Article

Classification and Analysis of *Agaricus bisporus* Diseases with Pre-Trained Deep Learning Models

Umit Albayrak [1], Adem Golcuk [2], Sinan Aktas [3], Ugur Coruh [4,*], Sakir Tasdemir [2] and Omer Kaan Baykan [5]

1   Ilgin Vocational School, Selcuk University, Konya 42600, Turkey; umitalbayrak@selcuk.edu.tr
2   Department of Computer Engineering, Selcuk University, Konya 42031, Turkey; adem.golcuk@selcuk.edu.tr (A.G.); stasdemir@selcuk.edu.tr (S.T.)
3   Department of Biology, Selcuk University, Konya 42031, Turkey; saktas@selcuk.edu.tr
4   Department of Computer Engineering, Recep Tayyip Erdoğan University, Rize 53020, Turkey
5   Department of Computer Engineering, Konya Technical University, Konya 42250, Turkey; okbaykan@ktun.edu.tr
*   Correspondence: ugur.coruh@erdogan.edu.tr

Abstract: This research evaluates 20 advanced convolutional neural network (CNN) architectures for classifying mushroom diseases in *Agaricus bisporus*, utilizing a custom dataset of 3195 images (2464 infected and 731 healthy mushrooms) captured under uniform white-light conditions. The consistent illumination in the dataset enhances the robustness and practical usability of the assessed models. Using a weighted scoring system that incorporates precision, recall, F1-score, area under the ROC curve (AUC), and average precision (AP), ResNet-50 achieved the highest overall score of 99.70%, demonstrating outstanding performance across all disease categories. DenseNet-201 and DarkNet-53 followed closely, confirming their reliability in classification tasks with high recall and precision values. Confusion matrices and ROC curves further validated the classification capabilities of the models. These findings underscore the potential of CNN-based approaches for accurate and efficient early detection of mushroom diseases, contributing to more sustainable and data-driven agricultural practices.

Keywords: *Agaricus bisporus*; mushroom diseases; deep learning; image processing; precision agriculture; smart farming; convolutional neural networks

1. Introduction

The cultivation of mushrooms, particularly *Agaricus bisporus* (J.E. Lange) Imbach—commonly referred to as the cultured mushroom, white-capped cultivated mushroom, meadow mushroom, or "champion mushroom"—is of profound economic and nutritional importance worldwide. Produced in more than seventy countries [1,2], *A. bisporus* is increasingly favored due to its low calorie, sodium, fat, and cholesterol content; its high levels of protein, carbohydrates, fiber, vitamins, and amino acids; its pleasant taste; widespread availability; and its evaluation as a functional food [3,4]. These attributes have intensified global demand, fostering growth in the mushroom market and supporting food security and sustainable agricultural practices.

Despite these advantages, *A. bisporus* cultivation faces substantial challenges from various fungal diseases—such as bacterial blotch (*Pseudomonas tolaasii*) [5], dry bubble (*Verticillium fungicola*) [6], cobweb (*Cladobotryum mycophilum*) [7], and wet bubble (*Mycogone perniciosa*) [8]—which severely reduce yield and quality if not identified and managed

promptly [9,10]. Such diseases undermine the economic viability of producers and limit consumer access to nutritious food. In response to increasing disease prevalence and pest pressure, the use of chemical pesticides has escalated. However, in regions like Turkey, the Ministry of Agriculture and Forestry has authorized only a limited range of pesticides for mushroom production [11]. This regulatory constraint compels producers to employ unapproved chemicals, posing significant risks. Moreover, the short growth cycle of button mushrooms can result in pesticide residues on the harvested product, raising potential health concerns for consumers [12].

Improving productivity and ensuring public health require rapid and accurate detection of diseased mushrooms. Early diagnosis allows for the prompt removal of infected compost or the application of suitable remedies, preventing further contamination and reducing the need for chemical interventions. This approach not only diminishes economic losses but also advances sustainable agricultural practices. Nonetheless, conventional diagnostic methods remain time-consuming, costly, labor-intensive, and subjective. Such inefficiencies highlight the urgent need for automated and objective detection systems capable of precise, swift, and cost-effective identification of fungal pathogens.

Convolutional Neural Networks (CNNs) [13] have revolutionized plant disease detection in crops like apples [14], tomatoes [15], and wheat [16] by enabling early intervention through accurate classification based on subtle visual differences. Their strength in extracting complex patterns from images makes CNNs well-suited for detecting minute distinctions between healthy and infected specimens [17,18].

However, while transformer-based architectures also show remarkable performance in computer vision, practical constraints in agricultural applications can limit their feasibility—reinforcing the appropriateness of CNNs for our study. Our study specifically focused on CNN architectures due to their proven efficiency in handling datasets of similar scales and their successful track record in agricultural applications. CNNs offer an optimal balance between accuracy and computational efficiency, making them particularly well-suited for the practical requirements of crop disease detection systems. However, we acknowledge the potential of transformer-based models and consider their evaluation as a valuable direction for future research, particularly as computational resources become more accessible and dataset sizes expand.

Despite these advances, the control of mushroom diseases has received comparatively little attention in current research. The unique challenges of mushroom farming—controlled growing environments, morphological similarities among diseases, and limited comprehensive studies—have inhibited the adaptation of CNN-based methods in this domain [19].

Our study shows that incorporating deep learning methodologies into detecting and classifying fungal diseases in *Agaricus bisporus* cultivation leads to faster, more efficient, and more objective results than traditional manual monitoring methods. By reducing diagnostic errors and facilitating timely interventions, these technologies can reduce chemical usage, enhance food safety, bolster economic stability for producers, and support sustainable agricultural systems [20]. This underscores the imperative for focused research to apply CNN-based solutions to the specific challenges of mushroom disease management.

This study aims to address critical gaps in the diagnosis and classification of mushroom diseases by focusing on three primary objectives:

Custom Dataset Development: The research involves constructing a high-quality dataset comprising 3195 images—2464 depicting diseased mushrooms and 731 representing healthy specimens. All images are captured under standardized conditions of lighting, color temperature, brightness, exposure, white balance, distance, and camera angle. This controlled setup not only simulates authentic agricultural scenarios but also ensures that

the resulting dataset is both durable and broadly applicable [21]. It thus provides a reliable foundation for training and evaluating machine learning models [22].

Comprehensive Evaluation Metric: This study systematically compares 20 cutting-edge CNN architectures, including DenseNet-201 [23], ResNet-50 [24], DarkNet-53 [25], EfficientNet-b0 [26], and MobileNet-v2 [27], to determine their effectiveness in mushroom disease classification. Multiple performance indicators—accuracy, precision, recall, F1-score, area under the ROC curve (AUC), and average precision (AP)—are employed. A weighted scoring method is introduced to synthesize these metrics into a single comprehensive indicator, thereby capturing each model's overall balance between accuracy and efficiency.

Enhanced Analytical Framework: By capitalizing on the strengths of both MATLAB and Python (version 3.9), the research adopts a hybrid analytical pipeline. MATLAB (version R2023b) is utilized for data preprocessing, model training, and generating advanced visualizations such as ROC and precision-recall curves. Python (version 3.9) supplements these efforts with additional visualization and analytical tools. This integrated approach enhances analytical precision, reproducibility, and operational efficiency [28].

The effective detection of diseases in *Agaricus bisporus* is of critical importance due to its significant role as a nutritional source and its substantial share in the food industry. The short cultivation cycle of *Agaricus bisporus* increases the risk of chemical residues from pesticide use, which can negatively impact human health. This highlights the urgent need for systems that can facilitate automated disease detection. While this study does not directly automate disease detection, it lays a solid foundation for future developments in this direction, offering significant potential to reduce producers' losses in terms of time, effort, and financial resources, thereby contributing to sustainable production practices.

In the existing literature, there are no studies that feature such a comprehensive and unique dataset for detecting *Agaricus bisporus* diseases. The novelty of our work lies in the creation of this dataset, which was developed over a period of more than one year through direct engagement with mushroom production facilities across various regions of Turkey. This extensive data collection effort ensured that the dataset transcended local limitations and provided a broad representation, significantly enhancing its value.

Furthermore, the custom dataset was subjected to training and evaluation using 20 different pre-trained deep learning models, an approach rarely seen in similar studies. Among these models, ResNet-50 achieved an impressive accuracy of 99.70%, underscoring the reliability and robustness of the dataset. This study not only advances the field of disease classification for mushrooms but also serves as a benchmark for agricultural classification problems more broadly. The unprecedented scope of model experimentation further underscores the study's contribution and sets it apart in the academic literature.

Additionally, prior investigations in plant pathology underscore the importance of early and accurate disease detection in various crops, illustrating how deep learning and imaging methodologies can significantly improve yield out comes and overall sustainability.

Over the last decade, a wealth of research has focused on leveraging advanced imaging techniques, convolutional neural networks (CNNs), and various machine learning (ML) methods to diagnose and classify diseases across a wide range of crops. Among these, imaging has shown significant potential in detecting plant diseases before fatal symptoms appear. For instance, Pane et al. utilized image analysis combined with Random Forest modeling to distinguish between healthy and powdery mildew-infected wild rocket leaves [29], demonstrating the potential of machine learning in disease classification. However, compared to the high accuracy achieved by DenseNet-201 in our study (99.98%), Random Forest's performance is limited by its reliance on manually selected features and less robust handling of diverse datasets. This highlights the advantage of deep learning models like DenseNet-201 in capturing complex patterns directly from raw image data,

offering superior scalability and precision for real-world applications. Similarly, Xie et al. highlighted that imaging can potentially identify wheat crown rot infections at an early stage, thus supporting preventative management strategies by using SVM [30].

Additionally, rice production has faced numerous challenges in recent years, where traditional methods are still being used to detect rice diseases. A recent study developed an automated rice blast disease diagnosis technique using deep learning, image processing, and transfer learning with pre-trained models such as Inception V3, VGG16, VGG19, and ResNet-50. The dataset included 2000 images (1200 blast-infected and 800 healthy). Among the models, ResNet-50 demonstrated the highest accuracy of 99.75%, with a loss rate of 0.33, and achieved superior validation metrics, including an F1-score of 99.70 and AUC of 99.83%. These findings underscore the applicability of ResNet-50 in high-precision agricultural disease detection scenarios, further supporting its potential for integration into automated systems [31].

Moreover, in medicinally significant crops such as black pepper, recent work has shown that transfer learning approaches (e.g., Inception V3, GoogleNet, SqueezeNet, and ResNet18) can achieve up to 99.67% accuracy in detecting various leaf diseases. By employing carefully selected hyperparameters and real-time annotated leaf images, such methods effectively diagnose anthracnose, slow wilt, and early phytophthora infection, thereby contributing to timely intervention in black pepper cultivation [32].

Deep learning, especially CNN-based frameworks, has emerged as a leading approach for plant disease detection. Chen et al. presented BLSNet, a UNet-based semantic segmentation network optimized with attention mechanisms and multi-scale feature extraction, to detect bacterial leaf streaks in rice [33]. However, compared to DenseNet-201, which demonstrated an accuracy of 99.98% in mushroom disease classification, BLSNet achieved a slightly lower accuracy of 99.33%, highlighting DenseNet-201's superior performance in handling complex inter-class similarities and achieving precise disease detection in challenging scenarios.

Chowdhury et al. adopted EfficientNet-based architectures and segmentation models (U-Net and modified U-Net) for tomato leaf disease classification, demonstrating high accuracy when trained on segmented leaf images [34]. Similar advancements were reported by Bansal et al. and Tan et al. for apple and tomato leaf disease classification, respectively, confirming the general effectiveness of CNNs in plant pathology [35,36].

In addition, Zhao et al. integrated attention mechanisms into CNNs to improve tomato disease diagnosis, achieving an average accuracy of 96.81% and real-time performance with a diagnosis speed of 31.68 ms [37]. However, compared to their results, our DenseNet-201-based model achieved a significantly higher accuracy of 99.98%, demonstrating its superior robustness and precision in disease classification. Peng et al. proposed a fused deep-feature approach combined with SVM classifiers for grape leaf disease identification, showing that the integration of multiple feature sets can lead to more robust disease recognition [38]. Studies by Li et al. and Wan et al. emphasized the importance of hyperspectral analysis in non-destructive and early disease detection [39,40]. Lin et al. and Wang et al. focused on lightweight models such as GrapeNet or optimized YOLOv5 architectures, which ensure efficient, real-time detection [41,42]. However, these models often trade off accuracy for efficiency. For instance, GrapeNet achieves an accuracy of 86.29% with only 2.15 million parameters, making it suitable for mobile and embedded systems. In contrast, DenseNet-201, as evaluated in our study, demonstrated a significantly higher accuracy of 99.98% across all mushroom disease categories. While DenseNet-201 has a larger parameter count, its superior performance highlights its capability to deliver exceptional precision and reliability, making it more suitable for applications where accuracy is paramount, such as early disease detection in critical agricultural systems.

Furthermore, multi-disease detection has recently become a prominent research topic, especially regarding sustainability goals and global food security. In one study, a set of nine pre-trained CNNs (e.g., DenseNet201, ResNet50, EfficientNetB7) was combined through early-fusion and voting-ensemble strategies, yielding accuracy scores as high as 97.79% on the PlantVillage dataset, which comprises 15 diverse classes. This ensemble methodology has been reported to enhance generalization and mitigate overfitting, a challenge frequently faced in plant disease detection workflows [43].

Wu et al. introduced DS-DETR, a model based on the DETR framework, which achieved a classification accuracy of 96.4% by efficiently segmenting and evaluating tomato leaf diseases through advanced pre-training and spatially modulated co-attention mechanisms. However, this accuracy is lower compared to the 99.98% accuracy achieved by DenseNet-201, highlighting the superior classification performance of the latter [44].

Yin et al. designed DISE-Net, a deep convolutional network incorporating attention mechanisms, to precisely classify maize leaf spots [45], achieving an accuracy of 97.12%; however, DenseNet-201 surpasses this with an accuracy of 99.98%, demonstrating superior performance in classification precision and feature extraction.

Zendler et al. applied shallow convolutional networks to assess downy mildew severity on grapevine leaf discs, achieving a validation accuracy of 95% [46]. However, DenseNet-201 demonstrated superior performance with an accuracy of 99.98% in our work, showcasing its advanced ability to handle complex visual patterns and achieve higher precision in classification tasks.

Beyond common field crops, recent work has extended to niche agricultural products like mushrooms. Zahan et al. explored deep learning models, including ResNet15, to classify several mushroom diseases, finding an accuracy of 88.40% [47]. Furthermore, Albayrak et al. constructed a comprehensive image-based dataset for *Agaricus bisporus* diseases, paving the way for CNN applications in mushroom disease detection and classification [3]. Gu et al., Orchi et al., and Mehmood et al. demonstrated high-precision results in multi-disease detection scenarios, achieving accuracies of 99.72%, 99.64%, and 99.00%, respectively, using various DL and ML architectures [48–50].

Likewise, in the domain of grape leaf disease detection, a strategy incorporating CNNs with Gaussian noise augmentation was shown to reduce overfitting and attain 99.88% accuracy when using PlantVillage data and various pre-trained models (e.g., VGG16, ResNet50, InceptionV3, DenseNet121). This approach broadens data diversity and has been noted to substantially improve model generalization [51].

Additionally, image segmentation techniques have proven essential for improving classification performance in complex backgrounds, as shown by Ngugi et al., who achieved an accuracy of 97.66% using their KijaniNet model, and Zhang et al., whose CRF_ResUNet++ model attained an accuracy of 99.11%, both of which emphasized effective image preprocessing methods [52,53]. Zhao et al. and Jiang et al. focused on tasks ranging from spore segmentation to weed and crop recognition, demonstrating that integrated feature extraction and semi-supervised learning approaches can further enhance precision agriculture technologies [54,55].

Overall, current literature underscores the capabilities of deep learning techniques, combined with advanced imaging methods and feature fusion strategies, in accurately diagnosing plant diseases. Such methodologies, adapted to specific species—from common field crops to niche produce like mushrooms—provide a foundation for implementing efficient, real-time disease management solutions in modern agriculture.

The current study stands out as a novel contribution to this field by presenting the most comprehensive and unique dataset developed specifically for the classification of *Agaricus bisporus* diseases. Unlike previous studies, our work incorporates the use of 20 pre-

trained deep learning models for analysis, offering an unprecedented exploration of model performance in this domain. Among these models, ResNet-50 demonstrated exceptional results, achieving an impressive accuracy of 99.70%. This study not only provides a robust framework for the detection and classification of *Agaricus bisporus* diseases but also facilitates a comprehensive evaluation of the performance of existing pre-trained models in this context. By addressing the limitations of prior research, this work sets a new benchmark in agricultural disease classification and contributes significantly to the development of automated systems for disease detection and management.

2. Materials and Methods

2.1. Dataset Creation Methodology

To capture high-quality mushroom images for dataset creation, a custom-designed portable imaging apparatus was developed, as illustrated in Figures 1–3. This apparatus was engineered to ensure controlled and uniform lighting conditions essential for consistent and reliable image acquisition.


(**a**)	(**b**)
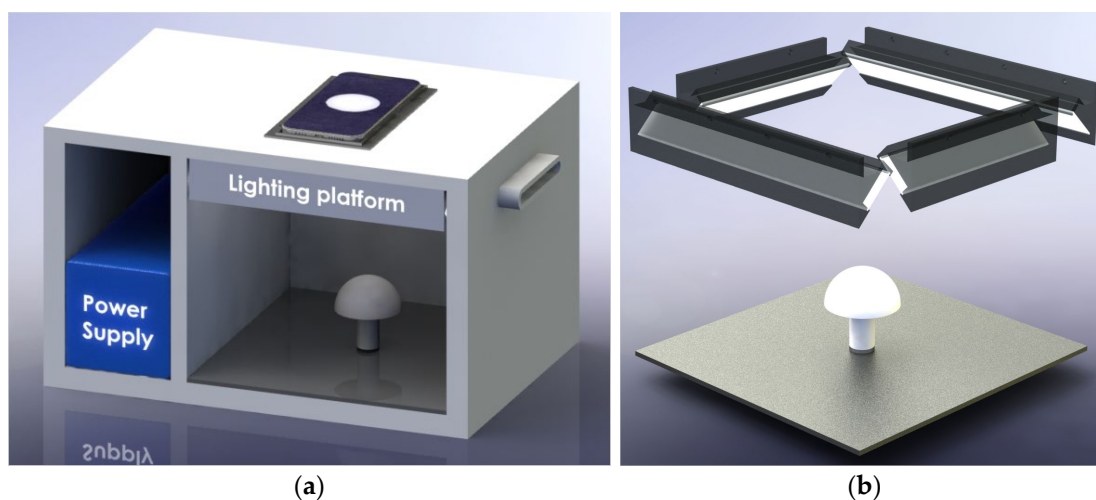
Figure 1. General View and Lighting System Details of the Portable Imaging Apparatus. (**a**) General View of the Portable Imaging Apparatus, highlighting the modular design, power supply compartment, and lighting platform setup. (**b**) Detailed View of the Lighting System of the Portable Imaging Apparatus, highlighting the specially designed 45-degree angled lighting channels equipped with diffusers to minimize glare and ensure uniform illumination, specifically tailored for optimal imaging conditions of the mushroom specimen.

The interior surfaces of the imaging device were coated with matte black bases and matte white walls to optimize lighting quality by reducing glare and reflections [56]. Adjustable lighting channels equipped with white diffusers provided homogeneous illumination, while a consistent 45° angle of light incidence minimized specular reflections and enhanced the visibility of critical features required for accurate disease classification [57,58]. The mushroom samples were positioned on a sturdy matte black felt background to maintain consistent visual context, which is visible in the imaging setup in Figures 2 and 3.

The imaging process was specifically designed to address the rapid progression of diseases in mushrooms, which begins immediately after harvest. To ensure that the disease characteristics were accurately documented in their early stages without delays that could compromise the integrity of the visual data, images had to be captured directly at the production facilities. To eliminate reliance on external lighting and power sources, the device was equipped with its own battery-powered lighting system.
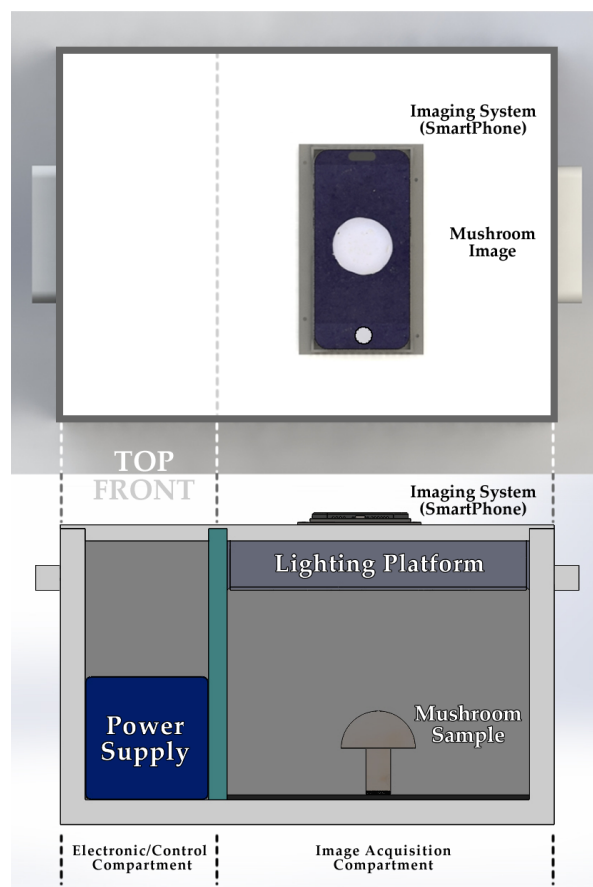
Figure 2. Top and Front View of the Custom Portable Imaging Apparatus, illustrating the internal lighting platform, smartphone-based imaging system, and mushroom placement.
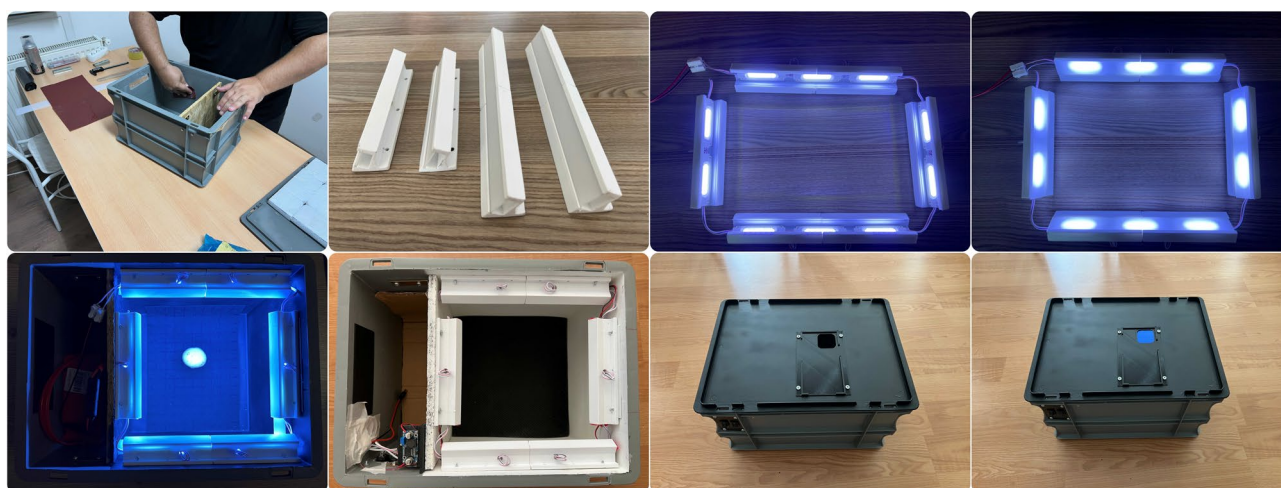
Figure 3. Development Process of the Portable Imaging Apparatus, illustrating key stages from development to final design.

Due to the need for portability and on-site image capture, a smartphone camera was chosen as the imaging device. While this decision was driven by practical constraints, the selected smartphone featured advanced lens technology and imaging capabilities that, in many cases, rival or surpass industrial cameras. This ensured that high-resolution images could be captured effectively in field conditions. Additionally, to preserve the raw integrity of the images, all automatic software corrections by the smartphone were disabled,

and a fully isolated and controlled lighting environment was created to prevent external light interference.

To support extended field operations, a 12 V battery was integrated into the system, enabling the device to function independently in various agricultural environments. This portable power supply, visible in the left compartment of Figure 1, provided the necessary energy for consistent and reliable imaging under field conditions.

The apparatus' modular design, as presented in Figure 1, includes clearly defined compartments for lighting and imaging systems, ensuring minimal interference and consistent, high-quality image acquisition. These carefully implemented measures enabled reliable dataset creation across diverse farming settings, establishing a robust basis for training and evaluating machine learning models. The development process of the portable imaging apparatus designed within the scope of this study is presented in Figure 3.

2.2. Dataset Composition

The dataset consisted of 3195 images classified into five categories: Healthy (731), Bacterial Blotch (576), Dry Bubble (665), Cobweb (664), and Wet Bubble (559). These categories are visually represented in Figure 4, showcasing examples of each condition captured under controlled white-light conditions. Images were taken at multiple stages of disease progression and from various angles, as illustrated in Figure 5, to ensure diversity and enhance the robustness of the dataset for training, validation, and testing. The inclusion of random angles ($\alpha°$) for diversity and upright positions ($90°$) provided a comprehensive representation of the mushroom samples, allowing the models to better capture disease characteristics. Proportional balancing among the classes, as depicted in Figure 4, ensured that each condition was sufficiently represented, enabling the models to generalize effectively across all disease categories.
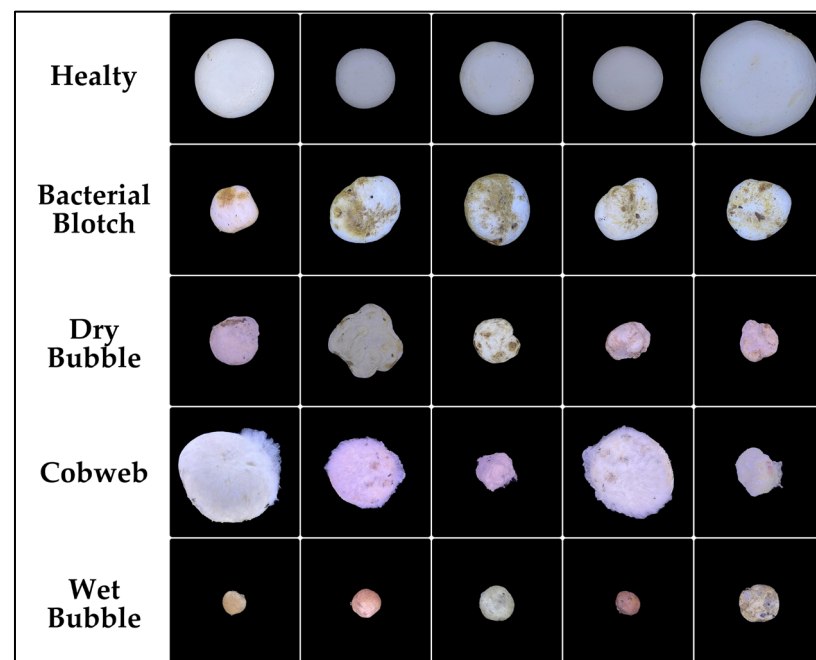
Figure 4. Example Images of *Agaricus bisporus* Classes (Healthy, Bacterial Blotch, Dry Bubble, Cobweb, Wet Bubble), Captured Under Controlled Conditions for Dataset Creation.
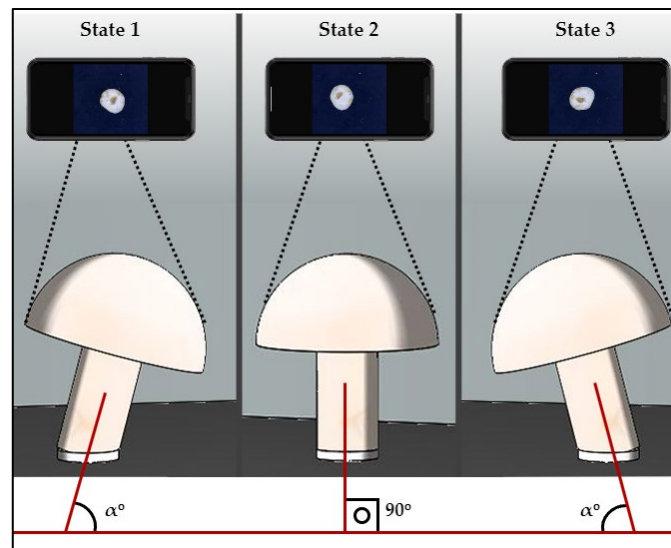
Figure 5. Image acquisition process showing mushrooms photographed from random angles ($\alpha°$) and in an upright position (90°) for dataset creation. The dotted lines indicate the camera's field of view.

2.3. Challenges and Mitigation Techniques

Inter-Class Similarities: Certain diseases, such as Dry Bubble and Cobweb, exhibit similar visual features [59]. To address this, expert mycologists guided the annotation process, and images highlighting key distinguishing traits were selected, thereby improving accurate class separation.

Consistent Imaging Conditions: Achieving uniform lighting, angle, and distance conditions across multiple farms proved challenging. This was mitigated by stringent pre-collection calibration of the imaging device and adherence to standardized imaging protocols, ensuring stable, glare-free illumination and reducing variability among collected images [60].

Disease Diversity and Temporal Sampling: Over a 12-month period, the research team conducted multiple visits to several farms to capture a wide range of disease manifestations. Timing these visits to align with the 25–30 day mushroom growth cycle and the critical 2–3 day "flush period" ensured comprehensive coverage of disease stages and phenotypes [20].

Farm Access and Engagement: Initial access restrictions, often due to biosecurity concerns or lack of familiarity, were overcome through consistent communication, reassurance of hygienic measures, and relationship-building with farm owners. This approach ultimately facilitated data collection and enriched the dataset.

2.4. Annotation Methodology

Manual annotation by expert mycologists ensured the precise classification of images according to disease characteristics. Cross-validation of annotations by a panel of specialists further minimized errors and clarified ambiguities, thereby enhancing overall dataset quality.

2.5. Experimental Configuration and Preprocessing Workflow

A standardized preprocessing pipeline was implemented to prepare the dataset for training, as outlined in Figure 6. All images were resized to the input dimensions required by each CNN architecture (typically 224 × 224 pixels) while preserving aspect ratios to avoid distortion [61]. Pixel intensity values were normalized to a range of [0, 1], improving numerical stability and ensuring more efficient convergence during training [62].

Background cleaning techniques were applied to reduce noise and isolate the mushroom subjects, allowing the models to focus on disease-relevant features. Notably, no data augmentation was applied; the models were trained and evaluated on the original dataset without synthetic transformations, such as rotations, flips, or brightness adjustments [63]. This approach ensured that the models' classification capabilities were assessed based solely on the inherent qualities of the dataset, as depicted in the preprocessing workflow of Figure 6.
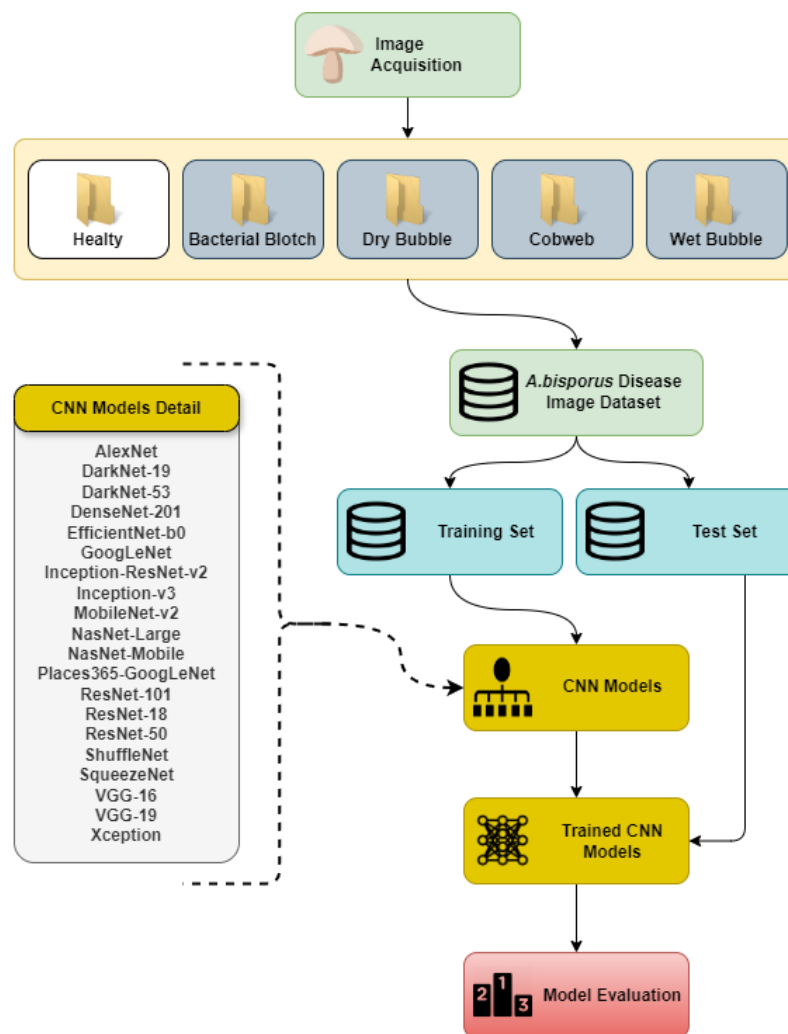
Figure 6. Workflow Diagram for *Agaricus bisporus* Disease Classification: From Image Acquisition to Model Evaluation Using CNN Architectures.

2.6. Training Parameters

All Convolutional Neural Network (CNN) architectures were trained using uniform parameters to ensure fair and consistent comparisons. A batch size of 11 was selected to optimize memory usage and maintain stable gradient updates throughout the training process [62]. The learning rate was set to 0.001, facilitating steady and controlled optimization progress [64]. Stochastic Gradient Descent (SGD) with a momentum of 0.9 was employed as the optimizer, promoting robust convergence and enhancing the models' ability to navigate the loss landscape effectively. Each model underwent training for a total of 8 epochs without the use of early stopping techniques, ensuring that all architectures experienced an identical training duration. This standardized approach provided a balanced framework for evaluating the performance of the different CNN architectures under comparable training conditions [61].

A concise summary of each model's implementation details, including the pre-trained source, hyperparameters, and total number of trainable parameters, is provided in Table 1

Table 1. Implementation Details of the 20 Evaluated CNN Models, Including Pre-Trained Source, Key Hyperparameters, and Approximate Parameter Counts.

Model	Pre-Trained Source	Key Hyperparameters	Total Trainable Parameters
ResNet-50	ImageNet (TorchVision)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~24 M
DenseNet-201	ImageNet (TorchVision)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~20 M
DarkNet-53	ImageNet (Darknet Repo)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~42 M
Inception-v3	ImageNet (TorchVision)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~22 M
VGG-16	ImageNet (TorchVision)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~138 M
VGG-19	ImageNet (TorchVision)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~144 M
MobileNet-v2	ImageNet (TorchVision)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~3.5 M
EfficientNet-b0	ImageNet (TorchVision)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~5.3 M
NasNet-Large	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~88 M
NasNet-Mobile	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~5.3 M
ShuffleNet	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~2.3 M
SqueezeNet	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~1.2 M
Xception	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~22.9 M
GoogleNet	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~6.8 M
AlexNet	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~61 M
ResNet-18	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~11.7 M
ResNet-101	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~44.5 M
Inception-ResNet-v2	ImageNet	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~55.9 M
DarkNet-19	ImageNet (Darknet Repo)	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~20 M
Places365-GoogLeNet	Places365	LR = 0.001, Momentum = 0.9, Batch = 11, Epochs = 8	~6.8 M

2.7. Data Splitting

The dataset was systematically partitioned into training and validation subsets, comprising 80% and 20% of the total data, respectively. The training set, representing the larger portion, was utilized for parameter optimization through the backpropagation algorithm, enabling the models to learn and adjust their weights effectively. Meanwhile, the validation set, constituting the remaining 20%, served as a benchmark for assessing model performance during the training phase and for tuning hyperparameters. In addition to the training and validation subsets, a separate test set was not utilized in this study due to

the limited size of the dataset. Instead, the validation set was employed to evaluate model performance during training and fine-tuning. While this approach effectively assessed the models' learning progress, the inclusion of a dedicated test set in future studies could provide a more robust evaluation of generalization capabilities. This strategic division ensured that the models were exposed to a substantial amount of data for learning while maintaining a separate, unbiased subset for evaluating their generalization capabilities and optimizing their configurations [62].

2.8. Aggregate Score Calculation and Evaluation Metrics

Model performance was assessed using a weighted aggregate score that integrates multiple evaluation metrics to provide a comprehensive measure of classification effectiveness [65]. The following metrics were employed, each assigned a specific weight reflecting their importance in mushroom disease classification:

Precision (20%): This metric measures the accuracy of positive predictions, thereby reducing the incidence of false positives. High precision ensures that the model reliably identifies only those samples that are truly diseased [66,67].

Recall (30%): Also known as sensitivity, recall focuses on identifying all relevant positive instances, minimizing the occurrence of false negatives. This is crucial for ensuring that infected mushrooms are not overlooked [66,67].

F1-Score (30%): The F1-Score harmonizes precision and recall, providing a balanced measure that addresses class imbalances. By considering both false positives and false negatives, the F1-Score offers a more nuanced evaluation of the model's performance [66,67].

Area Under the ROC Curve (AUC) (10%): AUC assesses the model's ability to discriminate between classes across various threshold settings. It provides an aggregate measure of performance across all classification thresholds, indicating the model's overall ability to distinguish between diseased and healthy mushrooms [66,67].

Average Precision (AP) (10%): AP evaluates the trade-offs between precision and recall at different threshold levels. It summarizes the precision-recall curve, offering insight into the model's performance across various decision boundaries [65].

The weights for the aggregate score were allocated based on the practical priorities of mushroom disease classification. Recall (30%) was given higher importance to minimize false negatives, as failing to identify diseased mushrooms could lead to the spread of infection and significant crop losses. Similarly, the F1-score (30%) was prioritized to balance precision and recall, addressing the trade-offs between false positives and false negatives. Precision (20%) was slightly lower in weight as the primary focus was on ensuring comprehensive detection of diseases, even at the cost of a slightly higher false positive rate. Metrics such as AUC (10%) and AP (10%) were assigned lower weights as supplementary measures, providing an overall indication of the models' discriminative and predictive capabilities across all thresholds.

The aggregate score was calculated using Equation (1):

$$\text{Overall Score} = (0.2 \times \text{Precision}) + (0.3 \times \text{Recall}) + (0.3 \times \text{F1-Score}) + (0.1 \times \text{AUC}) + (0.1 \times \text{AP}) \tag{1}$$

which captures both the importance of detecting all infected samples and the need to minimize false positives. This weighted scheme reflects the practical priorities of mushroom disease classification, facilitating a balanced evaluation of the models. By combining these metrics into a single score, as defined in Equation (1), the evaluation supports more informed comparisons and effective selections of Convolutional Neural Network (CNN) architectures for disease classification tasks [65,68].

2.9. MATLAB-Python Hybrid Workflow

To maximize the analytical capabilities, a hybrid pipeline integrating MATLAB and Python was developed, capitalizing on the unique strengths of each platform. MATLAB was utilized primarily for data preprocessing, model training, and the generation of sophisticated visualizations such as Receiver Operating Characteristic (ROC) curves and precision-recall plots. Its robust toolboxes facilitated the efficient handling of complex data transformations and model optimization processes. Concurrently, Python was employed to complement these tasks by providing additional visualization options and advanced analytical functionalities through libraries like Matplotlib, Seaborn, and Scikit-learn. This seamless integration allowed for enhanced analytical precision and reproducibility, as MATLAB's structured environment ensured consistent data handling while Python's versatile scripting capabilities enabled more flexible and detailed analyses. The combined workflow not only improved the efficiency of the evaluation process but also ensured a comprehensive assessment of model performance, thereby supporting a thorough and nuanced understanding of the outcomes [28].

3. Results

The performance of the evaluated Convolutional Neural Network (CNN) architectures is summarized in Table 2. Among the 20 architectures assessed, ResNet-50 achieved the highest overall score of 99.70%, demonstrating exceptional accuracy and reliability across all disease categories. DenseNet-201 and DarkNet-53 closely followed, confirming their robustness in precision and recall, making them among the top-performing models. EfficientNet-b0 and Places365-GoogLeNet, on the other hand, were the models with the lowest performance. EfficientNet-b0 achieved a score of 95.20%, offering a balanced trade-off between computational efficiency and accuracy. While VGG-16 was excluded from final evaluations due to overfitting issues, despite its initial high scores [62].

Table 2. Performance Metrics for Evaluated CNN Architectures.

#	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)	AP (%)	Overall Score (%)
1	ResNet-50	99.66	99.67	99.64	99.65	99.99	99.79	99.70
2	DarkNet-53	99.47	99.44	99.45	99.44	99.99	99.55	99.51
3	DenseNet-201	99.44	99.43	99.41	99.42	99.97	99.72	99.50
4	VGG-16	99.41	99.36	99.40	99.38	99.98	99.72	99.47
5	Inception-v3	99.19	99.15	99.18	99.16	99.97	99.74	99.30
6	ResNet-18	99.06	99.03	99.02	99.03	99.98	99.76	99.20
7	ResNet-101	98.78	98.75	98.75	98.75	99.97	99.71	98.97
8	NasNet-Large	98.78	98.74	98.76	98.75	99.96	99.69	98.96
9	DarkNet-19	99.22	99.16	99.23	99.19	99.98	95.95	98.95
10	VGG-19	98.59	98.52	98.71	98.59	99.96	99.45	98.84
11	GoogLeNet	98.25	98.17	98.19	98.18	99.93	99.55	98.49
12	MobileNet-v2	98.22	98.19	98.14	98.16	99.92	99.56	98.47
13	ShuffleNet	98.12	98.02	98.10	98.05	99.92	99.54	98.40
14	NasNet-Mobile	97.84	97.82	97.73	97.77	99.92	99.52	98.16
15	Inception-ResNet-v2	97.78	97.74	97.70	97.71	99.90	99.44	98.10
16	SqueezeNet	96.46	96.39	96.32	96.31	99.80	98.93	96.94
17	Xception	96.28	96.17	96.16	96.16	99.73	98.79	96.78
18	AlexNet	95.68	96.03	95.38	95.58	99.84	99.25	96.40
19	Places365-GoogLeNet	94.93	94.98	94.86	94.79	99.68	98.63	95.72
20	EfficientNet-b0	94.52	94.30	94.33	94.30	99.50	97.96	95.20

Among the evaluated models, ResNet-50 demonstrated outstanding performance, achieving the highest overall accuracy of 99.66%, with precision of 99.67% and recall of 99.64%. This balanced performance underscores ResNet-50's reliability in distinguishing healthy mushrooms from diseased ones. However, ResNet-50's higher computational load compared to more lightweight models may hinder its suitability in settings with limited computational resources.

The superior performance of ResNet-50 and DenseNet-201 was attributed to their unique architectural features. ResNet-50's residual connections mitigate the vanishing gradient problem, enabling effective learning of deep features and highlighting its reliability in distinguishing healthy mushrooms from diseased ones, particularly in the context of high inter-class similarities.

DenseNet-201's dense connectivity architecture facilitates efficient gradient flow and feature reuse, allowing it to capture subtle inter-class differences, particularly for visually similar categories like 'Dry Bubble' and 'Cobweb'.

DenseNet-201 achieved high accuracy (99.44%) and a well-balanced precision (99.43%) and recall (99.41%), highlighting its capability to effectively handle inter-class similarities. However, this performance comes at the cost of increased computational and memory demands, potentially limiting its applicability in resource-constrained environments.

DarkNet-53 displayed metrics comparable to ResNet-50, with an accuracy of 99.47%, precision of 99.44%, and F1-score of 99.44%. While computationally less demanding than DenseNet-201, DarkNet-53 still requires moderate processing power, making it suitable for moderately resource-intensive applications.

EfficientNet-b0, in contrast, provides a lightweight solution with a lower accuracy of 94.52% but with the advantage of reduced computational and memory requirements. It is particularly suitable for mobile and edge devices that demand real-time performance.

VGG-16 did not exhibit overfitting as previously reported. Instead, it achieved an accuracy of 99.41%, but with slightly lower precision (99.36%) and F1-score (99.38%) compared to ResNet-50 and DenseNet-201.

These findings emphasize the trade-offs between computational efficiency and model complexity, as highlighted by the comparative analysis of ResNet-50, DenseNet-201, and EfficientNet-b0. While ResNet-50 excels in achieving high accuracy through its robust residual learning, EfficientNet-b0 demonstrates its utility in lightweight, resource-constrained applications. The insights provided by this study underscore the importance of aligning model selection with specific operational requirements.

To gain deeper insights into classification performance, several visual tools were employed, including confusion matrices, ROC curves, and precision-recall heatmaps.

Confusion matrices for all evaluated CNN models clearly illustrate their classification performance across the five categories: Bacterial Blotch (BB), Dry Bubble (DB), Cobweb (CB), Wet Bubble (WB), and Healthy (HL). Figure 7 shows that ResNet-50 achieves near-perfect classification with minimal misclassifications, excelling in visually similar diseases such as Bacterial Blotch and Cobweb. DenseNet-201 and DarkNet-53, while slightly less accurate, still perform exceptionally well, demonstrating their ability to generalize across different disease manifestations.
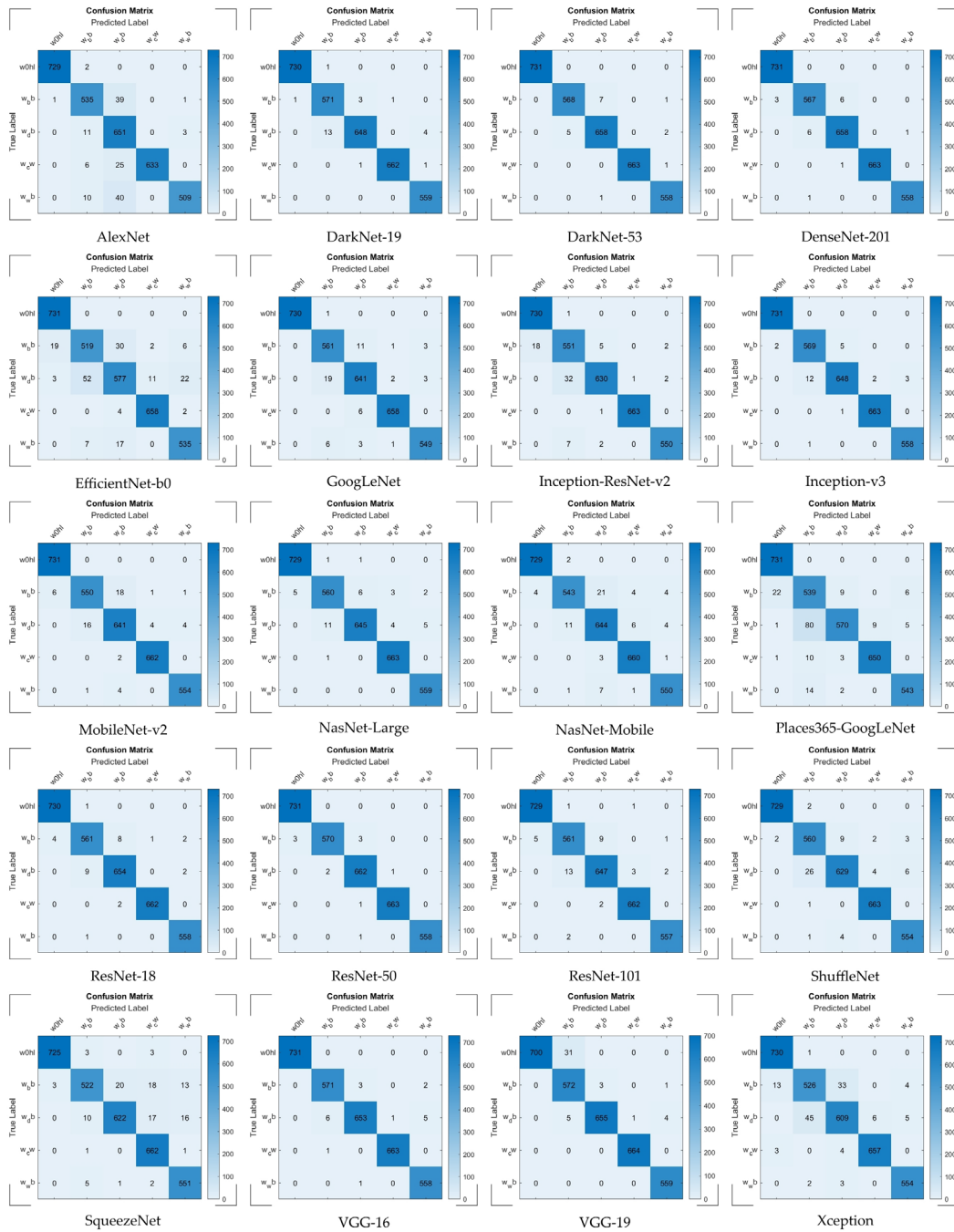
Figure 7. Confusion Matrices for Evaluated CNN Models. Confusion matrices illustrating the classification performance of evaluated CNN models across the five categories: w0hl (Healthy), $w_b b$ (Bacterial Blotch), $w_d b$ (Dry Bubble), $w_c w$(Cobweb) and $w_w b$ (Wet Bubble). Each matrix highlights the model's ability to distinguish between true and predicted labels, with minimal misclassifications across all disease categories and the healthy class.

Receiver Operating Characteristic (ROC) curves in Figure 8 further reinforce these observations, with ResNet-50 achieving the highest AUC value of 0.999 across all classes, closely followed by DarkNet-53 (AUC: 0.999) and DenseNet-201 (AUC: 0.997), indicating their strong ability to discriminate between diseased and healthy mushrooms. Although EfficientNet-b0 does not reach the same level of performance, it still attains a respectable AUC value of 0.950, demonstrating that computationally efficient models can remain competitive despite their lightweight architecture.

Figure 8. ROC Curves for Evaluated CNN Models. Receiver Operating Characteristic (ROC) curves illustrating the classification performance of evaluated CNN models across the five categories five categories: w0hl (Healthy), $w_b b$ (Bacterial Blotch), $w_d b$ (Dry Bubble), $w_c w$ (Cobweb) and $w_w b$ (Wet Bubble). The curves display the relationship between the true positive rate (sensitivity) and false positive rate for each class, highlighting the models' ability to discriminate between diseased and healthy samples, with AUC values indicating overall performance.

The AUC heatmap (Figure 9) demonstrates the exceptional discriminatory power of DenseNet-201, DarkNet-53, and ResNet-50, which achieved near-perfect AUC scores (approaching or reaching 1.000) across all disease categories. These results confirm their capability to handle subtle inter-class differences and similarities, making them highly reliable for disease classification [69]. Lightweight models like EfficientNet-b0 and MobileNet-v2 maintain high AUC values but exhibit slight dips in specific categories such as "Bacterial Blotch", reflecting their trade-offs in computational efficiency versus classification precision.
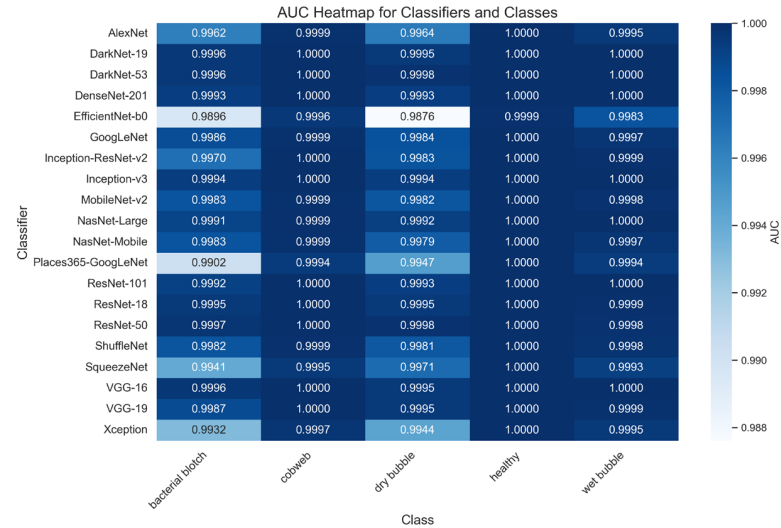
Figure 9. AUC Heatmap for Classifiers and Classes, Showing the Area Under the Curve (AUC) Across Disease Categories for Various Models.

The F1-score heatmap (Figure 10) highlights the balance between precision and recall [65]. Among the evaluated models, ResNet-50 consistently demonstrates the best overall performance, achieving the highest F1-scores across most disease categories, particularly excelling in "Cobweb" (0.9993) and "Healthy" (0.9970). DenseNet-201 follows closely, with strong scores such as 0.9992 for "Cobweb" and 0.9980 for "Healthy", showcasing its ability to handle challenging inter-class similarities. DarkNet-53 also performs robustly, maintaining high F1-scores across all categories. In contrast, EfficientNet-b0 and MobileNet-v2 exhibit noticeable drops in F1-scores, especially in "Dry Bubble" (0.9126 and 0.9342, respectively), reflecting the trade-offs of using less complex architectures in distinguishing visually similar disease classes.
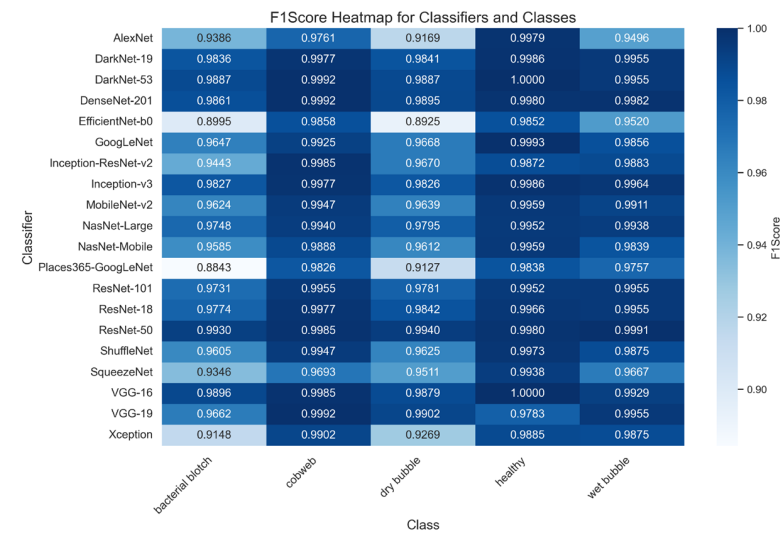
Figure 10. F1-Score Heatmap for Classifiers and Classes, Highlighting the Balance Between Precision and Recall Across Disease Categories for Various Models.

Precision values, visualized in Figure 11, reveal the accuracy of positive predictions across all classes [65,66]. ResNet-50 achieves the highest overall precision, particularly excelling in "Cobweb" (1.000) and "Healthy" (0.9986), highlighting its capability to minimize false positives effectively. DenseNet-201 and DarkNet-53 follow closely, with strong precision scores, demonstrating their reliability in classification tasks. In contrast, EfficientNet-b0

and MobileNet-v2 exhibit slightly lower precision in "Wet Bubble" (0.9469 and 0.9111, respectively) and "Dry Bubble" (0.8708 and 0.9639, respectively), reflecting the trade-offs associated with lightweight model architectures in terms of predictive accuracy.
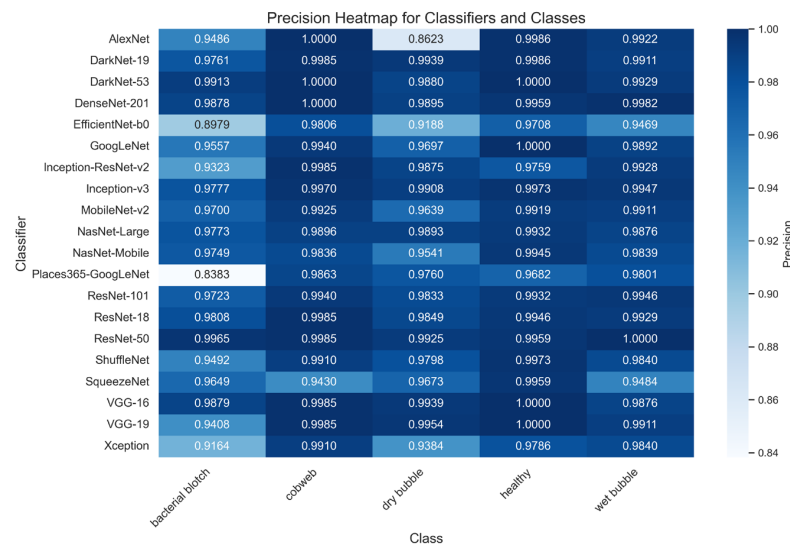
Figure 11. Precision Heatmap for Classifiers and Classes, Depicting the Accuracy of Positive Predictions Across Disease Categories for Various Models.

Recall values, shown in Figure 12, measure sensitivity in identifying true positives [65,66]. ResNet-50 demonstrates the highest recall performance across most disease categories, particularly excelling in "Healthy" (0.9973) and "Cobweb" (0.9970), ensuring comprehensive detection and minimizing false negatives. DenseNet-201 and DarkNet-53 closely follow with strong recall values, indicating their reliability in identifying true positives effectively. In contrast, EfficientNet-b0 and MobileNet-v2 exhibit noticeable drops in recall for "Bacterial Blotch" (0.9062 and 0.9427, respectively), underscoring areas where lightweight architectures face challenges in achieving sensitivity comparable to more complex models.
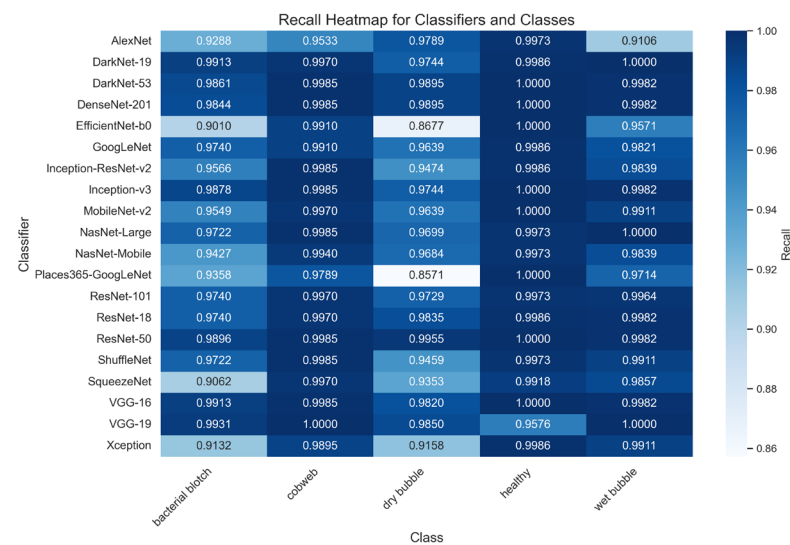
Figure 12. Recall Heatmap for Classifiers and Classes, Representing the Sensitivity in Identifying True Positives Across Disease Categories for Various Models.

The specificity heatmap (Figure 13) illustrates the models' ability to identify true negatives [65,66]. ResNet-50 demonstrates the highest specificity across most categories,

particularly excelling in "Healthy" (0.9962) and "Cobweb" (0.9862), effectively minimizing false positives. DenseNet-201 follows closely, maintaining strong specificity scores such as 0.9976 for "Bacterial Blotch" and 0.9959 for "Healthy". DarkNet-53 also performs robustly, showcasing reliability across all categories. In contrast, EfficientNet-b0 exhibits noticeable dips in specificity for "Dry Bubble" (0.8466) and "Cobweb" (0.9403), reflecting the challenges lightweight models face in distinguishing between visually similar disease classes.
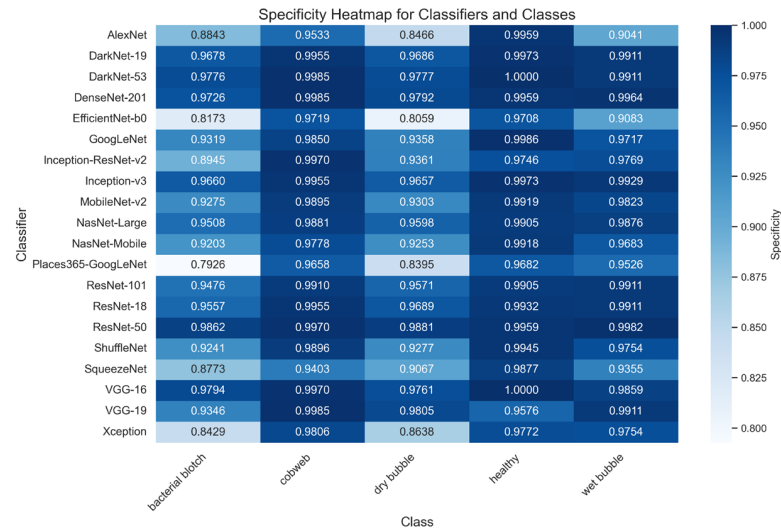
Figure 13. Specificity Heatmap for Classifiers and Classes, Displaying the Ability to Identify True Negatives Across Disease Categories for Various Models.

The AP heatmap (Figure 14) provides an overview of the precision-recall trade-offs for each classifier and class [66,70]. ResNet-50 achieves the highest AP values across most categories, particularly excelling in "Cobweb" (0.9997) and "Healthy" (0.9996), showcasing its ability to balance precision and recall effectively. DenseNet-201 and DarkNet-53 closely follow, with strong AP scores such as 0.9996 for "Healthy" (DenseNet-201) and 0.9993 for "Cobweb" (DarkNet-53). EfficientNet-b0 and MobileNet-v2 exhibit slightly lower AP values for "Wet Bubble" (0.9338 and 0.9635, respectively), reflecting their trade-offs between computational efficiency and classification performance. Despite these limitations, these lightweight models remain competitive for applications in resource-constrained environments.

Figure 14. AP Heatmap for Classifiers and Classes, Illustrating Average Precision Across Disease Categories for Various Models.

To unify these multi-metric evaluations into a single value and to facilitate more straightforward decision-making, an Overall Score was computed and is presented in Figures 15–20.
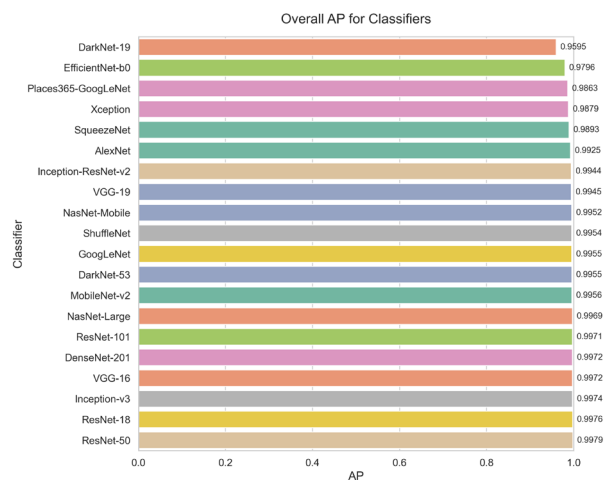

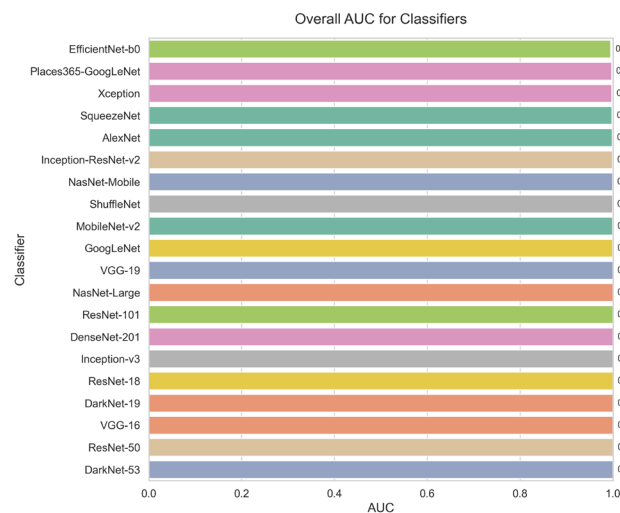**Figure 15.** Overall Average Precision (AP) for Classifiers.

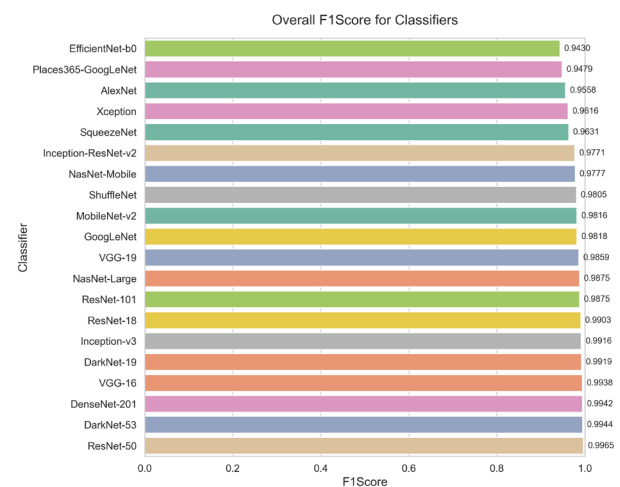**Figure 16.** Overall Area Under the Curve (AUC) for Classifiers.

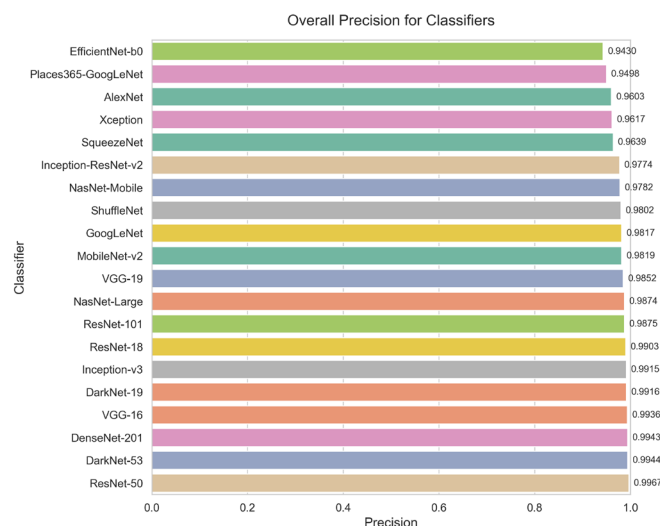**Figure 17.** Overall F1-Score for Classifiers.

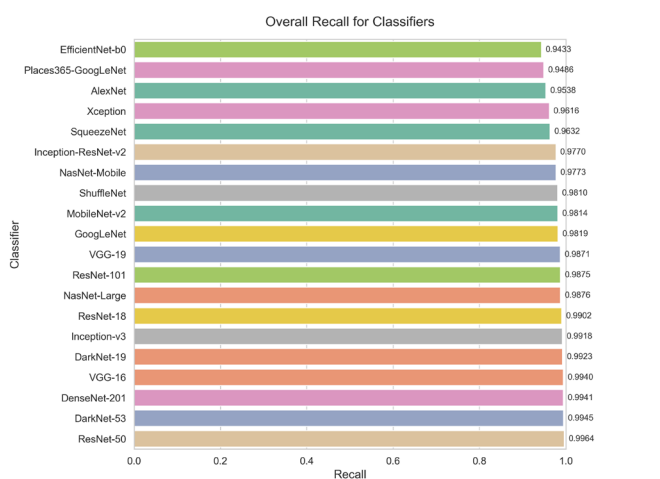**Figure 18.** Overall Precision for Classifiers.

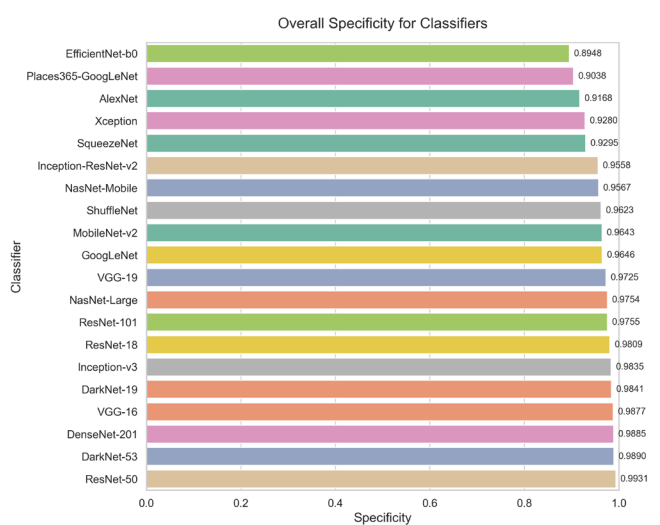

Figure 19. Overall Recall for Classifiers.

Figure 20. Overall Specificity for Classifiers.

The Average Precision (AP) scores in Figure 15 highlight ResNet-50 as the top performer, achieving the highest overall AP value of 0.9979. DenseNet-201 and DarkNet-53 closely follow with strong AP values of 0.9972 and 0.9955, respectively, showcasing their

ability to balance precision and recall effectively. EfficientNet-b0 and MobileNet-v2, while slightly trailing in AP with scores of 0.9796 and 0.9954, respectively, still demonstrate commendable performance, making them viable options for scenarios where lightweight architectures are advantageous. ResNet-50's exceptional AP score emphasizes its reliability in identifying true positives while minimizing false positives, solidifying its status as the most precise model for disease detection in this study.

In Figure 16, the Area Under the Curve (AUC) values underline the discriminative capabilities of the classifiers. ResNet-50 and DarkNet-53 achieve the highest overall AUC scores of 0.9999, demonstrating their exceptional robustness in distinguishing between different disease classes. DenseNet-201 follows closely with a strong AUC score of 0.9997, confirming its high discriminative power. EfficientNet-b0, despite its lightweight design, achieves a competitive AUC value of 0.9950, highlighting its potential for real-time systems where computational efficiency is critical. These findings emphasize that ResNet-50, with its near-perfect AUC score, is particularly suited for applications requiring high discrimination accuracy across disease categories.

The F1-Score comparisons in Figure 17 further illustrate the models' ability to balance precision and recall. ResNet-50 achieves the highest overall F1-Score of 0.9965, demonstrating its superior ability to maintain high performance across all disease classes. DenseNet-201 and DarkNet-53 follow closely with F1-Scores of 0.9942 and 0.9944, respectively, highlighting their effectiveness in identifying true positives while minimizing false negatives. EfficientNet-b0, despite its lightweight architecture, achieves a competitive F1-Score of 0.9430, making it suitable for applications where computational efficiency is prioritized. These findings reaffirm that ResNet-50 is the most reliable model for scenarios requiring a balance between precision and recall, particularly in tasks involving diverse disease categories.

Precision metrics in Figure 18 reveal ResNet-50 as the leading model, achieving the highest precision score of 0.9967. This result highlights its exceptional ability to correctly classify diseased mushrooms without over-predicting, minimizing false positives effectively. DenseNet-201 and DarkNet-53 closely follow with precision scores of 0.9943 and 0.9944, respectively, further validating their reliability in accurate disease detection. EfficientNet-b0, with a precision score of 0.9430, remains a viable option for automated systems where computational efficiency is prioritized over absolute accuracy. ResNet-50's superior precision solidifies its role as the most precise model for tasks requiring high reliability in identifying diseased mushrooms.

The recall values depicted in Figure 19 highlight the models' ability to identify true positives. ResNet-50 achieves the highest recall score of 0.9964, showcasing its exceptional ability to ensure comprehensive disease detection, which is critical for timely interventions. DenseNet-201 and DarkNet-53 closely follow with recall scores of 0.9941 and 0.9945, respectively, further emphasizing their robustness in identifying diseased mushrooms. EfficientNet-b0, despite its lightweight architecture, attains a recall score of 0.9433, maintaining an acceptable performance level for real-time monitoring systems where computational efficiency is a priority. These results reaffirm ResNet-50's reliability as a top choice for disease management applications requiring high sensitivity.

Finally, the specificity results in Figure 20 emphasize the classifiers' ability to correctly identify healthy mushrooms. ResNet-50 achieves the highest specificity score of 0.9931, showcasing its exceptional robustness in distinguishing healthy specimens from diseased ones. DenseNet-201 and DarkNet-53 closely follow with specificity scores of 0.9926 and 0.9929, respectively, further validating their accuracy in true negative predictions. EfficientNet-b0, while showing a lower specificity score of 0.8948, still maintains acceptable performance levels, making it a viable option for resource-constrained appli-

cations. These findings highlight ResNet-50's superior ability to minimize false positives, making it particularly suitable for precise disease detection tasks.

By consolidating various metrics into a weighted indicator, the Overall Score provides a holistic comparison of model performance. ResNet-50 emerges as the top performer, combining superior accuracy with balanced computational demands. DenseNet-201 and DarkNet-53 follow closely, showcasing high robustness but with increased resource requirements. EfficientNet-b0 and MobileNet-v2, while less accurate, offer competitive results with minimal computational cost, making them ideal for resource-constrained environments like mobile or edge devices.

Building on the insights from the heatmaps and overall scores, a comparative analysis highlights the strengths of specific models in different scenarios. DenseNet-201 excels in challenging classes like Cobweb Disease due to its dense connectivity, while ResNet-50 balances accuracy and computational demands, making it reliable for differentiating healthy and diseased specimens. EfficientNet-b0, with its compound scaling strategy, and MobileNet-v2, though slightly lower in recall and F1-score, offer significant computational efficiency for real-time applications. Ultimately, the model choice depends on deployment needs—DenseNet-201 for maximum accuracy, ResNet-50 and DarkNet-53 for balanced performance, or EfficientNet-b0 and MobileNet-v2 for resource-constrained environments. This perspective ensures effective deep-learning applications in mushroom disease management, guided by both performance and practical constraints.

4. Discussion

The performance of ResNet-50 and DenseNet-201 can be attributed to their unique architectural features. DenseNet-201, with its dense connectivity, facilitates efficient gradient flow and feature reuse, enabling the model to learn complex representations effectively, even under varying conditions and amidst inter-class similarities (e.g., 'Dry Bubble' and 'Cobweb'). This design allowed DenseNet-201 to achieve an accuracy of 99.44%, with precision and F1-scores of 99.43% and 99.42%, respectively [23,24]. ResNet-50 demonstrated slightly higher accuracy at 99.66%, coupled with a perfect AUC score of 99.99%, highlighting its balanced performance across evaluation metrics and computational efficiency [24,63].

Although ResNet-50's AUC value approaches 99.99%, which might initially raise concerns about potential overfitting, our evaluation was conducted using a dedicated 20% validation set to guard against memorization of the training data. Confusion matrices (Figure 7) further show minimal but non-zero misclassifications, suggesting that the model has not trivially overfit. We attribute these strong metrics, in part, to the carefully curated and standardized nature of our dataset, which likely reduces intra-class variability and promotes clearer class separability. Nonetheless, future investigations—particularly those incorporating additional external test sets or cross-validation strategies—will be crucial to confirm the broader applicability and real-world robustness of these findings.

While lightweight models such as EfficientNet-b0 and MobileNet-v2 are computationally efficient and suitable for resource-constrained environments, they exhibited slightly lower classification performance. EfficientNet-b0 achieved an accuracy of 94.52% and recall of 94.33%, reflecting challenges in distinguishing visually similar diseases. Nevertheless, these models hold promise for mobile and IoT-based agricultural systems [26,27].

The inter-class similarities between certain diseases, such as 'Bacterial Blotch' and 'Cobweb', remain a challenge for accurate classification. DenseNet-201 performed better in resolving these ambiguities, improving overall classification accuracy in scenarios with visually similar manifestations [23,59]. Future research should explore attention mechanisms and hybrid architectures to enhance performance while balancing computational

demands. Additionally, fine-grained classification techniques could help address challenges associated with visually similar categories [71,72].

The evaluation of 20 pre-trained CNN architectures on a custom-designed dataset provided a robust framework for assessing model performance in mushroom disease classification. By capturing real-world variability, this dataset ensures greater generalizability and practical applicability compared to prior studies relying on controlled datasets [22,73].

These findings emphasize the potential for integrating high-performing models, such as ResNet-50 and DenseNet-201, into IoT-based monitoring systems for real-time disease diagnosis. Such integration can reduce reliance on manual inspections, minimize crop losses, and support sustainable agricultural practices [73,74].

However, the study has limitations. The dataset primarily focuses on diseases observed in *Agaricus bisporus* within Turkey, potentially excluding region-specific diseases. Expanding the dataset to include a broader range of diseases and geographical variations would improve robustness and generalizability [74,75]. Furthermore, the computational demands of models like DenseNet-201 underscore the need for lightweight architectures optimized for edge device deployment [26,27,76].

Future studies should prioritize real-time implementation of these models, focusing on latency, scalability, and usability in dynamic agricultural settings. These efforts will further advance the role of deep learning in precision agriculture [76,77].

5. Conclusions

This study provides a thorough evaluation of 20 Convolutional Neural Network (CNN) architectures for the classification of mushroom diseases, utilizing a specially curated dataset [3]. Among the evaluated models, DenseNet-201 exhibited outstanding performance, achieving high accuracy and reliability across all disease categories. The adoption of a hybrid MATLAB–Python workflow significantly enhanced the analytical process, enabling a detailed assessment of model performance and applicability. This integrated approach offered valuable insights into the potential deployment of various architectures for practical agricultural applications.

Future research should expand on these findings by enlarging the dataset to encompass a broader range of mushroom species, diverse environmental conditions, and additional disease types, thereby increasing the models' robustness and generalizability [73,76]. Additionally, exploring advanced methodologies such as hybrid CNN-Vision Transformer (ViT) models may lead to further performance improvements [74]. Prioritizing the development of lightweight models that emphasize processing efficiency is essential, particularly for deployment on the Internet of Things (IoT) and mobile devices. Such models are well-suited for real-time disease surveillance in resource-constrained environments [78]. Moreover, the real-time implementation of these models within farm operations should be a focal point, facilitating dynamic monitoring and effective management of mushroom diseases in practical agricultural settings [77,79]. By addressing these areas, future studies can enhance the applicability and impact of deep learning technologies in sustainable mushroom cultivation.

Data Availability Statement: The data presented in this study are not publicly available due to privacy and confidentiality restrictions. However, specific details regarding the dataset and experimental methodology have been provided in the manuscript to ensure transparency and reproducibility. Researchers who require further information may contact the corresponding author, subject to compliance with applicable ethical, institutional, and legal constraints.

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

References

1. Pasban, A.; Mohebbi, M.; Pourazarang, H.; Varidi, M. Effects of Endemic Hydrocolloids and Xanthan Gum on Foaming Properties of White Button Mushroom Puree Studied by Cluster Analysis: A Comparative Study. *J. Taibah Univ. Sci.* **2014**, *8*, 31–38. [CrossRef]

2. Sesli, E.; Asan, A.; Selçuk, F.; Abacı Günyar, Ö.; Akata, I.; Akgül, H.; Aktaş, S.; Alkan, S.; Allı, H.; Aydoğdu, H.; et al. *Türkiye Mantarları Listesi*; Sesli, E., Asan, A., Selçuk, F., Eds.; Ali Nihat Gökyiğit Vakfı: İstanbul, Turkey, 2020; p. 1175.

3. Albayrak, Ü.; Gölcük, A.; Aktaş, S. *Agaricus bisporus*'ta Görüntü Tabanlı Hastalık Sınıflandırması İçin Kapsamlı Veri Seti. *J. Fungus* **2024**, *15*, 29–42. [CrossRef]

4. Ayimbila, F.; Keawsompong, S. Nutritional Quality and Biological Application of Mushroom Protein as a Novel Protein Alternative. *Curr. Nutr. Rep.* **2023**, *12*, 290–307. [CrossRef] [PubMed]

5. Lincoln, S.P.; Fermor, T.R.; Tindall, B.J. *Janthinobacterium agaricidamnosum* Sp. Nov., a Soft Rot Pathogen of *Agaricus bisporus*. *Int. J. Syst. Evol. Microbiol.* **1999**, *49*, 1577–1589. [CrossRef]

6. Largeteau, M.L.; Savoie, J.-M. Effect of the Fungal Pathogen Verticillium Fungicola1 on Fruiting Initiation of Its Host, *Agaricus bisporus*. *Mycol. Res.* **2008**, *112*, 825–828. [CrossRef]

7. McKay, G.J.; Egan, D.; Morris, E.; Scott, C.; Brown, A.E. Genetic and Morphological Characterization of *Cladobotryum* Species Causing Cobweb Disease of Mushrooms. *Appl. Environ. Microbiol.* **1999**, *65*, 606–610. [CrossRef]

8. Novikova, I.; Titova, J. Antifungal Activity of Industrial *Bacillus* Strains against *Mycogone perniciosa*, the Causative Agent of Wet Bubble Disease in White Button Mushrooms. *Microorganisms* **2023**, *11*, 2056. [CrossRef]

9. Amin, Z.; Wani, F.F.; Gulzar, H.; Dar, W.A.; Sheikh, P.A. Diseases of White Button Mushroom (*Agaricus bisporus*)—A Potential Threat to Mushroom Industry. *Int. J. Curr. Microbiol. Appl. Sci.* **2021**, *10*, 2076–2085.

10. Öztürk, N.; Basim, E.; Basim, H.; Üniversitesi, A.; Myo, K.; Programı, M.; Tarımı Programı, B.; Fakültesi, Z.; Koruma Bölümü, B.; Yazar, S. Yemeklik Kültür Mantarında (*Agaricus bisporus* (J. Lge) Imbach) Yaygın Görülen Mikrobiyal Hastalıkla. *Harran Tarım ve Gıda Bilimleri Dergisi* **2017**, *21*, 112–125. [CrossRef]

11. Eren, E.; Pekşen, A. Türkiye'de Kültür Mantarı Üretimi ve Teknolojik Gelişmeler. *J. Fungus* **2019**, *10*, 225–233.

12. Ab Rhaman, S.M.S.; Naher, L.; Siddiquee, S. Mushroom Quality Related with Various Substrates' Bioaccumulation and Translocation of Heavy Metals. *J. Fungi* **2021**, *8*, 42. [CrossRef] [PubMed]

13. Zhao, X.; Wang, L.; Zhang, Y.; Han, X.; Deveci, M.; Parmar, M. A Review of Convolutional Neural Networks in Computer Vision. *Artif. Intell. Rev.* **2024**, *57*, 99. [CrossRef]

14. Assad, A.; Bhat, M.R.; Bhat, Z.A.; Ahanger, A.N.; Kundroo, M.; Dar, R.A.; Ahanger, A.B.; Dar, B.N. Apple Diseases: Detection and Classification Using Transfer Learning. *Qual. Assur. Saf. Crops Foods* **2023**, *15*, 27–37. [CrossRef]

15. Brahimi, M.; Boukhalfa, K.; Moussaoui, A. Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. *Appl. Artif. Intell.* **2017**, *31*, 299–315. [CrossRef]

16. Khan, H.; Haq, I.U.; Munsif, M.; Mustaqeem; Khan, S.U.; Lee, M.Y. Automated Wheat Diseases Classification Framework Using Advanced Machine Learning Technique. *Agriculture* **2022**, *12*, 1226. [CrossRef]

17. Yasar, A.; Golcuk, A.; Sari, O.F. Classification of Bread Wheat Varieties with a Combination of Deep Learning Approach. *Eur. Food Res. Technol.* **2023**, *250*, 181–189. [CrossRef]

18. Golcuk, A.; Yasar, A. Classification of Bread Wheat Genotypes by Machine Learning Algorithms. *J. Food Compos. Anal.* **2023**, *119*, 105253. [CrossRef]

19. Abade, A.; Ferreira, P.A.; de Barros Vidal, F. Plant Diseases Recognition on Images Using Convolutional Neural Networks: A Systematic Review. *Comput. Electron. Agric.* **2021**, *185*, 106125. [CrossRef]

20. Gea, F.J.; Navarro, M.J.; Santos, M.; Diánez, F.; Carrasco, J. Control of Fungal Diseases in Mushroom Crops While Dealing with Fungicide Resistance: A Review. *Microorganisms* **2021**, *9*, 585. [CrossRef]

21. Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D.; Stefanovic, D. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Comput. Intell. Neurosci.* **2016**, *2016*, 3289801. [CrossRef]

22. Yin, H.; Yi, W.; Hu, D. Computer Vision and Machine Learning Applied in the Mushroom Industry: A Critical Review. *Comput. Electron. Agric.* **2022**, *198*, 107015. [CrossRef]

23. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2016; pp. 2261–2269.

24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

25. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. In *Computer Vision and Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–6.

26. Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 6105–6114.

27. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.

28. Bezanson, J.; Edelman, A.; Karpinski, S.; Shah, V.B. Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.* **2017**, *59*, 65–98. [CrossRef]

29. Pane, C.; Manganiello, G.; Nicastro, N.; Cardi, T.; Carotenuto, F. Powdery Mildew Caused by Erysiphe Cruciferarum on Wild Rocket (*Diplotaxis tenuifolia*): Hyperspectral Imaging and Machine Learning Modeling for Non-Destructive Disease Detection. *Agriculture* **2021**, *11*, 337. [CrossRef]

30. Xie, Y.; Plett, D.; Liu, H. The Promise of Hyperspectral Imaging for the Early Detection of Crown Rot in Wheat. *AgriEngineering* **2021**, *3*, 924–941. [CrossRef]

31. Shah, S.R.; Qadri, S.; Bibi, H.; Shah, S.M.W.; Sharif, M.I.; Marinello, F. Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease. *Agronomy* **2023**, *13*, 1633. [CrossRef]

32. Kini, A.S.; Prema, K.V.; Pai, S.N. Early Stage Black Pepper Leaf Disease Prediction Based on Transfer Learning Using ConvNets. *Sci. Rep.* **2024**, *14*, 1404.

33. Chen, S.; Zhang, K.; Zhao, Y.; Sun, Y.; Ban, W.; Chen, Y.; Zhuang, H.; Zhang, X.; Liu, J.; Yang, T. An Approach for Rice Bacterial Leaf Streak Disease Segmentation and Disease Severity Estimation. *Agriculture* **2021**, *11*, 420. [CrossRef]

34. Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Ayari, M.A.; Khan, A.U.; Khan, M.S.; Al-Emadi, N.; Reaz, M.B.I.; Islam, M.T.; Ali, S.H.M. Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques. *AgriEngineering* **2021**, *3*, 294–312. [CrossRef]

35. Bansal, P.; Kumar, R.; Kumar, S. Disease Detection in Apple Leaves Using Deep Convolutional Neural Network. *Agriculture* **2021**, *11*, 617. [CrossRef]

36. Tan, L.; Lu, J.; Jiang, H. Tomato Leaf Diseases Classification Based on Leaf Images: A Comparison between Classical Machine Learning and Deep Learning Methods. *AgriEngineering* **2021**, *3*, 542–558. [CrossRef]

37. Zhao, S.; Peng, Y.; Liu, J.; Wu, S. Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module. *Agriculture* **2021**, *11*, 651. [CrossRef]

38. Peng, Y.; Zhao, S.; Liu, J. Fused-Deep-Features Based Grape Leaf Disease Diagnosis. *Agronomy* **2021**, *11*, 2234. [CrossRef]

39. Li, J.; Wu, J.; Lin, J.; Li, C.; Lu, H.; Lin, C. Nondestructive Identification of Litchi Downy Blight at Different Stages Based on Spectroscopy Analysis. *Agriculture* **2022**, *12*, 402. [CrossRef]

40. Wan, L.; Li, H.; Li, C.; Wang, A.; Yang, Y.; Wang, P. Hyperspectral Sensing of Plant Diseases: Principle and Methods. *Agronomy* **2022**, *12*, 1451. [CrossRef]

41. Lin, J.; Chen, X.; Pan, R.; Cao, T.; Cai, J.; Chen, Y.; Peng, X.; Cernava, T.; Zhang, X. GrapeNet: A Lightweight Convolutional Neural Network Model for Identification of Grape Leaf Diseases. *Agriculture* **2022**, *12*, 887. [CrossRef]

42. Wang, H.; Shang, S.; Wang, D.; He, X.; Feng, K.; Zhu, H. Plant Disease Detection and Classification Method Based on the Optimized Lightweight YOLOv5 Model. *Agriculture* **2022**, *12*, 931. [CrossRef]

43. Shafik, W.; Tufail, A.; De Silva Liyanage, C.; Apong, R.A.A.H.M. Using Transfer Learning-Based Plant Disease Classification and Detection for Sustainable Agriculture. *BMC Plant Biol.* **2024**, *24*, 136. [CrossRef]

44. Wu, J.; Wen, C.; Chen, H.; Ma, Z.; Zhang, T.; Su, H.; Yang, C. DS-DETR: A Model for Tomato Leaf Disease Segmentation and Damage Evaluation. *Agronomy* **2022**, *12*, 2023. [CrossRef]

45. Yin, C.; Zeng, T.; Zhang, H.; Fu, W.; Wang, L.; Yao, S. Maize Small Leaf Spot Classification Based on Improved Deep Convolutional Neural Networks with a Multi-Scale Attention Mechanism. *Agronomy* **2022**, *12*, 906. [CrossRef]

46. Zendler, D.; Malagol, N.; Schwandner, A.; Töpfer, R.; Hausmann, L.; Zyprian, E. High-Throughput Phenotyping of Leaf Discs Infected with Grapevine Downy Mildew Using Shallow Convolutional Neural Networks. *Agronomy* **2021**, *11*, 1768. [CrossRef]

47. Zahan, N.; Zahid Hasan, M.; Sharif Uddin, M.; Hussain, S.; Fahmida Islam, S. A Deep Learning-Based Approach for Mushroom Diseases Classification. In *Application of Machine Learning in Agriculture*; Academic Press: Cambridge, MA, USA, 2022; pp. 191–212.

48. Gu, Y.H.; Yin, H.; Jin, D.; Zheng, R.; Yoo, S.J. Improved Multi-Plant Disease Recognition Method Using Deep Convolutional Neural Networks in Six Diseases of Apples and Pears. *Agriculture* **2023**, *13*, 300. [CrossRef]

49. Orchi, H.; Sadik, M.; Khaldoun, M.; Sabir, E. Automation of Crop Disease Detection through Conventional Machine Learning and Deep Transfer Learning Approaches. *Agriculture* **2023**, *13*, 352. [CrossRef]

50. Mehmood, A.; Ahmad, M.; Ilyas, Q.M. On Precision Agriculture: Enhanced Automated Fruit Disease Identification and Classification Using a New Ensemble Classification Method. *Agriculture* **2023**, *13*, 500. [CrossRef]

51. Sood, S.; Singh, H. A Comparative Study of Grape Crop Disease Classification Using Various Transfer Learning Techniques. *Multimed. Tools Appl.* **2024**, *83*, 4359–4382. [CrossRef]

52. Ngugi, L.C.; Abdelwahab, M.; Abo-Zahhad, M. Tomato Leaf Segmentation Algorithms for Mobile Phone Applications Using Deep Learning. *Comput. Electron. Agric.* **2020**, *178*, 105788. [CrossRef]

53. Zhang, D.; Zhang, W.; Cheng, T.; Lei, Y.; Qiao, H.; Guo, W.; Yang, X.; Gu, C. Segmentation of Wheat Scab Fungus Spores Based on CRF_ResUNet++. *Comput. Electron. Agric.* **2024**, *216*, 108547. [CrossRef]

54. Zhao, Y.; Liu, S.; Hu, Z.; Bai, Y.; Shen, C.; Shi, X. Separate Degree Based Otsu and Signed Similarity Driven Level Set for Segmenting and Counting Anthrax Spores. *Comput. Electron. Agric.* **2020**, *169*, 105230. [CrossRef]

55. Jiang, H.; Zhang, C.; Qiao, Y.; Zhang, Z.; Zhang, W.; Song, C. CNN Feature Based Graph Convolutional Network for Weed and Crop Recognition in Smart Farming. *Comput. Electron. Agric.* **2020**, *174*, 105450. [CrossRef]

56. Hornberg, A. *Handbook of Machine and Computer Vision: The Guide for Developers and Users*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2017.

57. Cubero, S.; Aleixos, N.; Moltó, E.; Gómez-Sanchis, J.; Blasco, J. Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables. *Food Bioprocess Technol.* **2011**, *4*, 487–504. [CrossRef]

58. Gómez-Sanchis, J.; Moltó, E.; Camps-Valls, G.; Gómez-Chova, L.; Aleixos, N.; Blasco, J. Automatic Correction of the Effects of the Light Source on Spherical Objects. An Application to the Analysis of Hyperspectral Images of Citrus Fruits. *J. Food Eng.* **2008**, *85*, 191–200. [CrossRef]

59. Carrasco, J.; Navarro, M.J.; Gea, F.J. Cobweb, a Serious Pathology in Mushroom Crops: A Review. *Span. J. Agric. Res.* **2017**, *15*, e10R01. [CrossRef]

60. Li, L.; Zhang, Q.; Huang, D. A Review of Imaging Techniques for Plant Phenotyping. *Sensors* **2014**, *14*, 20078. [CrossRef] [PubMed]

61. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

62. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

63. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

64. Smith, L.N. Cyclical Learning Rates for Training Neural Networks. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017, Santa Rosa, CA, USA, 27–29 March 2017; pp. 464–472.

65. Sokolova, M.; Lapalme, G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]

66. Davis, J.; Goadrich, M. The Relationship between Precision-Recall and ROC Curves. *ACM Int. Conf. Proc. Ser.* **2006**, *148*, 233–240.

67. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. *Lect. Notes Comput. Sci.* **2005**, *3408*, 345–359.

68. De Diego, I.M.; Redondo, A.R.; Fernández, R.R.; Navarro, J.; Moguerza, J.M. General Performance Score for Classification Problems. *Appl. Intell.* **2022**, *52*, 12049–12063. [CrossRef]

69. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]

70. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2009**, *88*, 303–338. [CrossRef]

71. Liu, C.; Zhao, C.; Wu, H.; Han, X.; Li, S. ADDLight: An Energy-Saving Adder Neural Network for Cucumber Disease Classification. *Agriculture* **2022**, *12*, 452. [CrossRef]

72. Simumba, N.; Okami, S.; Kodaka, A.; Kohtake, N. Alternative Scoring Factors Using Non-Financial Data for Credit Decisions in Agricultural Microfinance. In Proceedings of the 2018 IEEE International Systems Engineering Symposium (ISSE), Rome, Italy, 1–3 October 2018; pp. 1–8.

73. Mohanty, S.P.; Hughes, D.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]

74. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the ICLR 2021—9th International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021.

75. Huang, M.; Xu, G.; Li, J.; Huang, J. A Method for Segmenting Disease Lesions of Maize Leaves in Real Time Using Attention YOLACT++. *Agriculture* **2021**, *11*, 1216. [CrossRef]

76. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]

77. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [CrossRef]

78. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.

79. Dubovan, J.; Litvik, J.; Benedikovic, D.; Mullerova, J.; Glesk, I.; Veselovsky, A.; Dado, M. Impact of Wind Gust on High-Speed Characteristics of Polarization Mode Dispersion in Optical Power Ground Wire Cables. *Sensors* **2020**, *20*, 7110. [CrossRef]