Department of Computer Science and Engineering
Bangladesh University of Business and Technology (BUBT)

CSE 498: Literature Review Records

Student's Id and Name	(21224103162) Md Masud Rana
Capstone Project Title	**Leaf Diseases Detection**
Supervisor Name & Designation	M.M Fazle Rabbi, Assistant Professor
Course Teacher's Name & Designation	M.M Fazle Rabbi, Assistant Professor

Aspects	**Paper # 1 (Leaf Diseases Detection)**
Title / Question (What is problem statement?)	The problem arises in early detection of plant diseases is critical to secure agricultural yield and food supply. However, due to environmental challenges and limited technological awareness among farmers, diseases in plants often go undetected until they cause significant damage. This research aims to leverage machine learning and computer vision techniques to develop a highly accurate model that can identify diseases in tomato leaves early, enabling timely interventions for better crop management.
Objectives / Goal (What is looking for?)	This project aims to develop a machine learning model using computer vision for early detection of tomato leaf diseases, helping reduce crop losses. By applying advanced image processing techniques (DWT, PCA, GLCM), it enhances feature extraction for higher accuracy. The project evaluates classifiers like SVM, K-NN, and CNN to identify the best method for disease detection, promoting accessible agricultural technology for farmers
Methodology/Theory (How to find the solution?)	**Methodologies**

	• **Image Preprocessing:** Resize images, apply histogram equalization, and use K-means clustering to enhance quality and segment disease regions on leaf samples. Set objectives: real-time data, remote monitoring, decision support. • **Feature Extraction:** Utilize Discrete Wavelet Transform (DWT) for frequency analysis, Principal Component Analysis (PCA) for dimensionality reduction, and Gray Level Co-occurrence Matrix (GLCM) for texture analysis to capture disease-related features. • **Classification:** Implement multiple machine learning classifiers—Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Convolutional Neural Network (CNN)—to classify diseases in leaf samples. • **Evaluation Metrics:** Assess model performance using accuracy, precision, recall, and F1 score to determine the best classification approach for tomato leaf disease detection
Software Tools (What program/software is used for design, coding and simulation?)	**The software tools used for design, coding, and simulation in this project include:** • **Python** – The primary programming language for implementing machine learning algorithms and processing images. • **TensorFlow and Keras** – Deep learning libraries for building and training neural networks, particularly the CNN model. • **OpenCV** – A computer vision library for image preprocessing and feature extraction tasks. • **NumPy** – A library for numerical computations, used to handle data manipulation and array operations • **Matplotlib** – A plotting library for visualizing results and performance metrics. • **PlantVillage Dataset** – A dataset containing images of tomato leaves with various diseases, used for training and testing the model

Test / Experiment How to test and characterize the design/prototype?	**To test and characterize the design/prototype, the following steps will be performed:** • **Dataset Preparation:** Use a publicly available dataset of tomato leaf images with various diseases (e.g., the PlantVillage dataset) for training and testing. The images will be divided into training and testing sets to evaluate model performance. • **Preprocessing Validation:** Ensure the preprocessing steps (resizing, histogram equalization, and K-means clustering) effectively enhance image quality and segment disease regions. Visual inspection and quantitative metrics like segmentation accuracy can be used to validate preprocessing. • **Feature Extraction Testing:** Assess the effectiveness of feature extraction techniques (DWT, PCA, GLCM) by evaluating the features' ability to distinguish between healthy and diseased leaves. This can be done through statistical analysis and visual comparisons. • **Model Training and Testing:** Train the classifiers (SVM, K-NN, CNN) on the preprocessed and feature-extracted data. The models will be tested on the validation dataset, and their performance will be evaluated using accuracy, precision, recall, and F1 score. • **Performance Comparison:** Compare the results of the different classifiers to determine which one provides the highest accuracy and efficiency. Use confusion matrices and ROC curves to visualize classification performance. • **Real-World Testing:** If possible, test the prototype on real-world tomato leaf samples from farms to validate its practical applicability. • **Evaluation Metrics:** Final model characterization will be based on performance metrics like precision, recall, F1 score, and overall accuracy, ensuring the model can effectively detect diseases with high reliability
Result / Conclusion (What was the final result?)	The Smart Agriculture Android app improved farming efficiency, decision-making, and productivity by providing real-time data and actionable insights. Positive user feedback confirmed the app’s usability and relevance, leading to enhanced crop yield and resource optimization. Overall, the app successfully met its objectives, empowering farmers with accessible, data-driven tools for sustainable agriculture.

Obstacles/Challenges (List the methodological obstacles if authors mentioned in the article)	The final result showed that the Convolutional Neural Network (CNN) classifier achieved the highest accuracy (99.09%) for detecting and classifying tomato leaf diseases, outperforming other classifiers like SVM and K-NN. The combination of preprocessing techniques and advanced feature extraction methods (DWT, PCA, GLCM) significantly improved detection accuracy. This approach demonstrates a reliable solution for early disease identification, helping farmers manage plant health effectively.
Terminology (List the common basic words frequently used in this research field)	**Common terminology in plant disease detection and machine learning includes:** Preprocessing, Feature Extraction, Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), Discrete Wavelet Transform (DWT), Principal Component Analysis (PCA), Gray Level Co-occurrence Matrix (GLCM), Image Segmentation, Accuracy, Precision, Recall, F1 Score, PlantVillage Dataset.
Review Judgment (Briefly compare the objectives and results of all the articles you reviewed)	Smart agriculture apps enhance crop management, resource optimization, and yield, promoting sustainable practices and data-driven decision-making, enhancing crop management and user accessibility.
Review Outcome (Make a decision how to use/refer the obtained knowledge to prepare a separate and new methodology for your own research project)	We will develop a unique methodology for my smart agriculture mobile app that focuses on detecting crop leaf diseases using machine learning including decisions • **Dataset Selection** • **Image pre-processing** • **Model Selection** • **Performance Evaluation** • **Mobile Application Integration**

Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

CSE 498: Literature Review Records

Student's Id and Name	(21224103162) Md Masud Rana
Capstone Project Title	**Leaf Diseases Detection**
Supervisor Name & Designation	M.M Fazle Rabbi, Assistant Professor
Course Teacher's Name & Designation	M.M Fazle Rabbi, Assistant Professor

Aspects	Paper # 2 (Leaf Diseases Detection)
Title / Question (What is problem statement?)	The problem statement addressed in this research paper on plant disease detection is: Crop diseases pose a significant threat to food security, but their rapid identification remains challenging in many parts of the world due to the lack of necessary infrastructure. The paper aims to address this issue by developing an accurate and efficient method for detecting plant diseases using machine learning techniques, specifically focusing on leaf-based image classification.
Objectives / Goal (What is looking for?)	The main objectives/goals of the research presented in this paper are: • Develop an accurate method for rapid plant disease detection using machine learning. • Implement a Random Forest classifier to distinguish between healthy and diseased leaves.

	• Create a dataset of diseased and healthy leaf images for training. • Use Histogram of Oriented Gradient (HOG) for effective feature extraction from leaf images. • Provide a scalable solution for plant disease identification, especially in areas lacking agricultural infrastructure. • Contribute to improved food security by enabling earlier and more accurate detection of crop diseases.

Methodology/Theory
 	
(How to find the solution?)
	The methodology of this study focuses on detecting plant diseases by analyzing leaf images using machine learning techniques. Here's a summary of the approach:
	• **Dataset Preparation**: Healthy and diseased leaf images are gathered and labeled, creating a dataset for training and testing the machine learning model.
	• **Preprocessing**: Images are resized uniformly and converted to grayscale, enabling consistent feature extraction across the dataset.
	• **Feature Extraction**: Three main feature descriptors are applied:
	    • **Hu Moments**: Used to capture the shape of the leaf by calculating shape-based features.
	    • **Haralick Texture**: Measures texture differences, useful for distinguishing between healthy and diseased textures.
	    • **Color Histogram**: Representing color distribution, this histogram is computed in the HSV color space to match human color perception
	• **Classifier Training and Testing:** The performance of the Random Forest classifier is evaluated, showing about 70% accuracy

Software Tools (What program/software is used for design, coding and simulation?)	**The software tools used for design, coding, and simulation in this project include:** • **Python:** Widely used for machine learning tasks, Python is a common choice for implementing models due to its extensive libraries and frameworks. • **OpenCV:** This computer vision library in Python (or C++) is useful for image processing tasks such as resizing, color conversion, and feature extraction OpenCV – A computer vision library for image preprocessing and feature extraction tasks. • **Scikit-Learn:** A Python library specifically for machine learning, Scikit-Learn provides tools to implement and evaluate algorithms like Random Forests, SVM, K-Nearest Neighbors, and Naive Bayes. • **NumPy and Pandas:** Essential for handling and manipulating data arrays and data frames, these libraries are commonly used for feature processing and dataset management. • **Matplotlib and Seaborn:** These visualization libraries in Python help plot data and visualize results, such as displaying color histograms, comparing classifier performance, and plotting accuracy metrics. • **Jupyter Notebook or Spyder:** Both are integrated development environments (IDEs) for Python. Jupyter is particularly popular for iterative development, testing, and visualization, while Spyder is more suited for modular code development. • **TensorFlow or PyTorch (if deep learning is used):** If the study incorporates deep learning for disease classification, TensorFlow or PyTorch might be used to build neural network architectures, although simpler machine learning methods like Random Forests might eliminate the need for these.

Test / Experiment How to test and characterize the design/prototype?	**To test and characterize the plant disease detection model:** • **Split Dataset**: Divide the dataset into training and testing sets to evaluate the model's performance on unseen data. • **Train the Model**: Use the training set to teach the model to recognize healthy and diseased leaves • **Feature Extraction**: For each test image, extract features (Hu Moments, Haralick Texture, Color Histogram) and input them into the trained classifier. • **Prediction and Evaluation**: Predict disease status for test images and compare predictions with actual labels to calculate accuracy, precision, recall, and F1-score. • **Cross-Validation**: Perform k-fold cross-validation to assess the model’s robustness and generalization on different subsets of the dataset. • **Real-World Testing:** If possible, test the prototype on real-world tomato leaf samples from farms to validate its practical applicability.
Result / Conclusion (What was the final result?)	The final result showed that the Random Forest classifier achieved around 70% accuracy in distinguishing between healthy and diseased leaves. This accuracy suggests that the model can moderately identify plant diseases based on leaf image features (shape, texture, color). With a larger dataset and potentially additional features, the accuracy could improve further, enhancing its practical application for large-scale disease detection in crops.
Obstacles/Challenges (List the methodological obstacles if authors mentioned in the article)	**The main challenges mentioned include:** • **Limited Dataset Size:** The model's accuracy (around 70%) was impacted by the small dataset, which limits its generalizability and performance.

	• **Feature Extraction Complexity:** Properly extracting and combining shape, texture, and color features (Hu Moments, Haralick Texture, and Color Histogram) is complex and requires precise preprocessing. • **Model Selection and Performance:** While Random Forest showed reasonable accuracy, finding a model that balances accuracy with processing speed for large datasets remains challenging. • **Background Noise: Capturing leaf images in controlled settings was necessary to reduce background interference, but this is difficult to maintain in real-world conditions.**
Terminology (List the common basic words frequently used in this research field)	**Common basic words frequently used in plant disease detection and machine learning research include:** Dataset, Classifier, Feature Extraction, Accuracy, Image Processing, Neural Network, Segmentation, Classification, Histogram, Image Segmentation, Accuracy, Precision, Recall, F1 Score.
Review Judgment (Briefly compare the objectives and results of all the articles you reviewed)	The reviewed articles all focus on using machine learning to detect plant diseases from leaf images. The common objective is to develop an accurate, efficient method to differentiate between healthy and diseased leaves to support agricultural sustainability and early disease intervention • **Objective Comparison**: Each article aims to improve disease detection accuracy through different methodologies. Some emphasize specific classifiers like Random Forest, SVM, or Neural Networks, while others explore unique feature extraction methods, such as Hu Moments or Color Histograms, to enhance classification. • **Results Comparison**: Accuracy varied across studies: ➢ Studies using Random Forest often achieved moderate accuracy (~70%), suitable for general disease classification. ➢ Neural Network-based approaches demonstrated high potential accuracy, though they required more computational resources and larger datasets.

	➢ SVM and K-Nearest Neighbors yielded moderate to high accuracy, though performance fluctuated based on dataset and feature choices. In summary, while objectives across the studies align in improving disease detection, results depend heavily on the dataset, choice of classifier, and feature extraction method. Most studies suggest that combining advanced feature extraction techniques and increasing dataset size can improve model accuracy and real-world applicability.
Review Outcome (Make a decision how to use/refer the obtained knowledge to prepare a separate and new methodology for your own research project)	Based on insights from reviewed studies, I will structure my methodology with the following adaptations: • **Hybrid Classifier Model:** Combine multiple machine learning algorithms (Random Forest, SVM, and Neural Networks) in an ensemble approach to improve classification accuracy and robustness for plant disease detection. Image pre-processing • **Diverse Feature Extraction:** Use a combination of advanced feature extraction techniques—Hu Moments for shape, Haralick Texture for surface patterns, and Color Histogram in HSV space—to fully capture leaf characteristics and enhance detection accuracy. Performance Evaluation • **Preprocessing Optimization:** Apply detailed image preprocessing, including noise reduction, resizing, and background removal, to standardize inputs and improve feature consistency. • **Evaluation Strategy:** Implement k-fold cross-validation and precision-recall analysis to accurately assess model performance, ensuring it generalizes well to unseen data.

Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

CSE 498: Literature Review Records

Student's Id and Name	(21224103162) Md Masud Rana
Capstone Project Title	**Leaf Diseases Detection**
Supervisor Name & Designation	M.M Fazle Rabbi, Assistant Professor
Course Teacher's Name & Designation	M.M Fazle Rabbi, Assistant Professor

Aspects	**Paper # 3 (Leaf Diseases Detection)**
Title / Question (What is problem statement?)	The problem statement in the article centers on developing an automated system for detecting plant leaf diseases. Given the limitations of manual disease diagnosis, this system leverages machine learning and image processing to provide fast, accurate identification of diseases from leaf images, enabling timely intervention to improve crop health and yield. The framework involves preprocessing leaf images, segmenting them to highlight disease areas, extracting critical features, and classifying the disease using algorithms like SVM, Random Forest, and ID3.
Objectives / Goal (What is looking for?)	The paper outlines objectives and goals associated with detecting plant leaf diseases using machine learning and image processing techniques. The primary goals include: • **Automating Plant Disease Detection**: Developing a system that can automatically diagnose plant diseases by analyzing leaf images, aiming to support farmers with rapid and precise diagnoses to improve crop yield and management.

	• **Improving Diagnostic Accuracy**: Utilizing machine learning algorithms (such as SVM, RBF-SVM, random forest, and ID3) and image processing methods (like image segmentation and feature extraction) to enhance the accuracy, sensitivity, and specificity of disease detection. • **Enhancing Image Processing Techniques**: The study focuses on image preprocessing, noise removal, and segmentation, improving the quality of data fed into classification algorithms to achieve better disease identification results. These objectives collectively aim to support precision agriculture by providing a reliable, fast, and effective tool for detecting and managing plant diseases based on leaf images
Methodology/Theory (How to find the solution?)	**The methodology for plant leaf disease detection involves several key steps:** • **Image Acquisition and Preprocessing:** Collect images of diseased leaves and preprocess them by removing noise with a mean filter and enhancing image quality through histogram equalization. • **Image Segmentation**: Use the K-means clustering algorithm to divide the image into segments, helping to isolate and identify diseased regions. • **Feature Extraction:** Apply Principal Component Analysis (PCA) to extract relevant features from the segmented images, reducing data complexity and focusing on critical disease indicators: • **Classification:** Classify the images using machine learning models like RBF-SVM, SVM, random forest, and ID3 to identify disease type based on the extracted features.
Software Tools (What program/software is used for design, coding and simulation?)	**The study uses machine learning and image processing software tools for design, coding, and simulation, including:** • **MATLAB or Python:** Commonly used for image processing, machine learning algorithms, and data analysis

	• **OpenCV:** Utilized for image processing tasks like noise removal, segmentation, and feature extraction. • **Scikit-Learn**: A Python library that provides tools for implementing machine learning models such as SVM, random forest, and ID3
Test / Experiment How to test and characterize the design/prototype?	**To test and characterize the plant disease detection model:** • **Dataset Division:** Split the dataset into training and testing sets. For example, 80% of images are used for training machine learning models, while the remaining 20% are used for testing. • **Performance Metrics**: Evaluate the model's accuracy, sensitivity, and specificity by comparing predictions to actual disease classifications. Key metrics like True Positive, True Negative, False Positive, and False Negative rates are calculated. • **Feature Extraction**: For each test image, extract features (Hu Moments, Haralick Texture, Color Histogram) and input them into the trained classifier. • **Algorithm Comparison**: Test multiple machine learning algorithms (e.g., SVM, RBF-SVM, random forest, ID3) to identify which yields the best results for leaf disease classification. • **Real-World Testing:** If possible, test the prototype on real-world tomato leaf samples from farms to validate its practical applicability.
Result / Conclusion (What was the final result?)	The final result showed that the RBF-SVM algorithm provided the highest accuracy for detecting plant leaf diseases compared to other tested models (SVM, random forest, ID3). The automated system successfully identified different types of leaf diseases, achieving high accuracy, sensitivity, and specificity. This system offers a reliable and efficient tool for farmers to diagnose plant diseases quickly, aiding in better crop management and potentially increasing yield.

Obstacles/Challenges (List the methodological obstacles if authors mentioned in the article)	**The article notes several challenges in the methodology, including:** • **Image Quality and Noise**: Variations in image quality and noise can impact the accuracy of disease detection, requiring robust preprocessing. • **Segmentation Complexity**: Accurate segmentation of diseased areas in leaves is challenging, as it relies on the effectiveness of clustering algorithms like K-means. • **Feature Extraction:** Selecting relevant features for classification is crucial but challenging due to the complexity and variability in leaf textures and disease symptoms. • **Model Performance:** Achieving high accuracy with machine learning models depends on selecting suitable algorithms and fine-tuning them to avoid overfitting or underfitting.
Terminology (List the common basic words frequently used in this research field)	**Common basic words frequently used in plant disease detection and machine learning research include:** Image Processing, Segmentation, Classification, Machine Learning Algorithms, Specificity, Sensitivity, Accuracy, PCA, K-Means Clustering, RBF
Review Judgment (Briefly compare the objectives and results of all the articles you reviewed)	The objectives across reviewed articles commonly focus on developing automated, accurate, and efficient systems for plant leaf disease detection using image processing and machine learning. Each study aims to enhance diagnostic speed and accuracy to support farmers in managing crop health.

	Results show varying levels of success, with some models achieving higher accuracy through advanced techniques like RBF-SVM or ensemble methods (e.g., random forest). While traditional methods like K-means clustering and SVM are effective, studies integrating more sophisticated algorithms (e.g., neural networks) often report better performance metrics. Overall, advancements in image processing and machine learning have led to significant improvements in disease detection accuracy and reliability across different crops and disease types.
Review Outcome (Make a decision how to use/refer the obtained knowledge to prepare a separate and new methodology for your own research project)	To develop a new methodology for a research project on plant disease detection, the insights from these articles can be applied as follows: • **Image Preprocessing and Noise Reduction:** Implement noise reduction methods like the mean filter and histogram equalization to enhance image quality, which are essential for clear disease feature extraction. • **Advanced Segmentation Techniques**: Use or improve upon K-means clustering for segmenting diseased areas. Additional refinement techniques, such as adaptive thresholding or watershed algorithms, can help address segmentation accuracy. • **Enhanced Feature Extraction**: Utilize PCA for dimensionality reduction, focusing on extracting meaningful features like color, texture, and shape specific to each disease. • **Algorithm Selection**: Test a variety of machine learning models (e.g., RBF-SVM, random forest, neural networks) and tune them to improve accuracy, sensitivity, and specificity in disease classification. • **Performance Evaluation**: Define performance metrics clearly (accuracy, sensitivity, specificity) and optimize the model for high scores across these metrics to ensure reliability in practical applications

Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

CSE 498: Literature Review Records

Student's Id and Name	(21224103162) Md Masud Rana
Capstone Project Title	**Leaf Diseases Detection**
Supervisor Name & Designation	M.M Fazle Rabbi, Assistant Professor
Course Teacher's Name & Designation	M.M Fazle Rabbi, Assistant Professor

Aspects	**Paper # 4 (Leaf Diseases Detection)**
Title / Question (What is problem statement?)	The problem statement of the study is focused on the agricultural challenge of early and accurate detection of plant diseases. These diseases cause substantial losses in food production and economic value, affecting agricultural productivity and quality. Manual identification of plant diseases is subjective, time-consuming, and prone to error. The research aims to address this by employing deep learning models—specifically convolutional neural networks (CNNs)—to automate and improve the accuracy of leaf disease detection in crops using image analysis. By optimizing and testing various CNN models, the study seeks to create a reliable, efficient tool for early disease diagnosis, ultimately supporting agricultural sustainability and food security.
Objectives / Goal (What is looking for?)	The objective of the study is to develop and evaluate deep learning models, specifically using CNNs, to achieve high accuracy in detecting and classifying multiple plant diseases from leaf images. The goal is to determine the most effective model for reliable, automated disease detection, thereby enabling early intervention, improving crop health, and supporting agricultural productivity

Methodology/Theory (How to find the solution?)	The study uses a transfer learning approach with pre-trained CNN models (DenseNet-121, ResNet-50, VGG-16, and Inception V4) to classify plant diseases from leaf images. The models are fine-tuned and trained on the PlantVillage dataset, which includes various plant species and disease types. Data augmentation techniques are applied to prevent overfitting, and hyperparameters are optimized to improve model accuracy. The performance of each model is evaluated based on accuracy, sensitivity, specificity, and F1 score to identify the most effective model for plant disease detection.
Software Tools (What program/software is used for design, coding and simulation?)	The study uses several software tools for model design, coding, and simulation, including Anaconda3 for the Python environment, Keras for deep learning model implementation, OpenCV for image processing, NumPy for numerical operations, and CuDNN and Theano libraries optimized for GPU acceleration. These tools support efficient training and testing of deep learning models on large datasets like PlantVillage**.**
Test / Experiment How to test and characterize the design/prototype?	The design is tested by training each deep learning model (DenseNet-121, ResNet-50, VGG-16, and Inception V4) on the PlantVillage dataset, which is split into training, validation, and test sets. Each model's performance is characterized through evaluation metrics, including **accuracy**, **sensitivity**, **specificity**, and **F1 score**. The models are run for a set number of epochs, and their convergence is observed. Data augmentation is used to assess model robustness, and testing accuracy and loss are recorded to determine the best model for plant disease classification.

Result / Conclusion (What was the final result?)	The final result showed that the RBF-SVM algorithm provided the highest accuracy for detecting plant leaf diseases compared to other tested models (SVM, random forest, ID3). The automated system successfully identified different types of leaf diseases, achieving high accuracy, sensitivity, and specificity. This system offers a reliable and efficient tool for farmers to diagnose plant diseases quickly, aiding in better crop management and potentially increasing yield.
Obstacles/Challenges (List the methodological obstacles if authors mentioned in the article)	**The study mentions several methodological obstacles::** • Dataset Diversity: Limited diversity in the dataset images affects model generalizability, making it difficult to apply the model to real-world agricultural settings with varied backgrounds and lighting conditions. • **Overfitting:** The large size of the dataset made the model prone to overfitting, which required data augmentation techniques to ensure the model's robustness. • **Computational Complexity:** Training deep CNN models is computationally intensive, requiring substantial processing power, memory, and time • **Vanishing Gradient:** Deeper models like ResNet and DenseNet faced challenges with vanishing gradients, which were mitigated using batch normalization and skip connections in the network.
Terminology (List the common basic words frequently used in this research field)	**Here are some commonly used terms in the field of plant disease detection using deep learning:** CNN, [redacted] Transfer Learning**,** [redacted] Data Augmentation, Overfitting, Accuracy, Sensitivity**,** Specificity, Score, [redacted] PlantVillage Dataset, Hyperparameters, Fine-Tuning, Classification

Review Judgment (Briefly compare the objectives and results of all the articles you reviewed)	Based on a review of multiple articles on plant disease detection using machine learning, common objectives and outcomes can be summarized as follows. To develop a new methodology for a research project on plant disease detection, the insights from these articles can be applied as follows: **Objectives** • **High-Accuracy Disease Classification:** Most studies aim to achieve high accuracy in classifying plant diseases through image analysis, using machine learning or deep learning models. • **Efficient Feature Extraction:** Studies often seek to improve the feature extraction process, with a focus on using convolutional neural networks (CNNs) or other deep learning methods to automate and enhance the identification of specific visual patterns in leaf diseases. • **Comparison of Model Performance:** Many studies compare various machine learning and deep learning algorithms (e.g., CNNs, SVMs, Random Forests) to find the best model for plant disease classification. • **Real-World Applicability:** Some studies emphasize the need for real-time detection capabilities, which could allow farmers to diagnose diseases in the field using mobile applications. **Results:** • **High Accuracy with CNNs:** Across studies, CNN-based models, especially with transfer learning, consistently achieve high accuracy rates (often above 90%) in plant disease classification. • **Transfer Learning's Impact:** Models that utilize transfer learning with pre-trained CNN architectures (such as DenseNet, ResNet, and Inception) generally outperform models built from scratch in terms of both accuracy and training efficiency.

	• **Dataset Limitations and Challenges:** Many studies identify a lack of diversity in available datasets, such as the commonly used PlantVillage dataset, which affects model robustness in real-world conditions
Review Outcome (Make a decision how to use/refer the obtained knowledge to prepare a separate and new methodology for your own research project)	Based on the knowledge obtained from the review, here's a suggested approach to developing a new methodology: • **Utilize Transfer Learning with CNN Models:** Given the success of pre-trained CNNs (like DenseNet, ResNet, and Inception) in existing research, these models could be adopted for efficient feature extraction and high classification accuracy. Fine-tuning specific layers to adapt to the target dataset would allow leveraging these models' robust architecture while tailoring them to the unique disease characteristics in your dataset. • **Incorporate Data Augmentation Techniques:** To address dataset diversity limitations, incorporate data augmentation (e.g., rotations, flips, zoom) to create a more robust model capable of handling real-world variations. This would help reduce overfitting and improve the model's generalizability. • **Evaluate with Comprehensive Metrics:** Use accuracy, sensitivity, specificity, and F1 score as primary performance metrics to ensure a balanced evaluation of the model's ability to correctly identify both diseased and healthy samples. • **Benchmark with Real-World Dataset:** Consider using a dataset with varied backgrounds, lighting conditions, and real-world noise to validate the model. This step would ensure your model is more applicable outside controlled settings, providing reliable performance in practical agricultural environments. These steps would combine proven techniques with adjustments tailored to your project's unique requirements, helping to create a novel and effective methodology for plant disease detection.

Department of Computer Science and Engineering

Bangladesh University of Business and Technology (BUBT)

CSE 498: Literature Review Records

Student's Id and Name	(21224103162) Md Masud Rana
Capstone Project Title	**Leaf Diseases Detection**
Supervisor Name & Designation	M.M Fazle Rabbi, Assistant Professor
Course Teacher's Name & Designation	M.M Fazle Rabbi, Assistant Professor

Aspects	Paper # 5 (Leaf Diseases Detection)
Title / Question (What is problem statement?)	The problem statement in the document is to develop an automated, accurate, and efficient system for detecting and classifying diseases in potato leaves. This is important due to the challenges in manual detection, which is time-consuming and prone to human error. The proposed solution involves using a deep learning model that can classify potato leaf diseases into multiple classes with high accuracy, addressing limitations in existing methods that detect only a few disease types and often suffer from data imbalance issues.
Objectives / Goal (What is looking for?)	The goal of this study is to create an effective, automated deep learning framework for identifying and classifying potato leaf diseases into five specific categories: Potato Late Blight (PLB), Potato Early Blight (PEB), Potato Leaf Roll (PLR), Potato Verticillium Wilt (PVw), and Potato Healthy (PH). The proposed model aims to improve on existing methods by achieving high accuracy in disease detection, addressing class imbalance in datasets, and enhancing computational efficiency through a modified DenseNet-201 architecture.

Methodology/Theory (How to find the solution?)	The solution involves a deep learning-based methodology using a modified DenseNet-201 architecture. Here's the approach: • **Data Collection:** Images of potato leaves are gathered from the Plant Village dataset and additional manual captures to cover all disease categories. • **Data Preprocessing**: Images are preprocessed through normalization, CLAHE (Contrast Limited Adaptive Histogram Equalization), and resizing to enhance visibility and reduce computational load. Binary file conversion is applied to minimize image size. • **Model Architecture**: The DenseNet-201 model is modified by adding an extra transition layer to reduce feature map size, thus decreasing computational complexity and improving efficiency. • **Class Imbalance Handling**: A reweighted cross-entropy loss function is employed to counteract class imbalance, allowing the model to be more robust on underrepresented classes. • **raining and Testing**: The model is trained on a balanced dataset, with features automatically extracted through the DenseNet-201 layers. The softmax classifier is used for final classification across the five disease categories. • **Evaluation**: The model's performance is measured using accuracy, precision, recall, and F1 scores on a test set, achieving a reported accuracy of 97.2% in classification.
Software Tools (What program/software is used for design, coding and simulation?)	The software tools used for the design, coding, and simulation of this deep learning framework include: • **Python:** The primary programming language used for model development and data processing**.** • **Keras:** A high-level neural networks API in Python, used to implement and train the DenseNet-201 model

	• **NVIDIA GPU (Quadro GM107GL):** For accelerated training and testing of the model. • **TensorFlow:** Backend framework supporting Keras, enabling deep learning operations and GPU support. • **OpenCV and NumPy:** Libraries for image processing, including operations like resizing, CLAHE, and normalization. • **Matplotlib:** For visualizing data, model performance metrics, and training/testing accuracy curves
Test / Experiment How to test and characterize the design/prototype?	The design is tested by dividing the dataset into training, validation, and testing sets. The model's performance is evaluated on the test set using metrics such as accuracy, precision, recall, and F1 score. A confusion matrix is also used to assess classification accuracy across the five potato leaf disease categories, verifying the model's robustness and efficiency in disease detection.to determine the best model for plant disease classification.

<table>
<tr><td>Result / Conclusion

(What was the final result?)</td><td>The final result showed that the proposed deep learning model achieved an accuracy of 97.2% in detecting and classifying five types of potato leaf diseases. This model demonstrated high robustness and efficiency, outperforming existing methods due to its improved DenseNet-201 architecture and effective handling of class imbalance. The study concludes that the model is suitable for accurate, automated disease detection in potato leaves, which can aid in timely crop management.</td></tr>
<tr><td>Obstacles/Challenges

(List the methodological obstacles if authors mentioned in the article)</td><td>The main obstacles mentioned include:
Class Imbalance: The dataset had uneven distribution across disease categories, with fewer samples for some classes, which could affect model performance. This was addressed using a reweighted cross-entropy loss function.
Limited Datasets: Publicly available datasets covered only a few disease types (early and late blight), so additional data had to be manually collected for other disease types like Potato Leaf Roll and Potato Verticillium Wilt
Overfitting: Due to the small size of some training classes, there was a risk of overfitting. The DenseNet architecture with dense connections and regularization techniques helped mitigate this issue.
Computational Complexity: Deep models are computationally intensive, so an extra transition layer was added in the DenseNet-201 architecture to reduce feature map size and computational demands.</td></tr>
<tr><td>Terminology

(List the common basic words frequently used in this research field)</td><td>Here are some commonly used terms in the field of plant disease detection using deep learning:

1. Deep Learning, DenseNet , Classification , Potato Leaf Disease, Plant Village Dataset, Image Preprocessing , Convolutional Neural Network (CNN) , Cross-Entropy Loss , Class Imbalance, Early Blight, Late Blight , Potato Leaf Roll, Potato Verticillium Wilt, Accuracy, Precision</td></tr>
</table>

Review Judgment (Briefly compare the objectives and results of all the articles you reviewed)	The primary objective across the reviewed articles is to develop accurate, automated systems for detecting and classifying plant diseases, especially in potato leaves, using machine learning and deep learning methods. Most studies use image-based data, with a common reliance on the **Plant Village Dataset**, and aim to classify diseases like early and late blight. However, while previous models often focused on binary or limited multi-class classification, the proposed model extends classification to five categories, including Potato Leaf Roll and Potato Verticillium Wilt, addressing the need for broader disease coverage. Unlike simpler architectures or traditional machine learning approaches, which achieved accuracies between 90% and 97%, the proposed model with an enhanced **DenseNet-201** architecture achieved 97.2% accuracy, indicating superior robustness and precision. This study also introduced solutions to **class imbalance** and **overfitting**—challenges frequently noted but not fully resolved in previous work.
Review Outcome (Make a decision how to use/refer the obtained knowledge to prepare a separate and new methodology for your own research project)	To develop a methodology based on the provided study: • **Dataset and Data Preparation**: Use a relevant dataset, similar to the Plant Village dataset, and address any class imbalances with techniques like reweighting or augmentation. • **Image Preprocessing**: Apply techniques such as normalization, CLAHE, and resizing to enhance image quality before model input. • **Model Selection and Customization**: Choose an architecture (e.g., DenseNet, ResNet), customizing as needed for feature retention and efficiency. • **Training Strategy and Loss Function**: Use a suitable loss function (e.g., reweighted cross-entropy) to handle class imbalance, and experiment with training parameters (batch size, learning rate, etc.). • **Evaluation Metrics**: Apply accuracy, precision, recall, F1 score, and a confusion matrix to evaluate model performance comprehensively. • **Benchmarking**: Compare with existing models to highlight improvements or specific advantages • **Future Improvements**: Consider scalability for other plants, adaptability to varied imaging, or real-time detection. This approach will help create a targeted, robust methodology for your project based on this framework