User Experience Analysis with Fashion/Tourism Design

Evie Snuggle

A project proposal submitted as part of the requirements
for the degree of BSc. (Hons.) in Computer Science
at the School of Computing and Digital Technology,
Birmingham City University, United Kingdom

November 2021,
CMP6200 Individual Undergraduate Project 2021–2022

Abstract

An undergraduate honours project proposal for the investigation into the impact of various diseases and conditions affecting plants with a proposed artificial-intelligence based solution for plant disease identification.

As part of the proposed solution, this will cover the areas of computer vision, machine learning / artificial intelligence, deep learning, data exploration and convolutional neural networks.

Project Origin

The initial idea for this project originated from a tutor list by...

Contents

List of Figures

Chapter 1

Introduction

1.1 Project Title

Find & Identify plants diseases using AI.

1.2 Background

People growing plants at home have difficulties knowing the best course of action for taking care of their plants. For example, an incredibly common source of frustration with plants is when over-watering causes a disease known as 'root rot', which can lead to people worsening the condition by over-watering further. This leads to a positive feedback loop which can cause the plant to wither away. (Cockett, H., 2021)

In addition, there are other common pathological diseases in plants such as fungal diseases, blight, canker, or mildew which can be difficult for someone not familiar for plants to properly diagnose and know what to do. (Lambe, R.C., 1978; Cockett, H., 2021) This project aims to help people diagnose plant issues quicker and more effectively, to prevent plants from dying.

1.3 Key Themes/Topics

This project will explore the use of computer vision processing with convolutional neural networks (CNN) to aid diagnosis of various plant conditions using pre-existing datasets available within online repositories such as Kaggle. This project will be based on Python's TensorFlow library, also using supplementary libraries such as Keras, pandas and numpy. Google Colaboratory, an online service which provides the use of CPU & GPU resources for research & testing of machine learning, may also be used. Colaboratory would be a suitable choice as Google is known for being prominent in the field of artificial intelligence.

Chapter 2

Aims and Objectives

2.1 Project Aim

This project aims to assist in the identification and diagnosis of assorted plant diseases or conditions. If successful, this model will perform better at identifying the underlying cause of a plant's symptoms than a selected control group.

2.2 Project Objectives

- Understand the common problems that are faced by people taking care of plants and identify which plant diseases would be most important to recognise.
- Explore existing datasets available online and analyze those that would be most suitable for this project.
- Organize datasets as needed, ensure datasets are without defective data as well as tagged appropriately, and pre-process images as necessary.
- Create multiple machine learning models to identify plant diseases.
- Evaluate the effectiveness of different models in plant disease identification, with comparisons to a control group study.

Chapter 3

Initial Design Plan

To achieve the objectives above listed above, data exploration techniques will be used to analyze existing datasets to determine their suitability. A convolutional neural network will be used with TensorFlow to process images of plants using computer vision to achieve multiclass/multinomial classification of the plants into different disease categories.

After these images of plants are classified, then there will be an analysis of the accuracy of the results with an evaluation of the machine learning model's effectiveness for plant disease identification.

A major limitation could arise if the datasets that have been identified include inconsistent or improperly tagged data. This could be overcome by manually tagging images, however this may even require finding a new dataset. Another limitation, a lack of accuracy when identifying plant diseases, could hinder the ability for this project to be of any use. To ensure that this is not the case, a control group study of amateur home gardeners will be evaluated against multiple machine learning models.

3.1 Project Schedule

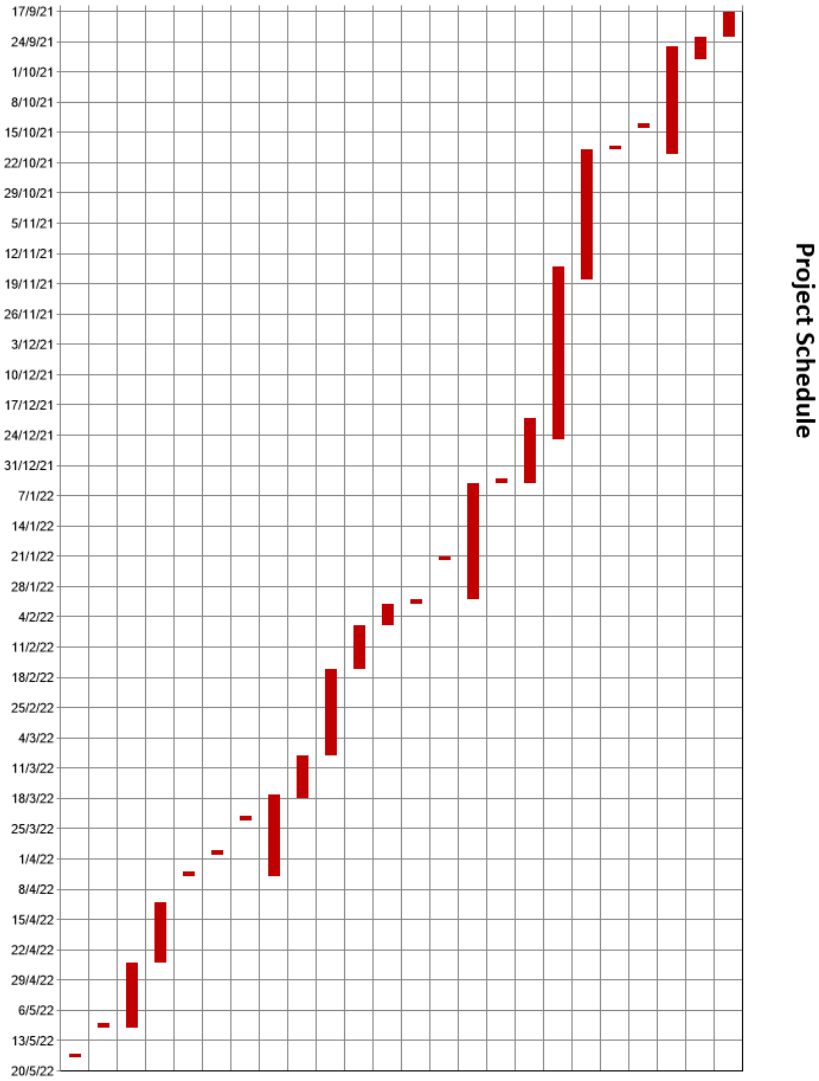

Figure 3.1: Project Schedule Gantt Chart

3.2 Rationale

While there is existing research into the common diseases that affect plants, there does not appear to be any proposed solutions based upon this existing research. Not only will this project help the general public but may be useful to further explore the current limits of computer vision processing.

3.3 Resources

3.3.1 Computer Hardware

- Dedicated graphics processing unit with 'tensor cores', either available locally or remotely using a service such as Google Colaboratory.
- Modern desktop computer.

3.3.2 Datasets

- Pre-existing plant disease dataset.
- Google Scholar.
- University library resources.
- Access to other student projects for evaluating and developing critique of final year project methodologies.

Chapter 4

Evaluating Previous Modules & Existing Research

4.1 Previous Modules Impact

Previous modules within my course impact the approach taken in this project, such as the 'Innovation Project' module in which groups had undertaken a project in a semi-independent fashion. This helped develop independent research methods and working towards achieving specific project goals. In addition, the 'Computer Programming' module helped solidify understanding in the Python programming language.

4.2 Critique of Past Final Year Projects

As part of the initial literature search, two past final year projects were particularly relevant to this project and within the field of agriculture. The first named "Smart Automated Plant Watering System" by Abid (2018), found in their research study, as part of figure 5, that the majority of participants practiced a lifestyle that was "too busy" to properly maintain plants. This evidence supports the aims and objectives of this project, making it easier for people to identify plant diseases which lowers the time investment required. It is notable that this questionnaire only included six participants for this question which severely limits the reliability of the results found.

A second final year project titled "Smart Irrigation System" by Bist (2018), identified that plants can sometimes be damaged if watering occurs above the soil, such as on the plant leaves. It also states that plants can be damaged from both over or under-watering, especially if unevenly distributed. Unfortunately when attempting to access the source for these assertions, which was last accessed Feburary 2018, it was no longer available online as of October 2021.

Both of these projects are thoroughly independently researched, although this project is sufficiently different enough that much cannot be reused, especially as the prior research conducted is specific and narrow-focused. For example, the design and implementation of various moisture sensors cannot be applied to this project. Research into the water requirements of various plants may have been valuable for this project, but this appears to not have been explored in detail.

Chapter 5

Literature Search Methodology

Online databases will be used for searching for existing literature and sources. Most importantly, Google Scholar will be used for searching for existing publications, as these are often from established academic journals. To determine the relevance of sources Google Scholar allows people to view how often a work has been cited, which is an indication of peer-review and reliability. In addition, I will contrast any literature I cite across multiple independent sources to verify each work's validity.

An online library, Kaggle, is available online to find existing datasets that are available for machine learning, as these datasets can be much larger than would be possible to create independently, which makes them more suitable for machine learning. (Brownlee, 2020) In addition, the Birmingham City University library will also be used to find prior student projects to analyze.

Key pieces of literature reviewed included Cockett, H. (2021), which stated: "Plants are living beings, they can get funny little bumps and marks occasionally, just like us and they can't always be explained. If you're confident that your environmental conditions are ideal, and what you're dealing with is indeed something out of the ordinary that can't be explained then how do you identify a fungal infection?".

In addition, Jones (2007) was referenced which included: "One of the errors made by home gardeners is to select a variety not well adapted to their growing conditions" which is later expanded upon, that home gardeners transferring instructions from one climatic zone to another will result in a disappointing crop.

Aworinde et al. (2013) and Stewart Jr., Abudayyeh and Stewart (2018) were both in the initial literature search but these were determined to not be relevant enough for inclusion, as there wasn't sufficient detail exploring the issues people face when home gardening.

5.1 Existing Datasets Identified

There is a large dataset, "PlantVillage", which includes thousands of plants categorised by different diseases. (Emmanuel, 2018) This will be split into training and testing data, which is often known as the train-test split procedure. (Brownlee, 2020)

5.2 Search Terms

- Google Colaboratory
- Keras
- TensorFlow
- Plant disease classification
- CNN model training & optimisation
- Data exploration machine learning
- Plant disease pathology
- Different plants and effect on disease resistance
- Global plant hardiness zones
- Plant disease epidemiology
- CNN Image pre-processing
- Computer vision
- Convolutional neural networks
- Machine learning model evaluation
- Deep learning
- Data classification
- Plants grown at home
- House plants disease

Chapter 6

Risk Assessments

6.1 Potential Issues for Project Completion

- If there is a lack of people available for control group study, then it will be more difficult to assess if the machine learning application is successful in the stated goals.
- Hardware available not being suitable for machine learning development could cause the project to slow.
- Lack of experience using convolutional neural networks could result in inadequate classification and misidentification.

None of these risks individually would cause the project to fail to complete. If any risks to the project above do cause issue, the project can still continue but may have a limited evaluation.

COVID-19 Contingency

This project should not be limited by restrictions from the COVID-19 measures in the United Kingdom, as it primarily focuses on house plants and artificial intelligence. Both of these will be unaffected by any home isolation measures in-place or in the future. If anything unforeseen causes disruption, the project advisor will be notified.

6.2 Health and Safety

ASSESSMENT No: 18128599-1	**HEALTH & SAFETY RISK ASSESSMENT**	HIGHEST RISK LEVEL 9 (Medium)

Activity:	*Individual Honours Project*	**Assessor:**	*Evie Snuggle*	**Date:**	*2021-10-18*
		Location:	*Birmingham, UK*		

Scope of assessment: *An individual honours project will be conducted. This could involve research into multiple different fields and data collection.*

Risk Assessment Review:
(Periodic, following any changes or accident/ incident)

Reason for Review	Date and Signature

SEVERITY AND PROBABILITY RATINGS

SEVERITY RATING (SR) (OUTCOME)		PROBABILITY RATING (PR) (LIKELIHOOD)	
SR	DESCRIPTION GUIDANCE	PR	DESCRIPTION GUIDANCE
1	None	1	Highly Unlikely
2	Minor Injury or illness	2	Unlikely
3	Over 7 Day Injury or illness	3	Possible
4	Major Injury or illness	4	Likely
5	Fatal	5	Highly Likely

RISK LEVEL MATRIX

Risk Level = Probability x Severity		SEVERITY (OUTCOME) 1	2	3	4	5
PROBABILITY (LIKELIHOOD)	5	Low 5	Med 10	High 15	High 20	High 25
	4	Low 4	Med 8	Med 12	High 16	High 20
	3	Low 3	Med 6	Med 9	Med 12	High 15
	2	Low 2	Low 4	Med 6	Med 8	Med 10
	1	Low 1	Low 2	Low 3	Low 4	Low 5

High 15-25	**High Risk** Control measures must be introduced to reduce the degree of risk prior to the activity/ process proceeding.
Medium 6-12	**Medium Risk** Activity/process can continue, but additional controls may need to be introduced to further reduce the risk prior to start of task.
Low 1-5	**Low Risk** Risk is being adequately controlled; but further control measures may further reduce the risk.

Figure 6.1: Risk Assessment Cover Sheet

HEALTH & SAFETY RISK ASSESSMENT

Hazard	Persons at Risk	Existing Control Measures (if any)	Initial Risk Rating SR	PR	Risk Level	Additional Control Measures (if required) *Transfer onto Action Plan Sheet 1*	Final Risk Rating SR	PR	Risk Level
Project Incomplete by Deadline	☐ *Staff* ☐ *Visitors* ☒ *Students* ☐ *Public*	A basic set of deadlines have been set which can give an outline of a project schedule.	1	4	4 (Low)	A comprehensive project schedule will be developed and as long as this is properly adhered to, this shouldn't be an issue.	1	2	2 (Low)
Mental Health Difficulties due to COVID-19 Lockdown (inc. stress)	☐ *Staff* ☐ *Visitors* ☒ *Students* ☐ *Public*	The University has a mental health & support team in-place which can help mitigate this risk.	3	3	9 (Medium)	Any students involved in this project will be made aware of the University's mental health facilities.	2	2	4 (Low)
Lack of People Available for Control Group Study	☐ *Staff* ☐ *Visitors* ☐ *Students* ☒ *Public*	People may be needed to act as a 'control' to test the project against. This will involve a minimal amount of anonymous data collection.	0	3	3 (Low)	Early contacting of any people who may be involved will help ensure that they are available for the control group.	0	2	2 (Low)
Unexpected Costs Associated with Project	☐ *Staff* ☐ *Visitors* ☒ *Students* ☐ *Public*	Some research has been done into the feasibility of the project, with a look into the hardware & costs required.	0	2	2 (Low)	Additional research into the hardware requirements of machine learning will be conducted before the project implementation is started.	0	1	1 (Low)
	☐ *Staff* ☐ *Visitors* ☐ *Students* ☐ *Public*								
	☐ *Staff* ☐ *Visitors* ☐ *Students* ☐ *Public*								

Figure 6.2: Risk Assessment Table

6.3 Ethical Considerations

Due to this project being in machine learning, there can be unconcious biases in the datasets used to build models. A common example is systemic racism, which could be introduced in this project by only including plants found in Western countries. To address this specific issue, there must be plants from a variety of climates included in the datasets used.

6.3.1 Data Collection

Some anonymous data will be collected from people to compare the effectiveness of my project to the general public. This will require consent forms to be completed for each person giving responses. This data will be anonymous by default, there is no need to know the name of each individual. There is very low risk of de-anonyimisation of people's identities from the data that would be collected.

The data collected will include people's assessment of a plant's condition, possible diseases and most appropriate action to take for the plant's health.

Bibliography

Abid, S., 2018. Smart automated plant watering system. *Birmingham City University Digital Library*, p.9. Available from: https://bcu-atom-arkivum-net.ezproxy.bcu.ac.uk/index.php/smart-automated-plant-watering-system.

Aworinde, D.O., Erinoso, S.M., Ogundairo, B.O. and Olanloye, A.O., 2013. Assessment of plants grown and maintained in home gardens in odeda area southwestern nigeria. *Journal of Horticulture and Forestry*, 5(2), pp.29–36. Available from: http://doi.org/10.5897/JHF2013.0294.

Bist, P., 2018. Smart irrigation system. *Birmingham City University Digital Library*, pp.2–4. Available from: https://bcu-atom-arkivum-net.ezproxy.bcu.ac.uk/index.php/smart-irrigation-system.

Brownlee, J., 2020. Train-test split for evaluating machine learning algorithms. *Machine Learning Mastery*. Available from: https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/.

Cockett, H., 2021. Root houseplants - a discussion on fungal disease in houseplants. *Root Houseplants*. Available from: https://www.root-houseplants.com/blog/houseplant-tips/the-f-word-a-discussion-on-fungal-disease-in-houseplants/.

Emmanuel, T.O., 2018. Plant disease detection using keras. Available from: https://www.kaggle.com/emmarex/plant-disease-detection-using-keras/data.

Jones, J.B., 2007. *Tomato plant culture*. 2nd ed. CRC Press. Available from: http://doi.org/10.1201/9781420007398.

Lambe, R.C., 1978. Identification and control of diseases of houseplants. *Virginia Polytechnic Institute and State University*, 166. Available from: https://vtechworks.lib.vt.edu/handle/10919/93657.

Stewart Jr., C.N., Abudayyeh, R.K. and Stewart, S.G., 2018. Houseplants as home health monitors. *Science*, 361(6399), pp.229–230. Available from: http://doi.org/10.1126/science.aau2560.