

Project Design Phase-II

Data Flow Diagram

Date	20 October 2023
Team ID	Team-592955
Project Name	Potato Disease Classification
Maximum Marks	4 Marks

Data Flow:

Data Collection: This phase entails gathering a diverse dataset of potato leaf images from sources such as field surveys, image repositories, or other means. The collected images are then stored in a raw data repository.

Data Pre-processing: Raw potato leaf images are pre-processed to prepare them for model training. This can include resizing images, normalizing pixel values, and applying data augmentation techniques to enhance dataset diversity.

Model Training: In this stage, the pre-processed data is utilized to train a deep learning model, which learns to classify potato leaf diseases. The trained model is saved for future use.

Model Evaluation: The performance of the trained model is assessed using a separate dataset not used in training, to gauge its accuracy, sensitivity, and specificity in disease classification.

Model Deployment: This step involves deploying the trained model to make it accessible for real-world disease classification applications, either on local devices or in the cloud.

User Interaction: End-users interact with the deployed model through a user-friendly application or API, enabling them to submit potato leaf images for disease identification and receive prompt results.

Flow in the model:

Data is sourced from the Data Repository, and it undergoes transformation in the Data Pre-processing phase to become pre-processed data.

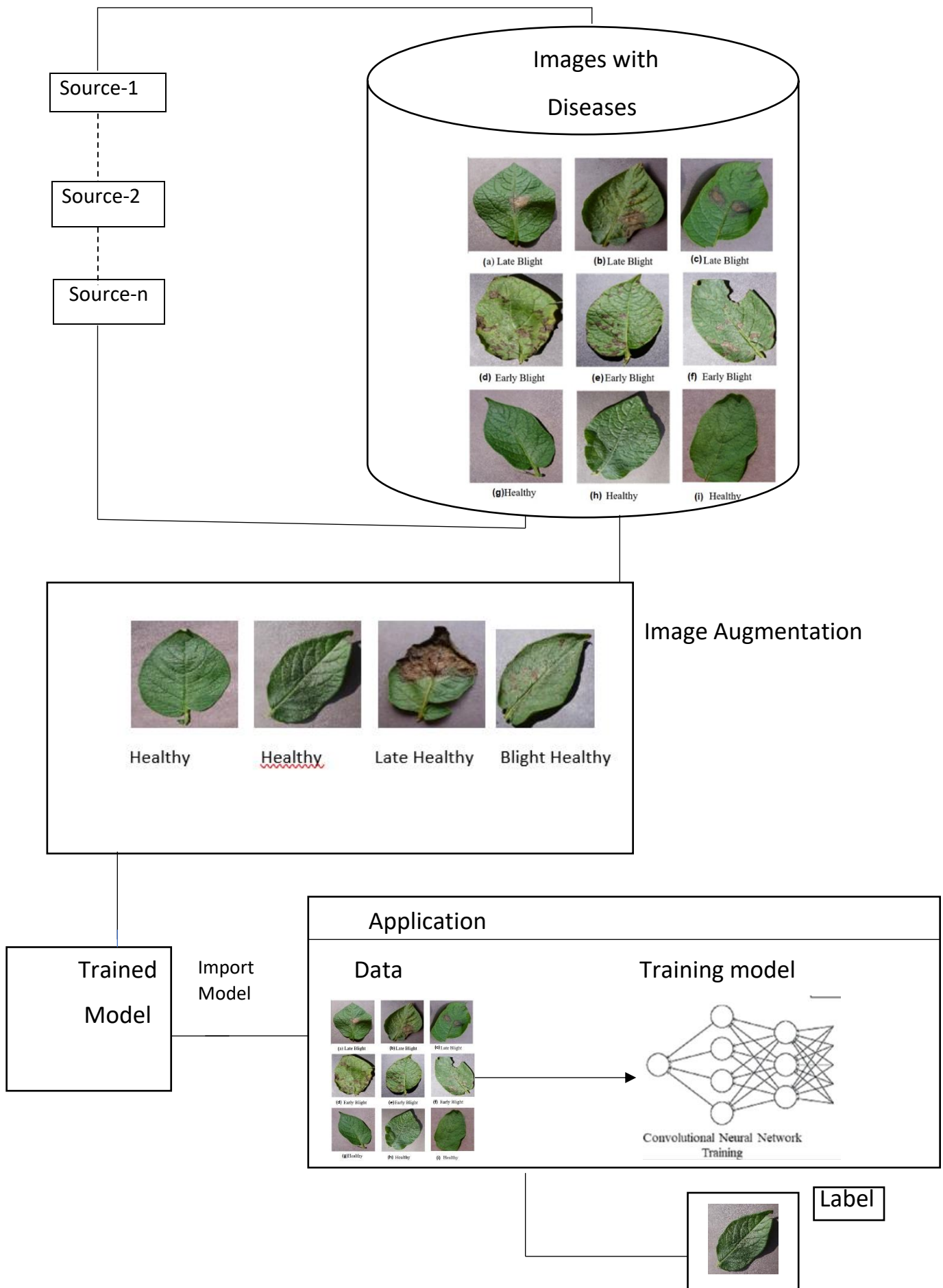
The pre-processed data flows to the Model Training process, where the model is trained using this data. The trained model can be used for classification.

In the Model Evaluation process, a distinct dataset is employed to gauge the model's accuracy and performance.

Once the model is ready, it can be deployed from the Trained Model Repository, making it accessible to users.

Users interact with the deployed model through the User Interaction process, submitting potato leaf images for classification.

Data Flow Diagram:



User stories:

User Type	Functional Requirement	User Story Number	User Story/Task	Acceptance criteria	Priority	Release
	Project setup and infrastructure	User – 1	To develop a project for potato leaf classification with the help of required tools and frame works.	Successful Configuration	High	Sprint 1
Coastal Guards	Development environment	User - 2	Compile comprehensive dataset of images showcasing various types of potato diseases to train a deep learning model for potato disease classification.	Gathered a diverse dataset of images depicting various types of diseases	High	Sprint 1
Individuals	Data collection	User – 3	Prior to model training, perform pre-processing tasks, including resizing the images, standardizing pixel values, and dividing the dataset into training and test subsets.	Pre-processing and the splitting of the plant village dataset	High	Sprint 2
Researchers and Academics	Data pre-processing	User – 4	Explore and evaluate various deep learning architectures (E.g.: CNN) to select the most suitable model for potato classification.	Exploring the different deep learning models	High	Sprint 2
Organizations	Model Development	User – 5	Train the designated deep learning model using the pre-processed dataset and closely observe its performance on the validation set.	Validation using the test part of the dataset	High	Sprint 3
Institutions	Training	User – 6	Integrate data augmentation methods to enhance the model's resilience and accuracy.	Test it with the augmented dataset	Medium	Sprint 4

	Model Deployment	User – 7	Deploy the trained deep learning model as an API or web service and make it accessible for potato classification.	Check the scalability of the model	Medium	Sprint 4
	Testing and Quality Assurance	User – 8	Thoroughly test the model and web interface, identify and report bugs, fine-tune parameters, and optimize performance based on user feedback.	Creating the web application	Medium	Sprint 6