TREE

CLASSIFICATION

SYSTEM

Team: FUTUREX

Member Details:

Reg. No.	Index No.	CN No.
ICT/19/20/058	4996	RK788
ICT/19/20/103	5040	KS2404
ICT/19/20/107	5043	PH573

GitHub Link:
https://github.com/ramba99/AI-Tree-classification.git

Contents

Introduction

The objective of this assignment is to develop a neural network model capable of accurately distinguishing between three tree species.

We have chosen our three tree species as Veralu, Uguressa, and Weliwenna. These species are known to be visually similar which makes it hard to identify one by one. Accurate classification is essential for biodiversity studies, conservation efforts, and educational purposes within the university garden. This report outlines the entire process, from data collection to model training and evaluation.

Overview of the selected species

All these three can be found in the university garden and all these three have some important values attached to them.

The uguressa(Flacourtia indica) fruit has countless health benefits and medicinal value. The fruit, seed, root and leaves can be used for arthritis. It has iron and is good for anemic conditions.

Veralu(Elaeocarpus serratus) fruit is rich in minerals, vitamins, fibre and valuable antioxidants. Due to its nutritional content it possesses anti-inflammatory, antibiotic, anti-anxiety, analgesic, antidepressant and antihypertensive properties. It is used in rheumatism and as an antidote for poison.

Weliwenna (dimorphocalyx glabellus) is used in treatment while the ayurvedic usage of it. includes Dyspepsia and Rheumatism.

Dataset Preparation

Data collection involved measuring specific leaf characteristics from multiple samples of each species.

To be exact ,we have measured certain leaf characteristics of about 30 leaves from each species and in total of 90 leaves.

We have selected our features of the leaves to be measured as ,

- Length of the leaf
- Width of the leaf
- Length to width ratio(l/w)
- Top angle of leaf
- Length of the petiole

The following sample images will illustrate the methodology we have used to collect data.
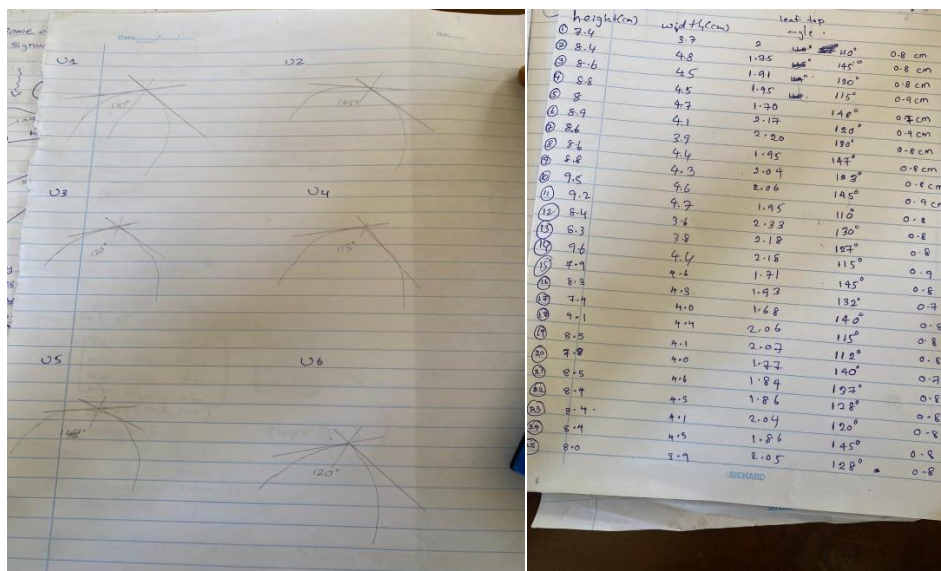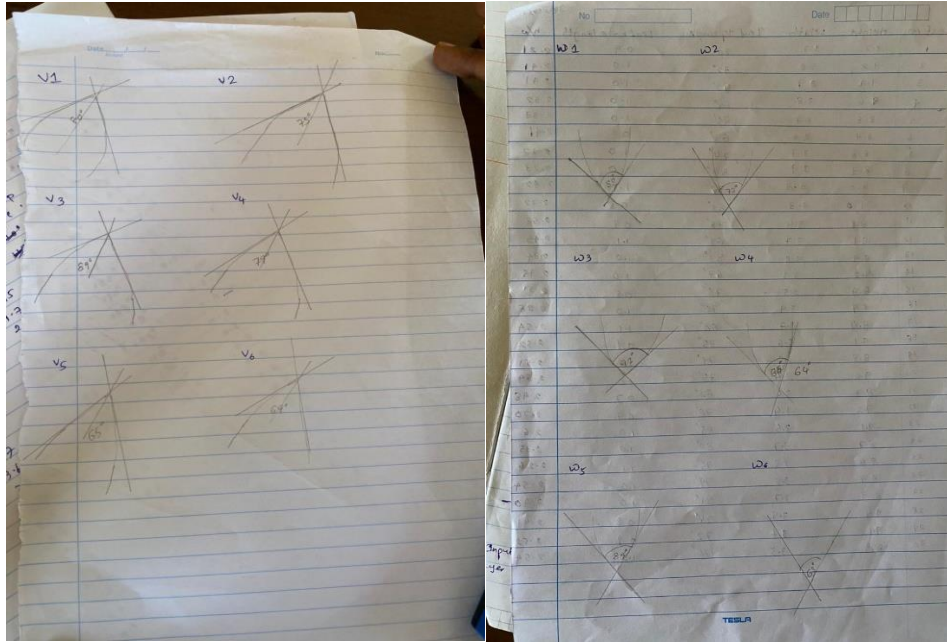
As the images illustrate we had named the leaves with letters and numbers and measured all the features mentioned above and made the dataset with the help of those measurements.

A total of 90 samples were collected 30 from each.

The dataset was split into training and testing sets with an 80-20 ratio.

Architecture of the proposed model

The neural network model for the identification process was designed using TensorFlow and Keras.

Input Layer: Accepts the five features.

Dense Layer 1: 128 neurons, ReLU activation, with dropout regularization (50%).

Dense Layer 2: 64 neurons, ReLU activation, with dropout regularization (50%).

Output Layer: 3 neurons (one for each species), softmax activation.

This design aims to balance model complexity and overfitting risk, given the relatively small dataset.

Training

The model was trained a laptop with an Intel i5 processor and 8GB RAM. The Adam optimizer was used with categorical cross-entropy loss. Hyperparameters such as batch size (32) and epochs (100) were chosen based on preliminary experiments. K-fold cross-validation (k=5) ensured robustness and generalization of the model.

The following steps shows the training process implemented in the Jupyter Notebook:

1) Loading and Preprocessing the Data: The dataset was loaded into a pandas DataFrame, and then the features were scaled using StandardScaler. The target variable was encoded with the help of LabelEncoder.

2) Building the Model: A sequential neural network model was constructed using Keras. The model architecture included two hidden layers with ReLU activation functions and dropout regularization to prevent overfitting.

3) Compiling the Model: The model was compiled using the Adam optimizer and categorical cross-entropy loss function.

4) Training the Model: The model was trained using k-fold cross-validation (k=5) to evaluate its performance. The final model was then trained on the full dataset with an 80-20 split for validation.

5) Evaluating the Model: The model's performance was evaluated using accuracy metrics. The final accuracy was calculated by averaging the validation accuracy from the k-fold cross-validation.

You can see all these features in our source code which was uploaded in github

https://github.com/ramba99/AI-Tree-classification.git

hyperparameter optimization techniques used and their impact on model performance

Experiments were used to fine-tune the hyperparameters of the model. Model performance and training time were balanced by the selected parameters, which were a batch size of 32 and 100 epochs. A 50% dropout rate was used to reduce overfitting.

Quantitative and Qualitative Results

During the k-fold cross-validation, the average validation accuracy was 92.34%. After training on the complete dataset, the final model obtained a validation accuracy of 93.21%. Strong model performance was also shown by the loss metric.

According to a qualitative evaluation, the model classified the majority of samples correctly; a small number of misclassifications were probably caused by feature overlap between species.

Discussion on Results

The great accuracy of the model shows that it can differentiate between the three species using observable characteristics. So these results can be very advantageous for many , as these species are very important biologically.

Summary

This study describes the creation of a neural network model for the classification of Veralu, Uguressa, and Weliwenna, three kinds of trees that are visually similar. An 80-20 split of a dataset consisting of 90 samples was made using measurable parameters such as petiole length, width, top angle, and leaf length. The training and testing sets were then separated. The model attained an average validation accuracy of 92.34% after being trained and validated in a Jupyter Notebook environment. The model's capacity to identify between the species is demonstrated by its excellent accuracy; nevertheless, more work with larger datasets is required to improve generalizability and resilience.

References

1) Srilankan floral and agricultural heritage -
   https://www.dilmahconservation.org/arboretum/plants-and-trees/uguressa-katulovi-dikpatana-waldivul--e3f56fe42ba85290173d7c46d0e98d61.html

2) Wikipedia-https://en.wikipedia.org/wiki/Elaeocarpus_serratus

3) Institute of Ayurweda -
   http://www.instituteofayurveda.org/plants/plants_detail.php?i=497&s=Local_name

4) Deep learning with python,tensorflow,Keras-

   https://www.youtube.com/watch?v=wQ8BIBpya2k&list=PLQVvvaa0QuDfhTox0AjmQ6tvTgMBZBEXN