Research Design II

Justin Borg, Mario Mallia-Milanes
Institute of Information & Communication Technology
Malta College or Arts, Science & Technology
Corradino Hill
Paola PLA 9032
{justin.borg.e21729, mario.mallia.milanes}@mcast.edu.mt

I. Introduction, Positioning, Research Onion

Nowadays botanical research and flower classification have become crucial for conservation efforts, various fields and medicinal research. Moreover, as research suggests, flowering plants are one of the more important species in the world, providing a source of life and food for a multitude of other living creatures like birds, insects, animals and also humans, furthermore there exist around 369000 different species of flowering plants, found across the world [1]. Thus, this study presents an approach of using deep learning models and Computer Vision to conduct flower image classification.

A. Description of Theme and Topic Rationale

The theme of the paper is on the use of deep learning models, in particular Convolutional Neural Networks (CNN), for the classification of Maltese endemic and non-endemic flowering plants. Furthermore, by comparing different iterations of the same model, through the application of data augmentation layers and preprocessing techniques, different results were created and evaluated. Ultimately, the goal of this study is to evaluate the feasibility and efficiency in using a CNN model to classify various flowering plants.

The topic rationale stems from the need to promote botanical research, protect biodiversity and support the process of identifying various flowering plant species. Most traditional classification methods found in the botanical sector, tend to be very labor-intensive and subjective, therefore the application of computer vision through a CNN model is a great step towards automating and subsequently improving the accuracy and efficiency of plant recognition, especially for the unique flora in Malta.

B. Positioning and Research Onion

1) Positioning: The positioning of this research situates itself within the interdisciplinary fields of botanical science, computer vision and environmental conservation. Its forefront ideology includes the application of deep learning technologies, specifically CNNs in the field of plant classification, specifically focusing on the diverse existing flora, Maltese endemic and non-endemic. This research sought out to bridge the gap between traditional methods of botanic classification and the more modern computational approaches, in aim of finding a more efficient and accurate plant identification process.

2) Research Onion:

a) Philosophical Stance: The Philosophical Stance at the outermost layer of this study, approaches a positivist methodology, relying on measurable data which is obtained from quantifiable data coming from the image classifier itself, in order to make unbiased and objective conclusions regarding the feasibility and effectiveness of using machine learning models in classifying Maltese endemic and non-endemic flowering plants.

b) Research Approach: The Research Approach layer reflects a great commitment to a direct and clear methodology. Wherein general theories of deep learning and computer vision are applied to the specific context of flowering plant classification. This approach is based on existing theories and literature, as it intends to test the hypotheses regarding the plausible use of CNNs in plant identification tasks.

c) Research Strategy: Moving inward, the Research Strategy layer of the study employs an experimental design by constructing different variations of the same deep learning model. By slightly changing variables and applying different preprocessing layers and data augmentation techniques, different accuracy results are presented and the model with the best performance data is gathered. Thus, this strategy is aligned with the scientific method, allowing for a systematic investigation into what creates a better performing CNN model and what are the important variables to be considered, in the context of their use in botanical classification.

d) Research Choices: Delving deeper into the Research Choices layer, the study engages in a mono-method quantitative kind of research. By utilising numerical data generated by the image classification models themselves, the performance of said CNN model is evaluated. Thus the choice is driven by the research questions, which are best answered through quantitative analysis and evaluation of results and testing, this clearly depicts whether the model is accurately and effectively providing a solution to the problem in hand, botanical classification.

e) Time Horizon: This study adopts a cross-sectional design in terms of time frames, capturing data at a single point in time, that being primarily after the model itself has been trained and is given flower images to classify. This design is appropriate to this study given the scope and objectives of the project, that focuses on how various implementations of image preprocessing techniques produce classification accuracy.

f) Techniques and Procedures: At the core layer, the Techniques and Procedures layer details the particular methods used in this study. The creation of the dataset,the building of the CNN model and the testing of its performance is discussed in detail. The deep learning model of this study follows the Keras Sequential model, running on the TensorFlow platform, and is designed to be able to learn train and predict on a dataset made up of 8 different flowering plant classes, combining a total of 6511 images. Furthermore, study variables such as; the preparation of the data and how the dataset was created, the construction of the model, the training and how it was executed, the validation of the training, and other testing and training procedures, were meticulously and deeply described to ensure study transparency and replicability.

C. Background to this research theme

1) Overview of Computer Vision and Image Recognition: Computer Vision and Image Recognition are two very powerful tools which are nowadays widely used in a lot of different fields, even in everyday applications and scenarios. An evident example of this would be the use of image recognition in the application called Facebook, this world-renowned application uses image recognition in order to recognise a human's face, with only a few labeled images and it has last been reported to have an approximate 98% of accuracy [2].

2) Overview of Image Classification within Deep Learning models: Building a neural network for the use of Image Classification is usually comprised of various structured steps. Firstly, the process starts by finding or building a sufficient dataset which the model will be able to use for training and validation, next an architecture is chosen, usually the Convolutional Neural Networks (CNN) architecture is chosen, if Image Classification is the target goal. According to [3], the CNN architecture has plausible evidence to be prominent and efficient for this particular task. The model is then initialised, optimised and trained with the existing dataset. Finally, testing with new data is carried out to evaluate the model's accuracy and performance [1].

3) Flower Classification: The term flower classification is defined as arranging different flowering plants, according to their unique characteristics. In order to successfully identify and group a specific plant's taxonomy, various characteristics are analysed, these characteristics include, petal arrangement, shape, size, colour, and overall structure. Expert taxonomy is considered to be a field with shortcomings, as there is not enough professionals in this area, thus the implementation of automated recognition and classification has provided a great alternative [1].

4) Exploring the use of Image Classification to classify flowering plants: Given its broad variety of applications in different areas such as, plant observation, gardening, botanical research, and ayurvadic treatment, the application of Image Recognition in Flower Classification proves to be a necessity of high importance [1] [4]. Authors in [5], conducted a study in which they aimed to address the use of Neural Networks in conjunction with image processing, for the understanding of unique flower image features, particularly the Malaysian Blooming Flower. They accentuated that since Neural Networks were designed with the aim to *"solve complex problems such as pattern recognition and classification"*, hence we can deduct that such Neural Networks are more than capable of classifying complex datasets, such as flower datasets. In another study conducted by [6], the authors investigated the process of carrying out Flower Image Classification through a system that uses edge and color characteristics of a flower image, in order to classify it. The final results of this study showed their system obtaining an accuracy greater than 80%, and the system was deemed to be successful at flower image classification.

D. Hypothesis

This paper hypothesizes that the use of deep neural networks and image classification methods can be used to classify Maltese endemic and non-endemic flowering plants. Furthermore, it proves that classification is improved through the incorporation of various data augmentation and preprocessing layers. This study suggests that by enhancing the quality, variation and size of the dataset, it will lead to a model which is able to learn and generalize from the input images better, leading to improved flower image classification.

E. Research Aim and Purpose statement

The aim and main focus of this research is to develop and test a deep leaning CNN model, which is able to carry out image classification on Maltese endemic and non-endemic flowering plants. The purpose behind this study is twofold. Firstly, to propose a technological solution that can assist in accurately and efficiently identifying and classifying various plant species, and secondly, to investigate how various data augmentation layers and preprocessing methods can improve the accuracy of said deep learning models, in the field of botanical image classification, therefore further informing the fields of computer vision and machine learning.

II. Review of research methodologies & map

III. Reflection on the chosen Methodology

IV. Results, analysis and discussion.

V. Conclusion

References

[1] C. Narvekar and M. Rao, "Flower classification using cnn and transfer learning in cnn- agriculture perspective," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 2020, pp. 660–664.

[2] J. Kothari, "A case study of image classification based on deep learning using tensorflow," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 6, pp. 3888–3892, 04 2018.

[3] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, 2014, pp. 844–848.

[4] C. Gamit, "Literature review on flower classification," *International Journal of Engineering Research Technology (IJERT)*, vol. 4, no. 2, 2015.

[5] F. Siraj, M. A. Salahuddin, and S. A. M. Yusof, "Digital image classification for malaysian blooming flower," in *2010 Second International Conference on Computational Intelligence, Modelling and Simulation*, 2010, pp. 33–38.

[6] T. Tiay, P. Benyaphaichit, and P. Riyamongkol, "Flower recognition system based on image processing," *Proceedings of the 2014 3rd ICT International Senior Project Conference, ICT-ISPC 2014*, pp. 99–102, 10 2014.