too few disease classes for comparable phylogenetic evaluation. In addition, the 28S rRNA marker provides only moderate phylogenetic resolution among fungal taxa.

Overall, these results indicate that CNN-derived phenotypic representations capture biologically meaningful information that is partially consistent with pathogen evolutionary relationships. Rather than serving as a direct surrogate for phylogeny, the learned embeddings provide complementary phenotypic information that reflects both evolutionary history and the complex biological processes underlying disease symptom development.

4. Discussion

4.1. Deep Learning Performance and Robustness in Disease Classification

The results demonstrate that the proposed deep learning framework achieved high classification performance for soybean disease recognition despite the pronounced class imbalance of the dataset [9]. The close agreement between the training and validation learning curves indicates stable convergence throughout the optimization process, suggesting that the adopted augmentation, weighted loss function, and regularization strategies effectively mitigated overfitting while promoting good generalization.

The model achieved a mean cross-validation accuracy of $98.72 \pm 1.17\%$ and a weighted F1-score of $98.74 \pm 1.16\%$, while the image-level held-out test set yielded an overall accuracy of 96.19%, a weighted F1-score of 96.28%, and a macro F1-score of 95.86%. These results indicate that the learned feature representations generalized well across independent data while maintaining consistent performance across repeated training partitions.

Although most disease classes were classified with high precision and recall, performance remained influenced by the severe class imbalance. Rare diseases, particularly Cercospora leaf blight, Septoria, and Mosaic virus, were represented by very few training and test images, limiting the reliability of class-specific performance estimates. Consequently, the perfect classification observed for some underrepresented classes should be interpreted cautiously and validated using larger and more balanced datasets. Nevertheless, the consistently strong overall performance demonstrates the ability of deep learning models to learn discriminative visual representations from complex plant disease symptoms [8,51]. Similar advantages of CNN-based approaches over conventional machine learning methods have been reported in agricultural image analysis, particularly for their ability to automatically extract hierarchical image features and achieve high predictive accuracy [11,13].

Despite these encouraging results, several limitations should be acknowledged. The dataset comprised only 703 images distributed across ten disease classes, with a largest-to-smallest class ratio of approximately 27:1, representing a highly imbalanced classification problem. Although weighted sampling, class-weighted loss, data augmentation, and regularization reduced the impact of this imbalance, larger and more balanced datasets will be essential for further improving model robustness, particularly for rare diseases, and for evaluating generalization across broader imaging conditions.

The cross-validation results further support the stability of the proposed framework. The low standard deviations observed across folds indicate limited sensitivity to the specific train–validation partition, suggesting that the network consistently learned reproducible phenotypic features rather than memorizing partition-specific characteristics. The modest reduction in performance on the independent test set relative to cross-validation is expected and reflects evaluation on previously unseen images, providing additional evidence of the model's generalization capability.

4.2. Biological Relevance of Learned Features

Beyond predictive performance, model interpretability provides important insight into the biological relevance of the learned feature representations. The Grad-CAM visualizations consistently highlighted disease-relevant image regions, including lesion boundaries, chlorotic tissue, necrotic areas, and other symptomatic leaf structures. These observations suggest that the model primarily relied on biologically meaningful phenotypic characteristics rather than irrelevant background information when generating its predictions [19,20].

The localization of the highlighted regions is consistent with established principles of plant pathology, in which visible disease symptoms reflect pathogen infection and host physiological responses [5,7]. Such biologically plausible attention patterns increase confidence that the model learned diagnostically relevant visual features and support its potential applicability to automated disease recognition in agricultural settings [52,53].

Nevertheless, the Grad-CAM analysis should be interpreted as qualitative evidence of model attention rather than a quantitative assessment of lesion localization. Because pixel-level or lesion-level annotations were not available, the highlighted regions cannot be considered a formal validation of localization accuracy. Future studies incorporating expert lesion annotations or segmentation masks would enable a more rigorous evaluation of the correspondence between model attention and disease symptoms.

4.3. Structure of the Phenotypic Feature Space

The t-SNE visualization of the learned embeddings revealed a structured and discriminative feature space, with clear separation among most disease classes. This organization indicates that the CNN learned compact feature representations capable of distinguish-

ing diverse soybean disease phenotypes while preserving within-class similarity. Such structured latent representations are characteristic of deep convolutional networks trained for image classification and reflect their ability to encode complex visual patterns into discriminative feature spaces [31,33].

Rather than clustering primarily according to pathogen taxonomy, the learned feature space was organized predominantly by visual symptom characteristics. Diseases exhibiting similar lesion morphology, chlorosis, discoloration, or tissue damage occupied relatively closer regions of the embedding space, whereas taxonomically related pathogens were not consistently grouped together. This observation is consistent with the objective of the CNN, which optimizes discrimination based on image appearance rather than evolutionary relationships.

From a plant pathology perspective, these findings suggest that CNN-derived embeddings capture biologically relevant phenotypic variation associated with disease expression. The resulting feature space therefore reflects both disease-specific visual signatures and symptom characteristics shared across different diseases, which may arise from common host physiological responses, similar pathological processes, or convergent symptom development rather than close evolutionary relatedness [38]. This interpretation is further supported by the moderate—but statistically significant—association observed between phenotypic similarity and fungal phylogeny, indicating that pathogen evolution contributes to, but does not solely determine, the organization of the learned feature space.

4.4. Phenotypic Similarity and Disease Relationships

The phenotypic similarity matrix and network analysis demonstrate that visual relationships among soybean diseases are structured but heterogeneous. While several disease pairs exhibited moderate phenotypic similarity, most pairwise relationships remained weak, indicating that the learned feature representations effectively distinguished the majority of disease classes despite the presence of shared visual characteristics.

The sparse connectivity observed in the phenotypic similarity network further supports this interpretation, with only a limited number of disease pairs exhibiting stronger associations. Rather than reflecting pathogen taxonomy, these localized relationships appear to be driven primarily by similarities in symptom expression, such as comparable lesion morphology, chlorosis, or tissue damage. This finding reinforces the interpretation that CNN-derived embeddings capture visual phenotypic characteristics more strongly than pathogen identity.

These observations are consistent with systems biology perspectives, in which complex biological traits are organized as partially connected networks rather than discrete or fully cohesive groups [37,53,54]. Nevertheless, phenotypic similarity alone cannot fully explain disease relationships because symptom expression is influenced by multiple interacting factors, including pathogen biology, host genotype, disease progression, and environmental conditions [28,55]. Consequently, visually similar diseases may arise from distinct pathogens, whereas closely related pathogens may produce substantially different symptom phenotypes under varying biological and environmental contexts.

4.5. Phylogenetic and Taxonomic Insights

The taxonomy-derived distance analysis provided a broad biological framework for interpreting relationships among the pathogens represented in the dataset, reflecting their hierarchical classification across fungal, bacterial, and viral groups. Complementing this analysis, the sequence-based phylogenetic reconstruction of the fungal subset recovered the expected evolutionary relationships among the fungal pathogens, providing an independent biological reference for comparison with the CNN-derived phenotypic representations.

The use of the 28S rRNA marker provided a well-established basis for inferring broad evolutionary relationships among fungi, consistent with its widespread application in fungal systematics [23]. However, because this marker is relatively conserved, it offers limited resolution for distinguishing closely related taxa [56]. Consequently, the inferred phylogeny is appropriate for evaluating general evolutionary relationships but may not capture more recent evolutionary divergence or species-level variation.

Together, the taxonomy-based and sequence-based analyses provide complementary biological context for interpreting the phenotypic similarity learned by the CNN. While taxonomy reflects hierarchical biological classification and phylogenetic analysis captures evolutionary relatedness within the fungal subset, neither alone fully explains the observed organization of the learned feature space, highlighting the complex relationship between pathogen evolution and disease symptom expression.

4.6. Relationship Between Phenotype and Phylogeny

A central finding of this study is that the relationship between phenotypic similarity and pathogen phylogeny is significant but incomplete. The Mantel analysis revealed a moderate positive association between CNN-derived phenotypic distances and sequence-based genetic distances, indicating that evolutionary relatedness contributes to the organization of the learned feature space. However, the moderate magnitude of the correlation demonstrates that phylogeny explains only part of the observed phenotypic variation, suggesting that the learned feature representations capture a detectable but not exclusive phylogenetic signal.

Several biological factors likely contribute to this partial correspondence. Closely related pathogens may produce distinct disease symptoms because of differences in infection strategies, virulence factors, tissue colonization, or interactions with the host. Conversely, phylogenetically distant pathogens may exhibit similar visual symptoms by inducing comparable physiological responses or affecting similar host tissues, resulting in convergent disease phenotypes [28,57]. Consequently, visual similarity does not necessarily imply close evolutionary relatedness. Moreover, the phenotype represented in leaf images reflects not only pathogen evolutionary history but also disease progression and the complex interactions between the pathogen and its host, further contributing to variation in symptom expression [58].

Host and environmental influences add another layer of complexity to disease expression. Host genotype, developmental stage, and immune responses interact with environmental factors such as temperature, humidity, and nutrient availability to influence symptom severity and appearance [6,59–62]. As a result, the visible phenotype captured in leaf images represents the combined effects of pathogen biology, host physiology, and environmental conditions rather than pathogen evolution alone.

The scope of the phylogenetic analysis should also be considered when interpreting these findings. The comparison was intentionally restricted to the fungal subset of the dataset because it represented the only pathogen group with sufficient taxonomic diversity for meaningful comparative phylogenetic analysis. Furthermore, the relatively conserved 28S rRNA marker is well suited for reconstructing broad evolutionary relationships but provides limited resolution for distinguishing closely related fungal taxa [55,63]. Future studies incorporating a broader diversity of pathogen groups together with higher-resolution molecular markers may provide a more comprehensive understanding of the relationship between pathogen evolution and disease phenotype.

Overall, the observed phenotype–phylogeny association supports the biological relevance of the CNN-derived embeddings while demonstrating that visual disease phenotypes represent an integrated outcome of evolutionary, physiological, and environmental

processes. Rather than serving as a direct surrogate for phylogeny, deep learning-based phenotypic representations provide complementary biological information that can enhance the interpretation of plant disease diversity when integrated with traditional evolutionary analyses [64].

4.7. Implications for Multi-Modal Plant Disease Analysis

The findings of this study highlight the value of integrating complementary sources of biological information to better understand plant disease systems. While deep learning effectively captured phenotypic characteristics for accurate disease classification, and phylogenetic analysis provided insight into pathogen evolutionary relationships, the moderate association between these two domains indicates that neither approach alone fully explains the complexity of plant–pathogen interactions. Instead, phenotypic and evolutionary information provide complementary perspectives on disease biology.

Future research could further strengthen this integrative framework by incorporating higher-resolution molecular markers, such as the internal transcribed spacer (ITS) region or whole-genome sequencing, to improve phylogenetic resolution [62,65]. Expanding the diversity of pathogen groups and increasing the number of representative disease classes would also enable broader evaluations of phenotype–phylogeny relationships. In addition, integrating environmental variables and host genotype information could improve the characterization of disease phenotypes by accounting for important biological and ecological factors influencing symptom development [3,66].

From a methodological perspective, multi-modal machine learning approaches that integrate image-based phenotyping with genomic, taxonomic, and environmental information represent a promising direction for advancing plant disease diagnostics and precision agriculture [11,16]. Such approaches could improve both predictive performance and biological interpretability, facilitating more comprehensive analyses of plant disease diversity and supporting decision-making in crop health monitoring and disease management.

4.8. Practical Implications for Precision Agriculture

Beyond its biological insights, the proposed framework has potential applications in precision agriculture and digital crop health monitoring. Recent advances in deep learning and computer vision have demonstrated considerable potential for supporting rapid, image-based plant disease diagnosis using smartphones, unmanned aerial vehicles (UAVs), and other field imaging platforms, thereby facilitating early disease detection and more timely crop management decisions. At the same time, recent reviews have emphasized that successful field deployment depends not only on classification accuracy but also on model interpretability, computational efficiency, robustness under variable environmental conditions, and seamless integration with agricultural decision-support systems [67].

In this context, the incorporation of explainable artificial intelligence techniques, such as Grad-CAM, provides an additional advantage by highlighting the symptomatic regions that contribute to model predictions, thereby increasing model transparency and supporting expert interpretation. Likewise, the phenotypic embedding analysis presented in this study demonstrates that deep learning models can learn structured representations of disease symptoms that extend beyond conventional classification tasks. Such latent representations may support future applications including disease similarity analysis, retrieval of visually related diseases, and integration with genomic, environmental, and agronomic information within multimodal decision-support systems. Recent studies have identified the integration of explainable artificial intelligence, multimodal sensing, and

biologically informed representations as important directions for the next generation of intelligent agricultural systems [68].

Although the present study focused on soybean diseases, the proposed analytical framework is largely model-agnostic and could potentially be adapted to other crops provided that sufficiently diverse, representative, and well-annotated image datasets are available. Future work should evaluate the proposed framework under practical agricultural conditions, including deployment on lightweight edge devices and mobile platforms, integration with UAV-based monitoring systems, and validation using multi-location datasets encompassing diverse crop species, cultivars, and environmental conditions. Such developments would facilitate the translation of deep learning-based disease phenotyping from controlled research settings to scalable real-time crop monitoring applications in precision agriculture [69].

4.9. Study Limitations

Several limitations of this study should be acknowledged. First, although the proposed framework achieved strong classification performance, the dataset remained relatively small for deep learning applications and highly imbalanced across disease classes, particularly for underrepresented diseases such as Cercospora leaf blight and Septoria. While weighted sampling, class-weighted loss, data augmentation, and regularization helped mitigate these challenges, larger and more balanced datasets are needed to improve model robustness, provide more reliable estimates of class-specific performance, and reduce potential sampling bias.

Second, the sequence-based phylogenetic analysis was intentionally restricted to the fungal subset of the dataset because it represented the only pathogen group with sufficient taxonomic diversity for meaningful comparative analysis. Although this enabled the detection of a significant phenotype–phylogeny association, the relatively small number of fungal taxa limits the broader generalizability of these findings. Expanding future studies to include a greater diversity of pathogen groups will allow more comprehensive evaluations of phenotype–phylogeny relationships.

Third, the use of the conserved 28S rRNA marker provided a robust framework for broad phylogenetic inference but offered limited resolution for distinguishing closely related fungal taxa. Higher-resolution molecular markers, such as ITS regions, multilocus datasets, or whole-genome sequences, may improve future evolutionary analyses by capturing finer-scale genetic variation.

Fourth, the Grad-CAM interpretability analysis was qualitative because manually annotated lesion masks were unavailable. Consequently, the highlighted regions should be interpreted as evidence of model attention rather than quantitative validation of lesion localization. Future studies incorporating expert lesion annotations together with larger multimodal datasets will enable more rigorous evaluation of the biological relevance of CNN-derived feature representations.

The generalizability of the proposed framework should be interpreted with appropriate caution. The present study was conducted using a publicly available dataset comprising 703 soybean leaf images acquired under a limited range of imaging conditions. Consequently, the model was not evaluated across different geographic regions, environmental conditions, soybean cultivars, growth stages, imaging devices, or naturally occurring mixed infections. Variations in illumination, background complexity, symptom severity, leaf orientation, camera characteristics, and cultivar-specific symptom expression may influence model performance under practical field conditions. Recent studies have emphasized that robust deep learning applications in plant pathology require larger, more diverse datasets collected across multiple environments and imaging scenarios [70,71]. Furthermore, emerg-

ing developments in computational plant pathology, including multimodal phenotyping, federated learning, transformer-based architectures, and concept-driven explainability frameworks, provide promising opportunities to improve the robustness, scalability, and biological interpretability of future disease analysis systems [72,73]. Accordingly, future research should validate the proposed framework using independent multi-location datasets encompassing diverse environments, soybean cultivars, growing seasons, and imaging conditions to assess its robustness and facilitate its translation into practical decision-support systems for precision agriculture.

Finally, dataset partitioning was necessarily performed at the individual-image level because the public dataset did not provide identifiers for individual leaves, plants, plots, fields, imaging sessions, soybean cultivars, or geographic locations. Although exact image reuse across partitions was prevented and available metadata were inspected, it was not possible to verify that all visually related images originated from independent biological specimens or collection environments. Consequently, residual dependence among images and learning of background-, specimen-, or acquisition-specific features cannot be completely excluded. The reported held-out performance should therefore be interpreted as internal validation within the available dataset rather than evidence of external generalizability. Future studies should employ group-aware splitting based on plant, leaf, field, location, and acquisition-session identifiers and should include external validation across independent regions, cultivars, growing seasons, devices, and field conditions.

5. Conclusions

This study presents an integrative framework combining deep learning-based phenotypic analysis with phylogenetic inference to investigate the relationship between soybean disease symptoms and pathogen evolutionary history. The proposed EfficientNet-B0 model achieved high classification performance across ten soybean disease classes, demonstrating that convolutional neural networks can effectively learn discriminative visual representations despite substantial class imbalance. Furthermore, the learned feature embeddings formed a structured phenotypic space that clearly separated most disease classes while preserving biologically meaningful relationships among visually similar disease symptoms.

Comparative analyses of phenotypic similarity, taxonomy, and sequence-based phylogeny showed that the learned feature space was organized primarily according to visual symptom characteristics rather than pathogen taxonomy. Nevertheless, the Mantel analysis identified a moderate and statistically significant association between phenotypic and phylogenetic distances, indicating that CNN-derived feature representations contain a detectable phylogenetic signal while also reflecting additional sources of biological variation. These findings suggest that pathogen evolutionary history contributes to disease phenotype but does not solely determine its visual expression.

The results further demonstrate that disease phenotypes arise from the combined influence of pathogen evolution, host responses, disease progression, and environmental conditions. Consequently, deep learning-derived phenotypic representations should not be regarded as direct surrogates for pathogen phylogeny but rather as complementary sources of biological information that can enhance the interpretation of plant disease diversity.

From a methodological perspective, this study illustrates the value of integrating image-based phenotyping with phylogenetic analysis to move beyond automated disease classification toward biologically informed interpretation of deep learning models. Such integrative approaches have the potential to improve both the interpretability of artificial intelligence systems and our understanding of plant–pathogen interactions.

Future research should expand both phenotypic and genomic datasets, incorporate a broader diversity of pathogen groups, employ higher-resolution molecular markers

such as ITS regions, multilocus datasets, or whole-genome sequencing, and integrate additional sources of information including host genotype and environmental variables. These advances will enable more comprehensive investigations of phenotype–phylogeny relationships and support the development of more accurate, interpretable, and biologically informed disease diagnostic systems for precision agriculture.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy16151486/s1, Figure S1: Pairwise genetic distance matrix among fungal pathogens computed using 28S rRNA sequence divergence (p-distance); Figure S2: Disease-level genetic distance matrix derived from fungal pathogen sequences, mapped to corresponding disease classes; Figure S3: Neighbor-Joining phylogenetic tree of fungal pathogens inferred from 28S rRNA sequences, showing evolutionary relationships among species; Figure S4: Comparison of clustering structures derived from phenotypic similarity (left) and sequence-based genetic distance (right) for fungal pathogens; Figure S5: Comparison between phenotype-based and taxonomy-derived dendrograms, highlighting differences between visual similarity and evolutionary relationships; Figure S6: Relationship between phenotypic and sequence-derived distances among fungal pathogens. Weak correlation suggests partial decoupling between visual symptoms and genetic relatedness; Figure S7: Phenotypic similarity matrix at the pathogen level derived from image embeddings, illustrating relationships between causal agents beyond disease labels. Table S1: Pairwise Euclidean distance matrix among disease-class embeddings derived from the trained EfficientNet-B0 model; Table S2: Phenotypic similarity matrix obtained by transforming Euclidean distances (Similarity = 1 − Distance); Table S3: Pairwise disease relationships retained in the thresholded phenotypic similarity network. Only disease pairs with cosine similarity greater than or equal to the network threshold are included; Table S4: Pairwise phenotypic and sequence-derived distances among fungal diseases used in the Mantel test. Phenotypic distances were calculated from disease embedding centroids derived from the trained EfficientNet-B0 model, whereas sequence-derived distances were obtained from the corresponding fungal pathogen phylogeny based on 28S rRNA sequences. Each row represents one pairwise comparison included in the Mantel correlation analysis after excluding unavailable sequence comparisons.

Author Contributions: Conceptualization, M.A.K.; methodology, D.K. and K.M.; formal analysis, M.A.K. and D.K.; investigation, M.A.K.; data curation, M.A.K.; writing—original draft preparation, M.A.K.; writing—review and editing, K.M. and D.K.; supervision, M.A.K.; project administration, M.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The image dataset used in this study was assembled primarily from publicly available images obtained through Kaggle.com, supplemented with sudden death syndrome images from our previous study. The source code and Jupyter notebook used for data preprocessing, deterministic partitioning, model training, evaluation, feature extraction, visualization, and phenotype–phylogeny analysis are publicly available at the following GitHub repository: [https://github.com/abdelmajidk/soybean-phenotype-phylogeny (accessed on 26 July 2026)]. The public dataset source and required directory organization are described in the repository README.

Acknowledgments: The authors acknowledge the developers and maintainers of open-source scientific software, including Python, Jupyter Notebooks, pandas, Matplotlib, and seaborn, whose tools enabled the reproducible computational analyses presented in this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CNN	Convolutional Neural Network

F1	Weighted F1-score
Grad-CAM	Gradient-weighted Class Activation Mapping
ITS	Internal Transcribed Spacer
ML	Machine Learning
NJ	Neighbor-Joining
PCA	Principal Component Analysis
r	Spearman correlation coefficient
rRNA	Ribosomal RNA
SDS	Sudden Death Syndrome
t-SNE	t-distributed Stochastic Neighbor Embedding

References

1. Hartman, G.L.; West, E.D.; Herman, T.K. Crops that feed the World 2. Soybean—Worldwide production, use, and constraints caused by pathogens and pests. *Food Secur.* **2011**, *3*, 5–17. [CrossRef]
2. Strange, R.N.; Scott, P.R. Plant disease: A threat to global food security. *Annu. Rev. Phytopathol.* **2005**, *43*, 83–116. [CrossRef] [PubMed]
3. Savary, S.; Willocquet, L.; Pethybridge, S.J.; Esker, P.; McRoberts, N.; Nelson, A. The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* **2019**, *3*, 430–439. [CrossRef] [PubMed]
4. Oerke, E.C. Crop losses to pests. *J. Agric. Sci.* **2006**, *144*, 31–43.
5. Agrios, G.N. *Plant Pathology*; Elsevier: Amsterdam, The Netherlands, 2005.
6. Jones, J.D.G.; Dangl, J.L. The plant immune system. *Nature* **2006**, *444*, 323–329. [CrossRef] [PubMed]
7. Bock, C.H.; Poole, G.H.; Parker, P.E.; Gottwald, T.R. Plant disease severity estimated visually, by digital photography and im-age analysis, and by hyperspectral imaging. *Crit. Rev. Plant Sci.* **2010**, *29*, 59–107. [CrossRef]
8. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
9. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]
10. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [CrossRef]
11. Kamilaris, A.; Prenafeta-Boldu, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
12. Too, E.C.; Yujian, L.; Njuki, S.; Yingchun, L. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* **2019**, *161*, 272–279. [CrossRef]
13. Farah, N.; Drack, N.; Dawel, H.; Buettner, R. A deep learning-based approach for the detection of infested soybean leaves. *IEEE Access* **2023**, *11*, 99670–99679. [CrossRef]
14. Liu, J.; Wang, X. Plant diseases and pests detection based on deep learning: A review. *Plant Methods* **2021**, *17*, 22. [CrossRef] [PubMed]
15. Sajitha, P.; Andrushia, A.D.; Anand, N.; Naser, M.Z. A review on machine learning and deep learning image-based plant disease classification for industrial farming systems. *J. Ind. Inf. Integr.* **2024**, *38*, 100572. [CrossRef]
16. Upadhyay, A.; Chandel, N.S.; Singh, K.P.; Chakraborty, S.K.; Nandede, B.M.; Kumar, M.; Subeesh, A.; Upendar, K.; Elbeltagi, A. Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture. *Artif. Intell. Rev.* **2025**, *58*, 92. [CrossRef]
17. Raval, H.; Chaki, J. Ensemble transfer learning meets explainable AI: A deep learning approach for leaf disease detection. *Ecol. Inform.* **2024**, *84*, 102925. [CrossRef]
18. Ashurov, A.Y.; Al-Gaashani, M.S.A.M.; Samee, N.A.; Alkanhel, R.; Atteia, G.; Abdallah, H.A.; Muthanna, S.A.M. Enhancing plant disease detection through deep learning: A Depthwise CNN with squeeze and excitation integration and residual skip connections. *Front. Plant Sci.* **2025**, *15*, 1505857. [CrossRef] [PubMed]
19. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.-R. Explaining deep neural networks and beyond: A review of methods and applications. *Proc. IEEE* **2021**, *109*, 247–278. [CrossRef]
20. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359.
21. Porfirio, R.P.; Madeira, R.N.; Santos, P.A. Exploring Explainable AI Techniques for Plant Disease Classification in Digital Agriculture. *Procedia Comput. Sci.* **2025**, *265*, 175–182. [CrossRef]
22. Kondaveeti, H.K.; Simhadri, C.G. Evaluation of deep learning models using explainable AI with qualitative and quantitative analysis for rice leaf disease detection. *Sci. Rep.* **2025**, *15*, 31850. [CrossRef] [PubMed]

23. Schocha, C.L.; Seifertb, K.A.; Huhndorfc, S.; Robertd, V.; Spougea, J.L.; Levesqueb, C.A.; Chenb, W.; Consortiuma, F.B. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6241–6246. [CrossRef] [PubMed]
24. Tekpinar, A.D.; Kalmer, A. Utility of various molecular markers in fungal identification and phylogeny. *Nova Hedwig.* **2019**, *109*, 187–224. [CrossRef]
25. Taylor, J.W.; Jacobson, D.J.; Kroken, S.; Kasuga, T.; Geiser, D.M.; Hibbett, D.S.; Fisher, M.C. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.* **2000**, *31*, 21–32. [CrossRef] [PubMed]
26. Naranjo-Ortiz, M.A.; Gabaldón, T. Fungal evolution: Diversity, taxonomy and phylogeny of the Fungi. *Biol. Rev.* **2019**, *94*, 2101–2137. [CrossRef] [PubMed]
27. Hosseini, B.; Voegele, R.T.; Link, T.I. Diagnosis of soybean diseases caused by fungal and oomycete pathogens: Existing meth-ods and new developments. *J. Fungi* **2023**, *9*, 587. [CrossRef] [PubMed]
28. Gilbert, G.S.; Parker, I.M. The evolutionary ecology of plant disease: A phylogenetic perspective. *Annu. Rev. Phytopathol.* **2016**, *54*, 549–578. [CrossRef] [PubMed]
29. Furbank, R.T.; Tester, M. Phenomics–technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **2011**, *16*, 635–644. [CrossRef] [PubMed]
30. Gallego-Ortiz, C.; Martel, A.L. A graph-based lesion characterization and deep embedding approach for improved computer-aided diagnosis of nonmass breast MRI lesions. *Med. Image Anal.* **2019**, *51*, 116–124. [CrossRef] [PubMed]
31. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
32. Carter, H.; Hofree, M.; Ideker, T. Genotype to phenotype via network analysis. *Curr. Opin. Genet. Dev.* **2013**, *23*, 611–621. [CrossRef] [PubMed]
33. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]
34. Sun, H.; Chu, H.-Q.; Qin, Y.-M.; Hu, P.; Wang, R.-F. Empowering smart soybean farming with deep learning: Progress, Challenges, and Future Perspectives. *Agronomy* **2025**, *15*, 1831. [CrossRef]
35. Li, X.; Chen, M.; He, S.; Xu, X.; He, L.; Wang, L.; Gao, Y.; Tang, F.; Gong, T.; Wang, W.; et al. Estimation of soybean yield based on high-throughput phenotyping and machine learning. *Front. Plant Sci.* **2024**, *15*, 1395760. [CrossRef] [PubMed]
36. Yu, H.; Weng, L.; Wu, S.; He, J.; Yuan, Y.; Wang, J.; Xu, X.; Feng, X. Time-Series Field Phenotyping of Soybean Growth Analysis by Combining Multimodal Deep Learning and Dynamic Modeling. *Plant Phenomics* **2024**, *6*, 0158. [CrossRef] [PubMed]
37. Barabasi, A.-L.; Gulbahce, N.; Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **2011**, *12*, 56–68. [CrossRef] [PubMed]
38. Mahlein, A.-K. Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* **2016**, *100*, 241–251. [CrossRef] [PubMed]
39. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **1967**, *27*, 209–220. [PubMed]
40. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual, Version 3.10*; Python Language Reference; Python Software Foundation; CreateSpace: Scotts Valley, CA, USA, 2009. Available online: https://www.python.org (accessed on 1 June 2026).
41. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.E.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks—A publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing, Göttingen, Germany, 7–9 June 2016*; IOS Press: Amsterdam, The Netherlands, 2016; p. 87.
42. Harris, C.R.; Millman, K.J.; Van Der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef] [PubMed]
43. McKinney, W. Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 51–56. [CrossRef]
44. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
45. Waskom, M.L. Seaborn: Statistical data visualization. *J. Open-Source Softw.* **2021**, *6*, 3021. [CrossRef]
46. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*; Neural Information Processing Systems Foundation, Inc.: San Diega, CA, USA, 2019.
47. TorchVision Maintainers and Contributors. TorchVision: PyTorch's Computer Vision Library. GitHub Repository. 2016. Available online: https://github.com/pytorch/vision (accessed on 1 June 2026).
48. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
49. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

50. Cock, P.J.A.; Antao, T.; Chang, J.T.; Chapman, B.A.; Cox, C.J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422. [CrossRef] [PubMed]
51. Goodfellow, I. Nips 2016 tutorial: Generative adversarial networks. *arXiv* **2016**, arXiv:1701.00160.
52. Ribeiro, M.T.; Singh, S.; Guestrin, C. Model-agnostic interpretability of machine learning. *arXiv* **2016**, arXiv:1606.05386.
53. Tjoa, E.; Guan, C. A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4793–4813. [CrossRef] [PubMed]
54. Loscalzo, J.; Kohane, I.; Barabasi, A.L. Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. *Mol. Syst. Biol.* **2007**, *3*, MSB4100163. [CrossRef] [PubMed]
55. Newton, A.C.; Gravouil, C.; Fountaine, J.M. Managing the ecology of foliar pathogens: Ecological tolerance in crops. *Ann. Appl. Biol.* **2010**, *157*, 343–359. [CrossRef]
56. Nilsson, R.H.; Kristiansson, E.; Ryberg, M.; Hallenberg, N.; Larsson, K.-H. Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evol. Bioinform.* **2008**, *4*, EBO-S653. [CrossRef] [PubMed]
57. Brown, J.K.M.; Tellier, A. Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annu. Rev. Phytopathol.* **2011**, *49*, 345–367. [CrossRef] [PubMed]
58. Dodds, P.N.; Rathjen, J.P. Plant immunity: Towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **2010**, *11*, 539–548. [CrossRef] [PubMed]
59. Burdon, J.J.; Thrall, P.H. Pathogen evolution across the agro-ecological interface: Implications for disease management. *Evol. Appl.* **2008**, *1*, 57–65. [CrossRef] [PubMed]
60. Garrett, K.A.; Dendy, S.P.; Frank, E.E.; Rouse, M.N.; Travers, S.E. Climate change effects on plant disease: Genomes to ecosystems. *Annu. Rev. Phytopathol.* **2006**, *44*, 489–509. [CrossRef] [PubMed]
61. Paradis, E. *Analysis of Phylogenetics and Evolution with R*; Springer: New York, NY, USA, 2006.
62. Hibbett, D.S.; Ohman, A.; Glotzer, D.; Nuhn, M.; Kirk, P.; Nilsson, R.H. Progress in molecular and morphological taxon dis-covery in Fungi and options for formal classification of environmental sequences. *Fungal Biol. Rev.* **2011**, *25*, 38–47. [CrossRef]
63. Losos, J.B. Convergence, adaptation, and constraint. *Evolution* **2011**, *65*, 1827–1840. [CrossRef] [PubMed]
64. McDonald, B.A.; Linde, C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu. Rev. Phytopathol.* **2002**, *40*, 349–379. [CrossRef] [PubMed]
65. Fisher, M.C.; Henk, D.A.; Briggs, C.J.; Brownstein, J.S.; Madoff, L.C.; McCraw, S.L.; Gurr, S.J. Emerging fungal threats to animal, plant and ecosystem health. *Nature* **2012**, *484*, 186–194. [CrossRef] [PubMed]
66. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
67. Ghimire, S.; Lamsal, R.; Sankuratri, A.; Lakkarsu, P.; Vutla, S. Artificial Intelligence-Driven Plant Disease Detection and Diagnosis: A Comprehensive Review of Deep Learning Approaches, Multimodal Sensing Technologies, and Future Perspectives in Precision Agriculture. *Plant Pathol. J.* **2026**, *42*, 121–137. [CrossRef] [PubMed]
68. Anam, I.; Arafat, N.; Hafiz, M.S.; Jim, J.R.; Kabir, M.M.; Mridha, M.F. A systematic review of UAV and AI integration for targeted disease detection, weed management, and pest control in precision agriculture. *Smart Agric. Technol.* **2024**, *9*, 100647. [CrossRef]
69. Gudge, S.; Tiwari, A.; Ratnaparkhe, M.; Jha, P. On construction of data preprocessing for real-life SoyLeaf dataset & disease identification using Deep Learning Models. *Comput. Biol. Chem.* **2025**, *117*, 108417. [CrossRef] [PubMed]
70. Nyawose, T.; Maswanganyi, R.C.; Khumalo, P. A Review on the Detection of Plant Disease Using Machine Learning and Deep Learning Approaches. *J. Imaging* **2025**, *11*, 326. [CrossRef] [PubMed]
71. Chihaoui, N.; Waleed, A.; Mufleh, A.S.S.; Alshraah, S.M.; Alajmi, L.H.R.; Altuwayjiri, S. A comprehensive review on AI-based crop disease detection using leaf image classification and explainable AI. *Discov. Appl. Sci.* **2026**, *8*, 663. [CrossRef]
72. Amara, J.; Kong-Ries, B.; Samuel, S. Explainability of Deep Learning-Based Plant Disease Classifiers Through Automated Concept Identification. *arXiv* **2024**, arXiv:2412.07408.
73. Shawon, S.M.; Niha, F.L.; Sinan, O.H.; Abed, M.J.; Al Qwaid, M.; Sarker, M.T.; Zubair, H.T. Integrating Federated Learning and Explainable AI for Plant Disease Detection: A Systematic Review of Recent Advances in Precision Agriculture. *Int. J. Comput. Intell. Syst.* **2026**, *19*, 138. [CrossRef]