

DS6050 Project Literature Review: Toxic Plant Classification

Katherine Kelleher
School of Data Science
University of Virginia
Charlottesville, USA
kbk8vh@virginia.edu

Margaret Lindsay
School of Data Science
University of Virginia
Charlottesville, USA
jsk2zs@virginia.edu

Brian Nolt
School of Data Science
University of Virginia
Charlottesville, USA
frv3fp@virginia.edu

Abstract—Poisonous plants present health hazards to individuals who spend time outdoors across diverse geographic regions. While experts may have the knowledge required to identify poisonous plants, many individuals remain vulnerable to the misidentification of toxic vegetation, such as hikers, gardeners, and backyard enthusiasts. Drawing on a wide array of literature related to current machine-learning methods that have been applied to toxic plant classification, this paper proposes a project to leverage publicly available data to develop a deep-learning model that classifies images of plants by species and toxicity status. This type of model has the potential to empower individuals by allowing them to identify vegetation that pose health risks and prevent negative encounters with toxic vegetation.

I. INTRODUCTION

Poisonous plants growing around the world pose not only a significant health risk but also an economic threat to communities. While the old saying “leaves of three, let them be” may be useful for children in the continental United States to avoid an unfortunate encounter with poison ivy in the backyard, more sophisticated means of identifying plants from a wide range of regions, particularly whether a plant is poisonous, can protect individuals from physical harm and economic loss.

In a study investigating classifying medicinal and poisonous plants, researchers explain a farmer must recognize the plants growing in and around harvested fields to ensure poisonous plants are not harvested along with crops [1]. Ingesting poisonous plants can cause “burning of the mouth, skin irritation, diarrhea, headache, blurred vision, nausea, arrhythmia, dizziness, tremors” or in severe cases lead to death [2]. Moreover, farmers risk serious financial loss if their livestock die after consuming poisonous plants [1].

The need for developing mechanisms to identify poisonous plants is underscored by the reality that scientists continue to discover new plant species. Between 2004 and 2018, roughly 2,100 to 2,600 new plant species were discovered each year [3]. While botany experts have advanced methodologies for identifying a plant and whether it is poisonous, non-expert outdoors people and farmers would benefit significantly from having an accessible, non-destructive method of identifying plants they encounter as poisonous or not poisonous.

II. LITERATURE REVIEW

A. Summary of Existing Literature

Given the potentially grave consequences of misidentifying a poisonous plant, it is unsurprising there have been numerous studies that seek to improve on traditional plant identification methods using machine learning. Traditional methods of identifying plants can be time-consuming, requiring years of experience to develop expertise [1]. Deep learning computer vision models offer a faster alternative that would allow non-experts to safely and accurately classify plants as poisonous or non-poisonous.

Most studies published in the past five years focus on a limited subset of plant species, sometimes restricted to a particular region. Zuhri et al., for example, limited the study to three species of poisonous plants and three species of non-poisonous plants [4]. The dataset contained 50 images of each of the six species, totaling 300 images. In another study, Noor, Noor, and Elmezain focus their research on the Arabian Peninsula, including 2,500 images of 50 unique Arabic plant species [5]. Another study that sought to categorize poisonous plants and visually similar plants used for medicinal purposes used a 900-image dataset comprised of three classes: oregano, poisonous, and weed [1]. A recent publication considers the same publicly available dataset used in this paper. The dataset contains between 950 and 1,000 images of five toxic and five non-toxic plants [6].

Zuhri et al. placed an emphasis on leaf texture and therefore sought to classify plants by first extracting features from the Gray Level Co-occurrence Matrix (GLCM), commonly used in texture analysis [4]. GLCM “calculate[s] the probability of the relationship between pixels that have exactly the same value as the pixel distance and use angles of 0°, 45°, 90° and 135°” [4]. Six attributes were used from GLCM: IDM (a measure of homogeneity), contrast, energy, entropy, and correlation [4]. The study also used five leaf shape features: slimness, roundness, diameter, perimeter, large (calculation of area value), and width. These 11 total parameters to develop a neural network.

In the proposed SCAM-Herb model developed by Azadina et al., the authors agree texture represents a pivotal feature when distinguishing between plant species and thus conclude

“more attention should be paid to modules that consider texture features” [1]. Based on the ResNetSt model, the SCAM-Herb model is a variation on ResNet with channel attention (CA) and spatial attention (SA) blocks [1]. The authors of this study sought to increase model performance by extracting meaningful features without sacrificing efficiency. Their “main objective was to detect and extract key features and local information from plant images, without the requirement for preliminary pre-processing procedures” [1].

In the study focused on classifying poisonous plants in the Arabian Peninsula, the authors also incorporated a residual neural network architecture; however, instead of focusing on attention mechanisms, they created a hybrid model that used ResNet50, EfficientNetB0, MobileNet, Xception, NAS-NetLarge, MobileNetV2, and InceptionResNetV2 to extract features. The features were then inputted into a convolutional neural network (CNN) along with the support vector machine (SVM) for prediction [5]. The SVM was incorporated to address the issue of overfitting and predict data [5]. The authors compared their CNN-SVM model to six pre-trained models NASNetLarge, InceptionResNetV2, ResNet50, Xception, MobileNetV2, and EfficientNetB0. The study’s proposed CNN-SVM model had the highest accuracy of the seven models tested [5].

Marwa et al. acknowledge CNNs have “been widely adopted for plant identification because they can automatically analyze important features like texture, color, and patterns with high accuracy” [6]. The authors developed a model using EfficientNet-B7, as they identified a lack of previous work applying this architecture to the challenge of poisonous plant classification. This study also took the unique approach of making the model easily accessible to users who might need real time assistance in identifying poisonous plants. Moreover, the model incorporated first-aid instructions in the event a user had contact with a poisonous plant [6]. While some other studies limited species to a particular region, Marwa et al. designed the protocol so the model could generalize to a wide range of plant species. They used data augmentation to expand the training dataset and make the model more applicable to a variety of different plants.

Azadnia et al. also used data augmentation to “increase and diversify” training data [1]. Specifically, the authors used Fast AutoAugmentation (FAA) to minimize computational complexity and improve performance. While some studies changed the image background to white during pre-processing steps, other studies intentionally chose datasets with a variety of different backgrounds to improve model performance in real world scenarios. As Shrikrishna Kolhar and Jayant Jagtap suggest, developing computer vision models that can accurately identify plant species despite complications such as “light conditions, complex background, overlapping leaves, and circadian movement of plant leaves” is challenging but wields the potential to protect against the health and economic risks of misidentifying poisonous plants [7].

B. Gaps in Existing Literature

The studies considered in this literature review all resulted in accuracy scores of 92% and above, which suggests neural networks, specifically CNNs, can be used to develop highly successful models for classifying plants as poisonous or non-poisonous.

While all studies reviewed for this project use neural networks to develop computer vision models that classify a plant as poisonous or not, each study uses a different architecture to accomplish the task. There appears to be a need to attempt to reproduce the results of these studies and evaluate model performance on a range of plant species from different regions. Many of the papers considered did not discuss in detail the process for how the suggested model was developed in terms of tuning hyperparameters and inclusion of components. A more detailed investigation into hyperparameter optimization and an ablation study could be useful to determine if any further improvements could be made to the accuracy and efficiency of the suggested models.

III. PROPOSED FUTURE RESEARCH

A. Motivation

The motivation for this project is to apply modern deep learning models to the identification of common plants that are poisonous to humans. While adventure enthusiasts and horticulturalists might have the knowledge and skillset required to recognize different species of plants they encounter, the sporadic hiker, weekend gardener, or anyone who finds themselves in a brush for any reason (perhaps to fetch a ball at a park) most likely does not. This project aims to enable individuals with the capability to identify whether a plant captured within a photo is poisonous or not by utilizing advanced deep learning architectures and techniques.

B. Dataset

This project will utilize the **Toxic Plant Classification** dataset found on Kaggle [8]. This dataset contains nearly 10,000 images of 10 different classes of plants: five which are toxic and five which are not but are often mistaken for toxic variety. Each class of plant has 1000 images (a couple of them a few less). The toxic plant classes are western poison oak, eastern poison oak, western poison ivy, eastern poison ivy, and poison sumac. The nontoxic plant classes are virginia creeper, boxelder, Jack-in-the-pulpit, bear oak, and fragrant sumac. The images were scraped from the iNaturalist site. The dataset includes the slang name (like poison sumac), the scientific name (like *Toxicodendron vernix*) and its Herbarium 2022 Category ID for each image.

C. Project Proposal

The ultimate goal of this project is to build a model capable of correctly classifying images as either poisonous or nonpoisonous plants, so users have the ability to use the model to identify whether they have recently encountered a toxin. We understand that there are limits to the utility of simply classifying a plant as poisonous (or not). Someone who

has recently been in close proximity to a plant that our model deems poisonous may also be interested in understanding what type of plant is contained in the image, to better understand the optimal treatment plan for irritation caused by the plant. For this reason, we aim to both correctly classify the class of a plant contained within an image and whether it is toxic to humans. Our intention is to develop a model that can accurately characterize the plant's class while still reliably determining whether it is toxic to humans.

We plan to start with known image recognition models to see if we identify methods to replicate similar studies' accuracy. From there, we will experiment with the architecture and the hyperparameters in order to achieve the best accuracy balanced with efficiency for our test dataset. We will evaluate our model's results and accuracy by creating the appropriate confusion matrix for the plant classes that are identified by the model for each image.

D. Research Questions

1. How does the choice of deep-learning architecture and hyperparameter tuning affect the accuracy of poisonous vs non-poisonous plant classification across diverse species and backgrounds?
2. How does the choice of deep-learning architecture and hyperparameter tuning affect the accuracy of the classification of plant species?
3. How does the choice of training data, deep-learning architecture, and hyperparameter tuning affect the generalizability of poisonous vs non-poisonous plant classification for plants from a specific region?

REFERENCES

- [1] Azadnia, Rahim, Faramarz Noei-Khodabadi, Azad Moloudzadeh, Ahmad Jahanbakhshi, and Mahmoud Omid. "Medicinal and Poisonous Plants Classification from Visual Characteristics of Leaves Using Computer Vision and Deep Neural Networks." *Ecological Informatics* 82 (September 2024): 102683. <https://doi.org/10.1016/j.ecoinf.2024.102683>.
- [2] Hridoy, Rashidul Hasan, Fatema Akter, and Maisha Afroz. "An Efficient Computer Vision Approach for Rapid Recognition of Poisonous Plants by Classifying Leaf Images Using Transfer Learning." 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, July 6, 2021, 01–07. <https://doi.org/10.1109/ICCCNT51525.2021.9580011>.
- [3] Cheek, Martin, Eimear Nic Lughadha, Paul Kirk, et al. "New Scientific Discoveries: Plants and Fungi." *Plants, People, Planet* 2, no. 5 (2020): 371–88. <https://doi.org/10.1002/ppp3.10148>.
- [4] Zuhri, Mohammad Faishol, S. Kholidah Rahayu Maharani, Affandy, Aris Nurhindarto, Abdul Syukur, and Moch Arief Soeleman. "Classification of Toxic Plants on Leaf Patterns Using Gray Level Co-Occurrence Matrix (GLCM) with Neural Network Method." *Journal of Development Research* 6, no. 1 (2022): 1–5. <https://doi.org/10.28926/jdr.v6i1.202>.
- [5] Noor, T.H.; Noor, A.; Elmezain, M. Poisonous Plants Species Prediction Using a Convolutional Neural Network and Support Vector Machine Hybrid Model. *Electronics* 2022, 11, 3690. <https://doi.org/10.3390/electronics11223690>
- [6] Marwa, Muskaan, P. Aman Reddy, M. Nithish Reddy, and M. Satyanarayana Reddy. "Toxic Plant Classification Using Convolutional Neural Networks." *REST Journal on Data Analytics and Artificial Intelligence* 4, no. 1 (2025): 192–200. <https://doi.org/10.46632/jdaai/4/1/23>.
- [7] Kolhar, Shrikrishna, and Jayant Jagtap. "Plant Trait Estimation and Classification Studies in Plant Phenotyping Using Machine Vision – A Review." *Information Processing in Agriculture* 10, no. 1 (2023): 114–35. <https://doi.org/10.1016/j.inpa.2021.02.006>.
- [8] Elliot, Hans. "Toxic Plant Classification." 2022. <https://www.kaggle.com/datasets/hanselliott/toxic-plant-classification/data>