# Abstract

**\*\*This project also aims to demonstrate how machine learning can be a powerful tool in the hands of novice developers, making it accessible and scalable for real-world horticultural applications.\*\***

Apple is among the one of the most nutritious fruits across the globe and most economic driven fruit for New Zealand. However, there are several challenges in recognising different types of apple leaf diseases such as the apple scab, cedar apple rust, and black rot, which can severely reduce the apple yield and the overall quality. This project proposes to deploy and compare in detail two different supervised machine learning models for apple leaf disease detection using Support Vector Machine and Convolutional Neural Network (CNN). Both proposed machine learning models will be trained and evaluated using the dataset. The primary objective of project is to identify the effectiveness of both the proposed machine learning models for the apple leaf disease classification, and also to highlight the overall performance, complexity, and suitability for deployment in real-world agricultural scenarios.

**\*\*SVM model hyperparameter tuning (e.g., kernel choice, C value) may introduce additional learning complexity for new users.\*\***

**\*\*Another challenge is dataset imbalance, which may require resampling or class weighting strategies for more robust learning.\*\***

**\*\*For local deployment, the final CNN model may be converted into a standalone executable using Python-to-EXE converters for offline farm use.\*\***

# Introduction

**\*\*As the apple industry continues to adopt digital solutions, this research also aligns with national initiatives encouraging the integration of AI in agriculture.\*\***

**\*\*Recent studies have emphasized the growing relevance of machine learning in real-time plant health monitoring systems, particularly in orchard management environments.\*\***

Agricultural productivity is highly important to the global food security, yet it is also constantly threatened by various plant diseases, which decreases the productivity yields of the apples. [1]. Specially in the horticulture sector, apple cultivation is more famous and plays a significant economic role in many countries across the globe, including New Zealand [2]. However, apple trees are often vulnerable to different types of leaf diseases such as apple scab, cedar apple rust, and black rot, which can drastically affect yield quality and quantity [3]. In the past more traditional methods have been used by the farmers for disease detection based on manual inspections and self-knowledge, which has been cumbersome, time consume and not practical in a large-scale farm setup.

With the emergence of latest artificial intelligence (AI), particularly with different types of Machine Learning (ML) model have been introduced solve problems such as disease detection in farming. ML models help to optimised to automate disease detection through image-based classification systems[4]. In general, ML models not only improve the speed and consistency of diagnosis but also has the capability to enable early intervention, reducing large amount of crop loss and help the farmer to minimise the use of harmful pesticides. Among the most popular ML techniques which is most suitable for image classification are Support Vector Machines (SVM) and Convolutional Neural Networks (CNN)[4], [5].

This proposal outlines a supervised learning project aimed at detecting apple leaf diseases using image dataset. The proposed project will implement and compare two different ML models, which are SVM classifier, and a custom CNN developed from scratch using TensorFlow. Both models will be trained and evaluated using the obtained dataset, which includes four different classifications labelled as apple scab, cedar apple rust, black rot, and healthy leaves.

The key objective of this project is to assess the effectiveness, complexity, and real-world applicability of both traditional and deep learning approaches in agricultural disease detection. This project provides a solid foundation for beginners in machine learning while contributing to the broader field of precision agriculture.

# Overview of Data Set

**\*\*Class distribution appears balanced enough to support both classical and deep learning algorithms without substantial bias.\*\***

**\*\*The images were captured using uniform backgrounds, which allows for better feature extraction during preprocessing.\*\***

The dataset selected for this project has been obtained from [Kaggle](https://www.kaggle.com/datasets/lavaman151/plantifydr-dataset) website. This dataset is popular and have been widely used in different plant-based disease detection research and also provides labelled collection of different categories of apple leaf images that are most suitable for training supervised ML models for the delivery of this project. All the dataset images are in JPEG format, and they were captured under better consistent lighting and better background conditions to reduce noise and enhance clarity.

For the purpose of this project, only the apple leaf subset were used to the ML models, which contains approximately 13,000 images distributed across four distinct classes, each representing a specific disease condition or a healthy state:

* Apple Scab – 3232 images
* Cedar Apple Rust – 2553 images2.
* Black Rot – 3105 images
* Healthy Apple Leaves – 4234 images

In this dataset each of the images has been labelled with its relevant disease classification. The key objective is to build a model that can accurately categorise any given apple leaf image into a one of the four identified category based on the visual symptoms on the leaf such as the spots, colour differentiation and leaf textures.

# Methodology

**\*\*The implementation will also include code modularization to allow easy swapping and tuning of different model architectures.\*\***

**\*\*The two selected models represent a contrast between classical ML and modern deep learning, providing educational value and performance benchmarking.\*\***

## Preferred Learning Methods

In this project the student will implement and compare the following two supervised ML models, which will be explained in below section 3.1.1 and 3.1.2

### Support Vector Machine (SVM) with HOG Features

The SVM ML algorithm is mostly used for classification related tasks. This model performs well on moderately sized datasets, and it is very easy to implement. SVM ML model will use manually extracted the leaf image features, specifically containing the Histogram of Oriented Gradients (HOG), which can also capture leaf shapes and textures and the characteristics of apple leaves from the obtained dataset images. The features will then be used to train an SVM classifier to distinguish between four apple leaf conditions, which are Apple Scab, Cedar Apple Rust, Black Rot, Healthy Apple Leaves.

### Custom Convolutional Neural Network (CNN)

CNNs is a well-known ML model, which is mostly used for image classification due to their capability to automatically learn hierarchical visual features and patterns from raw images. For the purpose of this project a simple CNN model will be built using the tool such as TensorFlow/Keras. This approach will make it easier for the CNN to learn directly from the pixel values without manual feature extraction.

### Justification of Learning Methods

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Strengths | Weaknesses | Fit for This Project |
| Support Vector Machine (SVM) with HOG | - Easy to understand and implement- Low computational cost- Performs well on small, clean datasets- Interpretable results | - Requires manual feature extraction- Less accurate on complex image patterns- Struggles with large-scale or high-resolution data | - Suitable for benchmarking- Ideal for learning core ML principles- A solid classical baseline for image classification using extracted HOG features |
| Custom Convolutional Neural Network (CNN) | - Automatically learns complex visual features- High accuracy in image classification- Scalable to larger datasets- No need for manual feature engineering | - Requires more computational resources- Longer training time than SVM- Slightly more complex to implement | - Well-suited for detecting subtle visual symptoms in leaf images- Enables end-to-end learning- Provides a foundation for future expansion or deployment on mobile/web platforms |

Both models fall within the supervised learning paradigm, where labelled examples of apple leaves are used to train the model to predict new unseen labels. The project will compare their performance and computational efficiency to understand the strengths of each approach in an agricultural setting.

# Evaluation of Proposed ML Models

**\*\*The evaluation will also involve monitoring the models’ prediction consistency across different lighting conditions using augmented test images.\*\***

**\*\*In addition, Area Under Curve (AUC) and Receiver Operating Characteristic (ROC) metrics will be considered to analyze the true positive rate against the false positive rate.\*\***

To assess the performance of both models—Support Vector Machine (SVM) and Convolutional Neural Network (CNN)—a set of standard supervised learning metrics will be used. The dataset will be divided into training (70%), validation (15%), and testing (15%) sets.

Evaluation Metrics:

* Accuracy – the proportion of correctly classified leaf images.
* Precision – the ability of the model to avoid false positives.
* Recall – the ability of the model to capture true positives.
* F1-Score – the harmonic mean of precision and recall.
* Confusion Matrix – a visual representation of true vs. predicted classifications to identify class-specific performance.

For the CNN, performance will also be tracked across epochs using training and validation loss and accuracy curves. For the SVM, cross-validation may be used to reduce the risk of overfitting and improve generalization. The comparative analysis of both models will help determine which method is better suited for detecting apple leaf diseases in terms of accuracy, simplicity, and usability.

# Deployment

The final CNN model, if it demonstrates strong performance, will be prepared for deployment. Deployment scenarios may include:

* A desktop GUI using Python libraries such as Tkinter or Streamlit, allowing users to upload images and receive predictions.
* A web-based tool hosted on platforms like Heroku or Hugging Face Spaces for real-time usage.
* Future potential deployment on mobile devices using TensorFlow Lite (optional, not in scope for this beginner-level project).

Although the SVM model is not ideal for real-time deployment due to manual feature engineering requirements, it will remain useful as a benchmark or lightweight desktop tool for educational purposes.

# Challenges

Several challenges are anticipated during the project:

1. Limited Dataset Variability: The PlantVillage dataset was captured under controlled conditions. This may limit the models' ability to generalize to real-world leaf images with varied lighting, backgrounds, or leaf damage.
2. Image Preprocessing and Feature Extraction: Preparing images for the SVM model (grayscale conversion, HOG extraction) can be error-prone for beginners and may impact the quality of the input features.
3. Model Overfitting: The CNN, if not properly regularized, may overfit to training data. Techniques such as dropout, data augmentation, and early stopping will be considered.
4. Resource Constraints: Training CNNs on personal hardware may be slow. This will be addressed using cloud resources like Google Colab with free GPU support.
5. Model Comparison: Ensuring a fair and unbiased comparison between SVM and CNN models requires consistent data preprocessing, evaluation metrics, and validation protocols.

Addressing these challenges will enhance the learning experience and lead to a more robust and well-documented machine learning project.

# Requiored Tools

2.4 Preferred Programming Language and Tools

| Tool / Library | Purpose |
| --- | --- |
| Python | Primary language for model development |
| TensorFlow / Keras | CNN model creation and training |
| scikit-learn | SVM classifier implementation |
| OpenCV / skimage | Image preprocessing and HOG feature extraction |
| NumPy / Pandas | Data manipulation |
| Matplotlib / Seaborn | Visualization and result plotting |
| Google Colab | Free GPU support for CNN training |

Python is chosen for its simplicity, readability, and the availability of mature libraries for both traditional ML and deep learning workflows.

3. Summary

**\*\*The comparative evaluation of the models will not only guide implementation decisions but also support future research in AI-assisted agriculture.\*\***

This project will develop and compare two beginner-accessible supervised learning models for apple leaf disease detection using image data: an SVM with manually extracted HOG features and a basic CNN built from scratch. Both models will be trained and validated on the apple subset of the PlantVillage dataset, using four common disease categories. Performance metrics such as accuracy, precision, recall, and F1-score will be used to evaluate each model. The project aims to provide practical insights into machine learning techniques for precision agriculture while building foundational skills in image classification, feature extraction, and deep learning.

[1] R. Gebbers and V. I. Adamchuk, “Precision Agriculture and Food Security,” *Science (1979)*, vol. 327, no. 5967, pp. 828–831, Feb. 2010, doi: 10.1126/SCIENCE.1183899.

[2] AI Forum NZ, “Artificial Intelligence for Agriculture in New Zealand.” Accessed: Mar. 17, 2025. [Online]. Available: https://aiforum.org.nz/wp-content/uploads/2019/10/Artifical-Intelligence-For-Agriculture-in-New-Zealand.pdf

[3] H. G. Lawrence, “ADOPTION OF PRECISION AGRICULTURE TECHNOLOGIES FOR FERTILISER PLACEMENT IN NEW ZEALAND,” 2007.

[4] N. Zhang, M. Wang, and N. Wang, “Precision agriculture—a worldwide overview,” *Comput Electron Agric*, vol. 36, no. 2–3, pp. 113–132, Nov. 2002, doi: 10.1016/S0168-1699(02)00096-0.

[5] P. Usendi Research Scholar *et al.*, “Precision farming in horticultural crops”, doi: 10.33545/26174693.2025.v9.i2h.3834.