CHAPTER 7

TRANSFER LEARNING FOR AUTOMATED MEDICINAL PLANT CLASSIFICATION

Dewi Nasien, M. Hasmil Adiya,
Rio Juan Hendri Butar-Butar, and Zuriahati Mohd Yunos

7.1 INTRODUCTION

Indonesia is a country with a rich biodiversity. In Indonesia's vast and diverse botanical landscape, the identification and classification of medicinal plants hold immense significance for traditional and modern healthcare practices. The rich biodiversity of this archipelago has endowed it with many plant species, many of which possess valuable therapeutic properties. Accurate and efficient classification of these medicinal herbs is pivotal for preserving indigenous knowledge, pharmaceutical research, and the sustainable utilisation of botanical resources. However, recognising medicinal plants may pose difficulties, particularly for individuals lacking expertise in the field, due to the similarity of the medicinal plants (Noor et al., 2022).

Machine learning (ML) is the most suitable method to automate the identification of medicinal plants. In recent years, ML, specifically transfer learning, has emerged as a promising tool in automated plant classification (Patel & Chaware, 2020; Wagle et al., 2022). Transfer learning leverages pre-trained models on extensive datasets to empower the recognition and classification of novel plant species, often with limited annotated data

(Goodfellow et al., 2016; Gulzar, 2023; Wani et al., 2019). This approach can potentially revolutionise the identification of medicinal plants, particularly in regions like Indonesia, where the vast array of species presents a formidable challenge for traditional taxonomic methods.

This paper presents an innovative method for automating the classification of medicinal plants in Indonesia by utilising the potential of transfer learning. Our approach utilizes a convolutional neural network (CNN) model previously trained on an extensive dataset of natural images. Subsequently, this pre-trained CNN is refined through fine-tuning using a dataset of nine indigenous medicinal plant species from Indonesia. The dataset comprises 1300 images of medicinal plants, divided into two sets: 80% for training and 20% for testing. This paper also compares three pre-trained models which are MobileNetV2, InceptionV3, and Xception, to get the best performance.

Compared to existing methods, our approach boasts several advantages. Firstly, it attains high accuracy even with a limited dataset of medicinal plants. Secondly, it excels at identifying medicinal plants under diverse conditions, including varying lighting and backgrounds. Lastly, it offers computational efficiency, rendering it suitable for real-time applications.

The potential application of our approach extends to the development of a mobile app for automated medicinal plant classification, which could empower non-experts to identify medicinal plants in the field. Integrating modern ML techniques with Indonesia's botanical heritage promises to enhance healthcare and conservation efforts by bridging the gap between traditional knowledge and cutting-edge technology. This research aims to facilitate a more efficient and accurate identification of medicinal plant species.

7.2 MACHINE LEARNING RESEARCH IN MEDICINAL PLANTS

In recent years, so much research has applied ML algorithms to automate the identification process of medicinal plants. Wagle et al. (2022) and colleagues conducted research that categorised medicinal plants using two methodologies: Utilising a CNN and implementing transfer learning with the Alexnet model. The CNN architecture used three layers with varying parameters. Notably, the CNN approach yielded an accuracy of 99.65%, while using transfer learning with Alexnet resulted in an even higher accuracy of 99.73% (Wagle et al., 2022). In 2018, Sabarinathan et al. (2018) presented a system employing CNN technology for identifying medicinal plants, which covered 50 different medicinal plant varieties and achieved an impressive accuracy rate of 97.80%.

A study by Akiyama et al. (2019) examined the effectiveness of CNN for mobile leaf identification in Hokkaido. The research team evaluated the performance of three specific models–VGG19, MobileNet, and MobileNetV2–In their ability to identify plant leaves accurately. The dataset employed in this investigation comprised 4,483 instances representing 33 distinct types of plant leaves. The training images for medicinal plants had dimensions of 320 × 240 pixels. Among the three models under scrutiny, MobileNetV2 outperformed the others, delivering the highest precision score of 0.993, a recall score of 0.992, and an F1-Score of 0.992.

7.3 MODEL ARCHITECTURE

This section explained the process involves which are data preparation, pre-processing, and model architecture.

7.3.1 Medicinal Plant Image Dataset

The study utilised a dataset comprising images of nine distinct medicinal plants, namely Roselle (*Hibiscus sabdariffa*), Butterfly pea (*Clitoria ternatea*), Binahong (*Anredera cordifolia*), Longevity spinach (*Gynura procumbens*), Bay leaf (*Syzygium polyanthum*), Daun sirih (*Piper betle*), Chinese hibiscus (*Hibiscus rosa-sinensis*), Cat's whiskers (*Orthosiphon aristatus*), and Mengkudu (*Morinda citrifolia*). Following data collection, unsuitable and redundant images were removed, resulting in a dataset consisting of 1,300 images. This dataset was