IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 5, October 2025, pp. 4123~4136
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i5.pp4123-4136

Enhanced classification of aromatic herbs using EfficientNet and transfer learning

Samira Nascimento Antunes, Madalena De Oliveira Barbosa Divino, Luana Dos Santos Cordeiro, Fernanda Pereira Leite Aguiar, Marcelo Tsuguio Okano
Program of Production Engineering, Paulista University, São Paulo, Brazil

Article Info

Article history:

Received Sep 21, 2024
Revised Jun 20, 2025
Accepted Jul 10, 2025

Keywords:

Aromatic herbs
Classification
Convolutional neural network
Deep learning
EfficientNet
Image identification
Plant identification

ABSTRACT

Herbs have long been used for culinary and medicinal purposes, as well as in religious rituals, due to their essential oils and aromatic properties. However, distinguishing between aromatic and medicinal herbs based on visual characteristics alone can be challenging. With recent advances in computer vision, plant identification from images has seen significant growth, offering promising applications in several domains. This article aims to evaluate the classification of aromatic herbs using the EfficientNet convolutional neural network (CNN) technique with transfer learning. The methodology used is experimental research, systematically manipulating variables to observe their effects on the object of study. The researcher plays an active role in this process, rather than being a passive observer. Based on the results and the literature review, it is evident that the objective of this research was achieved, as despite the opportunities for improvement in training to achieve accuracy above 0.8, it was possible to evaluate the classification of aromatic herbs using EfficientNet CNN through the transfer learning technique.

Corresponding Author:

Marcelo Tsuguio Okano
Program of Production Engineering, Paulista University (UNIP)
Rua Dr. Bacelar, No. 1212, Vila Clementino, São Paulo, Brazil
Email: marcelo.okano@unip.br

1. INTRODUCTION

The identification and classification of visually similar herbs present a significant challenge due to their resemblance in characteristics such as color, shape, and texture. Traditional methods often fail to distinguish these species with high accuracy, highlighting the need for more advanced technologies. Studies have shown that models based on convolutional neural networks (CNNs) can enhance this process, enabling automated and more precise identification. Herbs play a crucial role in both nutrition and phytotherapy, possessing historical and cultural significance. They have been widely used in rituals and religious ceremonies due to their aromatic and medicinal properties [1].

The wide variety of aromatic and medicinal herbs with similar shape, color, and geometry characteristics makes their visual differentiation difficult. However, recent advancements in computer vision have spurred rapid growth in plant identification research, with promising results that demonstrate high accuracy, precision, and practical real-world applications [2]. With digital technologies like artificial intelligence (AI), transfer learning, deep learning (DL), CNNs, and mobile devices, automating the classification process of aromatic and medicinal herbs has become feasible. Herb classification holds particular importance in medicine, plant science, and the food industry. While various parts of plants, such as leaves, flowers, fruits, seeds, and roots, can be used for species identification, leaves are especially advantageous due to their availability throughout much of the plant's lifecycle [3].

This study aims to evaluate specific classification techniques using the EfficientNet CNN and transfer learning to overcome these challenges. The research adopts an experimental methodology that allows for the controlled manipulation of study variables to analyze their impact on the object of investigation. This approach enables a comprehensive assessment of the system's behavior under different conditions, ensuring that the effects of each variable are systematically evaluated. The validation of the methodology is carried out through experiments that simulate various operational conditions, allowing the researcher to actively observe the influence of each variable [4].

The structure of this paper is as follows: section 2 presents the theoretical framework for the proposed topics including the methodology. Section 4 details the results and discussion. Finally, section 5 concludes with final observations and directions for future research.

2. THEORETICAL FRAMEWORK

2.1. Classification vs. image identification with EfficientNet convolutional neural network

The distinction between image classification and image identification is fundamental to understanding computer vision tasks. Image classification refers to assigning an image a single class or label. In contrast, image identification involves identifying one or more objects through bounding boxes associated with object classes. According to Russakovsky *et al*. [5], image classification is a problem where the goal is to assign a single class to an entire image. This process involves determining the general category to which the image belongs without considering the precise location of objects within the image. On the other hand, image identification, as defined by Bochkovskiy *et al*. [6], involves the localization and identification of specific objects or individuals within a complex image. In this context, the task is to recognize objects' presence and determine their locations and often their specific identities. The swift advancement of digital technology, combined with the growing integration of AI in agriculture, has opened new paths for innovative solutions in plant identification [7]. This process typically involves using object detection algorithms, such as you only look once (YOLO) or faster region-based convolutional neural network (Faster R-CNN), to identify regions of interest in the image. While image classification focuses on assigning labels to entire images, image identification involves detecting and identifying specific objects within the image. These distinctions are essential for understanding the different approaches and techniques used in machine learning.

This research focused on classification using the EfficientNet CNN. According to Tan and Le [8], the EfficientNet CNN proposes an innovative approach to designing scalable and computationally efficient CNNs. They introduce a scaling method consisting of depth, width, and resolution scales, allowing the model to be resource-efficient while maintaining or improving classification performance. Furthermore, Chenna [9] emphasize that the EfficientNet CNN represents a significant advancement in the design of CNN architectures. They note that the scaling method proposed by EfficientNet allows for effective adaptation to different dataset sizes and computational constraints, making it an attractive choice for various practical applications. The EfficientNet CNN utilizes conventional metrics for evaluating classification models, such as accuracy, to measure its effectiveness in image classification. According to Tan and Le [8], the image levels in EfficientNet refer to different variants of the EfficientNet architecture that are scaled in depth, width, and image resolution. These levels are represented by letters and numbers, such as B0, B1, B2, and B7, where B0 refers to the lowest resolution, and B7 refers to the highest resolution and requires greater complexity and computational capacity. Each image level has a specific combination of network depth, channel width, and input resolution. The higher the level number (e.g., B7 compared to B0), the greater the image's depth, width, and resolution, resulting in a more complex and powerful network capable of handling more demanding computer vision tasks. Table 1 summarizes this dynamic.

Table 1. EfficientNet model base according to image resolution

Model base	Resolution
EfficientNetB0	224
EfficientNetB1	240
EfficientNetB2	260
EfficientNetB3	300
EfficientNetB4	380
EfficientNetB5	456
EfficientNetB6	528
EfficientNetB7	600

This scalability approach allows EfficientNet CNNs to be adapted to different available computational resources and precision requirements, providing a flexible and effective solution for various DL applications in computer vision. According to Tan and Le [8], accuracy is an essential metric for evaluating the performance of the EfficientNet CNN in object classification tasks. They state that EfficientNet was designed to maximize accuracy on reference datasets such as ImageNet, ensuring that the model can make accurate and reliable predictions across various object classes. Furthermore, Antunes *et al.* [7] note that accuracy is a crucial metric for determining EfficientNet's practical utility in different applications. They also emphasize that a prominent level of accuracy is essential to ensure that the model is dependable enough to be used in real-world environments, where prediction accuracy is paramount.

2.2. Deep learning, convolutional neural network, and deep transfer learning

The concepts of DL, CNNs, and deep transfer learning (DTL) have been crucial in the field of AI. They demonstrate the ability to manage complex data and increasingly sophisticated tasks, resulting in a positive impact due to their wide range of applications in various fields. The growing popularity of classification methods such as deep CNNs enhances AI's potential in multiple disciplines, particularly computer vision [10].

DL is a computer-implemented algorithm that utilizes a neural network to select data features automatically. AI does not require any extra enhancements. It combines low-level attributes to create high-level attributes that describe the distributed nature of the data and sample attributes. Compared to traditional image recognition and detection methods, DL is more effective in image recognition and target detection. Today, the main types of networks are multilayer perceptron's, CNNs, and recurrent neural networks (RNNs) [11]. The most common type of DL for spatial pattern analysis is a CNN (also called ConvNets). CNNs aim to acquire spatial skills, such as edges, corners, textures, or other shapes that more accurately describe the intended class or volume [11]. DL methods, especially those based on CNNs, are popular in the agricultural domain for detection and classification purposes [11]. The core to learning these features is a series of transformations of the input data that occur at different spatial scales (e.g., through pooling). This facilitates identifying and combining low-level properties and conceptual ideas [11].

According to Lu *et al.* [11], CNNs typically comprise convolutional, pooling, and fully connected layers. The convolutional layer considers the local association of image content to derive features. Like other typical neural network models, CNNs consist of neurons organized in layers and thus capable of learning complex representations, with neurons between layers connected through weights and biases. The first layer is the input layer, for example, remote sensing data, and the final layer is the output, such as a predicted classification of plant species. Between the layers are hidden transformations that alter how the input space is transformed into one that aligns with the output. CNNs have at least one convolutional layer as a hidden layer that explores patterns. Other non-convolutional layers can also be incorporated. The convolutional layers include various filters that can be optimized; these layers transform the input or previous hidden layer information. The number of filters determines the thickness of a convolutional layer. The resulting transformations aim to expose significant patterns that address the problem at hand [12].

DTL is a method that attempts to reduce dependence and costs by utilizing knowledge from a source task to train a target task [13]. Transfer learning relaxes the requirement that training and testing data be independent and identically distributed, which motivates the use of transfer learning to address the problem of insufficient training data. In transfer learning, the process of training and testing data does not need to be identically distributed, and the model in the target domain does not need to be trained from scratch, which can significantly reduce the demand for training data and the time required to train the model in the target domain [14], [15]. Transfer learning allows you to leverage what you have learned in one situation to learn more rapidly in another. It is commonly used for object recognition and image classification through pre-trained CNN models. Different methods of altering a pre-trained CNN and when to use each method are of great interest for research. The effect of each method on training time and test set accuracy affects the way transfer learning is employed. The herbs selected for analysis included *Laurus nobilis, Rosmarinus officinalis*, and *Mentha spicata*. The bay leaf (*Laurus nobilis*) is characterized by lanceolate leaves with an alternate arrangement and short petioles, measuring 5-8 cm in length and 3-4 cm in width. These leaves are leathery, punctate, with revolute and wavy entire margins. The upper surface is glabrous and glossy, ranging from olive green to brown, while the lower surface exhibits a dull olive green to brown coloration with a prominent midrib and veins. *Laurus nobilis* is widely used in the food and pharmaceutical industries due to its antioxidant and antimicrobial properties [16]. *Rosmarinus officinalis*, from the Lamiaceae family, is an aromatic perennial plant with erect stems and blue-white flowers. Commonly known as rosemary and native to the Mediterranean region, its fresh and dried leaves are widely used as a seasoning and for herbal tea preparation. It is particularly noted for its anti-inflammatory properties demonstrated in preclinical in vivo models [17]. Finally, spearmint (*Mentha spicata*) is distinguished by its opposite leaves with short petioles and oblong to oval shapes with serrated margins. This herb plays a crucial role due to its rich composition of phenolic compounds, which are effective in treating cardiovascular diseases. Additionally, it is frequently used as a seasoning and in infusions [18].

2.3. Methods

This research followed a structured experimental procedure to assess the classification of aromatic herbs using the EfficientNet CNN with transfer learning. The methodology is presented in two main phases to allow for replication by other researchers: i) literature review and dataset preparation, and ii) experimental model training and evaluation.

2.3.1. Literature review and dataset preparation

Initially, a comprehensive literature review was conducted to provide a solid theoretical foundation on relevant topics, including DL, CNN architectures, transfer learning, and EfficientNet's specific application to plant classification. This review aimed to identify knowledge gaps and validate EfficientNet's suitability for this task. The literature review is essential for understanding the evolution of CNNs, enabling the identification of advancements, challenges, and research opportunities. Recent studies emphasize their importance in analyzing architectures, optimizing models, and exploring practical applications [19]. Contemporary authors emphasize the relevance of literature reviews in constructing scientific knowledge. Literature reviews is a study that examines and synthesizes an existing body of literature by identifying, questioning, and developing the foundations of a theory through the analysis of a body (or multiple bodies) of prior work [20]. Additionally, Ebidor and Ikhide [21] highlight the importance of including diverse sources, such as technical reports and specialized journals, to ensure a comprehensive approach in the literature review. The aim of this literature review was to seek the most recent contributions to understanding the topic. According to Barry *et al.* [22], regularly updating the literature review is crucial to contextualizing the study within the state of the art. The careful selection of source ensures the reliability and relevance of the information, which are essential for building a consistent theoretical framework [23].

Zoph *et al.* [24] introduced groundbreaking approach to scalable image recognition by developing transferable architectures through automated neural architecture search, significantly improving efficiency and accuracy in image recognition. This work laid foundation for further advancements in efficient model design, such as EfficientNet, particularly in resource-constrained environments. These improvements in network architectures have facilitated the development of more specialized techniques, such as semantic segmentation, which are essential for applications requiring high precision in visual detail. Consequently, the adoption of advanced semantic segmentation techniques to ensure the preservation of fine details in images is crucial for differentiating visually similar herbs. The use of methodologies like those presented by Shivaprasad and Wadhawan [25] highlights segmentation strategies that emphasize the preservation of subtle and essential features. This approach is vital for maintaining the integrity of visual information during image processing, enabling more accurate and effective classification by enhancing the critical details that distinguish each species. The dataset selected for model training consisted of images of three herb species: rosemary, mint, and bay leaf. The herbs were chosen based on their visual similarities in color, texture, and shape, which present challenges for image classification. The images were gathered from open sources and categorized based on their visual quality and relevance, containing individual leaf images as well as instances of leaves in bundles and mixed with other objects. This diversity was intended to replicate real-world conditions but also led to some classification challenges.

Given the challenges often imposed by the dataset size and the need to increase model robustness, the incorporation of self-supervised learning (SSL) techniques can be highly effective. Abdulrazzaq *et al.* [26] highlighted methods like missing part prediction and automatic colorization of black-and-white images offer promising approaches. These self-supervised methods enable the model to develop a deeper understanding of the intrinsic features of herbs without relying exclusively on labeled data. SSL has shown great potential in extracting discriminative representations, which are crucial for the precise distinction between visually similar aromatic herbs, allowing for more efficient learning and more accurate classification. In parallel, the precision in neural network selection and feature optimization plays a key role in enhancing classification performance. This is well illustrated in [27], which demonstrates the cross-disciplinary applicability of these techniques. The study shows how strategic adjustments in model selection and data treatment can significantly improve classification outcomes, offering valuable insights for addressing similar challenges in other technical domains.

Moreover, image quality is fundamental to the accuracy of the classification process. The integration of enhancement methods, such as deblurring and asymmetric spatial attention, as presented in the study [28], can significantly improve the sharpness and clarity of herb images. These enhancement techniques, applied during the preprocessing phase, are essential for addressing issues related to blurry or low-quality images, ensuring that the model receives higher-quality visual data. By improving data input quality, these methods enhance classification accuracy and reduce errors caused by inadequate image conditions.

2.3.2. Experimental model training and evaluation

The model training was conducted on Google Colab, using available graphics processing unit (GPU) resources to facilitate efficient processing. Python programming language was utilized, alongside key libraries such as TensorFlow, NumPy, Pandas, Matplotlib, and OpenCV. These tools allowed for precise manipulation and analysis of the image data.

– Transfer learning with EfficientNet: the EfficientNet CNN was pre-trained with ImageNet weights to leverage transfer learning, which accelerates training by adapting learned features from a large, general dataset to specific herb classification task. The model was fine-tuned for 50 epochs, and variants B0 through B7 were tested with resolutions ranging from 224 to 600 pixels. Each variant's configuration, including depth, width, and resolution scaling, was adjusted to achieve optimal accuracy and computational efficiency.

– Custom model training: in addition to transfer learning, a custom model was trained from scratch using only the dataset of rosemary, mint, and bay leaf images. This approach aimed to address the challenges encountered in transfer learning due to dataset limitations and verify the model's ability to learn exclusively from this specific dataset. For comparability, the custom model was also trained for 50 epochs, and variants B0 through B4 were tested.
– Metrics for evaluation: the primary evaluation metric was accuracy, defined as the proportion of correctly classified images out of the total classifications. Loss was also monitored to assess model convergence and detect potential overfitting. These metrics were recorded throughout training and on test sets to provide a comprehensive view of model performance.

The model's hyperparameters, such as learning rate, batch size, and number of epochs, were kept consistent across all training sessions to ensure comparability. The image data were split into training and test sets using an automated algorithm, maintaining balanced representation for each herb category. This setup is intended to enable other researchers to replicate the experiments and observe similar results under controlled conditions. This structured methodology allows for replication by detailing each experimental step, from dataset preparation to model evaluation, enabling future studies to build on these findings and refine herb classification techniques using CNNs and transfer learning.

2.4. Consolidation of AI techniques applied to research

The application of advanced AI techniques has become a robust and efficient approach for complex image classification tasks, especially in areas where subtle visual differences make recognition challenging. In this study, we used the EfficientNet architecture combined with transfer learning and data augmentation techniques to classify aromatic herbs, meeting the demand for an automated and precise solution. The EfficientNet architecture represents an innovation in the design of CNNs by introducing a compound scaling method that automatically adjusts the network's depth, width, and resolution based on the available computational resources and task complexity [8]. This approach is particularly valuable for plant classification problems, where the model needs to capture subtle variations in texture and color between different species. Recent studies have highlighted that compound scaling allows for high accuracy with optimized resource use, which is essential in large-scale image processing environments [25], [26].

Transfer learning was applied to leverage prior knowledge from models trained on large datasets, such as ImageNet, and adapt them for specific herb classification tasks. Transfer learning is a technique in which an algorithm learns image features within one domain and can then apply this acquired knowledge to a new domain, even with a smaller dataset [27]. The technique of transfer learning with fine-tuning enables starting with a model pre-trained for a specific task and then adjusting only certain layers of the neural network to adapt it for a similar yet distinct target task [28]. This technique is widely recognized for accelerating training and improving accuracy, especially when applied to limited or highly visually similar datasets [29].

Additionally, to improve generalization and reduce overfitting, data augmentation techniques were applied, including transformations such as rotation, brightness adjustment, and cropping. Data augmentation creates artificial variations of the training images, simulating a more robust dataset and helping the model handle variations in angle and lighting [30], [31]. Studies have shown that data augmentation is effective in plant classification domains, where the model needs to adapt to subtle changes in leaf appearance [32]. Finally, using accuracy and loss metrics was essential for evaluating and adjusting the model throughout the training process, helping to detect issues like overfitting or underfitting. In summary, combining EfficientNet, transfer learning, and data augmentation provides a solid framework for complex visual recognition tasks, such as the classification of aromatic herbs, demonstrating the potential of these techniques to enhance the accuracy and applicability of AI models in challenging contexts.

2.5. Comparative evaluation of convolutional architectures for aromatic herb classification

This study focused exclusively on using EfficientNet for the classification of aromatic herbs due to its compound scaling ability, which automatically adjusts the model's depth, width, and resolution according to the task's requirements. This characteristic makes EfficientNet computationally efficient and especially useful in resource-limited environments. However, a comparative analysis with other modern CNN architectures could provide a more comprehensive evaluation of the model's effectiveness and identify alternatives that may bring distinct advantages for herb classification. ResNet, for instance, stands out for its use of residual connections, which facilitate gradient flow in deep networks and mitigate the vanishing gradient problem. The model has one MaxPool layer, 48 convolutional layers, and a standard pool layer. In each identification block, as well as in each convolution block, there are three convolution layers [33]. Comparing ResNet with EfficientNet would help assess whether the generalization capacity of residual connections benefits the classification of classes with high visual similarity, such as aromatic herbs. DenseNet, in turn, adopts a densely connected layer architecture, promoting feature reuse and improving information propagation throughout the network. This can be particularly advantageous in datasets where fine visual details, such as texture and edges, are crucial for classification. The ability to reuse intermediate information allows DenseNet to capture subtle visual nuances, offering an interesting alternative to EfficientNet in terms of differentiating visually similar species [34].

MobileNet, a lightweight architecture, was specifically designed for efficiency on mobile devices, making it ideal for applications requiring low resource consumption, such as field-based plant classification. In practical use situations, such as identifying herbs in remote locations, MobileNet could be a beneficial choice, balancing accuracy and computational efficiency on devices with lower processing power. Studies indicate that MobileNet performs well on mobile devices, maintaining competitive accuracy levels compared to more complex networks [35]. Architectures such as Inception, which combines convolutions of various sizes to capture a wide range of visual features, and Xception, which enhances this approach with separable convolutions for greater efficiency, could also enrich this analysis. Inception is effective with images exhibiting high pattern variability, while Xception combines accuracy and efficiency, making it suitable for contexts where balancing computational cost with feature extraction capacity is essential [36].

3. RESULTS AND DISCUSSION

3.1. Results

Table 2 presents the test set's results regarding the model training through the transfer learning process, considering the CNN EfficientNet pre-trained with the weights of the ImageNet dataset. Table 3 (see in Appendix) presents the evolution of accuracy and loss throughout the training process, providing a detailed view of the model's performance under different configurations. Analyzing these data is essential to understand how variations in network architecture and training parameters impact the final results. This analysis is crucial for optimizing the machine learning process, ensuring that the model is not only efficient but also capable of achieving the highest possible accuracy. Due to the models' accuracy being close to 0.5, the pronounced fluctuations of accuracy throughout the training, the average accuracy between 0.6 and 0.7 considering the test sets, and the confusion of results when testing with different images, it was decided to train the EfficientNet CNN without weights. Training a model considered customized, meaning the model learns only from the input dataset and with null input weight, aimed to identify the probable causes for the accuracy values under 0.8 and the different confusions during testing. Table 4 presents the test set's results concerning the model's training without the transfer learning process, considering the EfficientNet CNN without input weights and with only the selected aromatic herbs dataset. Notably, due to computational capacity, the models were trained to consider resolutions B0 through B4. Table 5 shows the evolution of accuracy and loss throughout the training sessions.

Regarding the training time, it is noted that it increased according to the resolution. Additionally, the time is considerably longer for a model learning without input weights. This dynamic is expected because, according to Tan and Le [8], the image levels in EfficientNet refer to different variants of the EfficientNet architecture, and the higher the level number, the greater the depth, width, and resolution of the image, resulting in a more complex and powerful network capable of handling more demanding computer vision tasks, hence the longer training time. When analyzing the loss, it is noticeable that both in the transfer learning process and in the custom process, it increases as the resolution increases. This is also expected due to the better resolution of the image pixels, which enables objects to be identified more accurately.

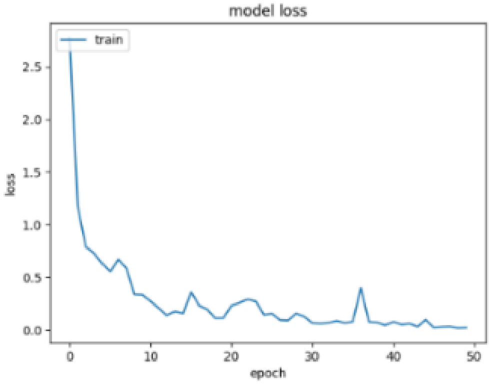
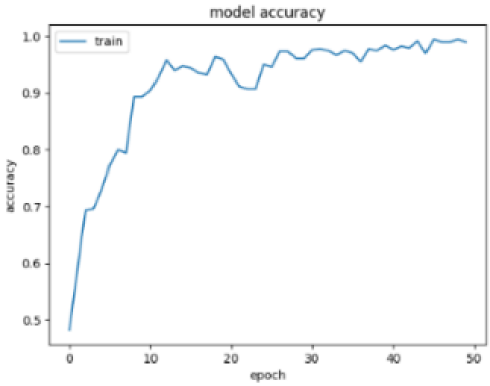
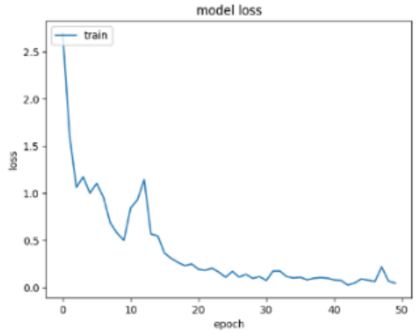
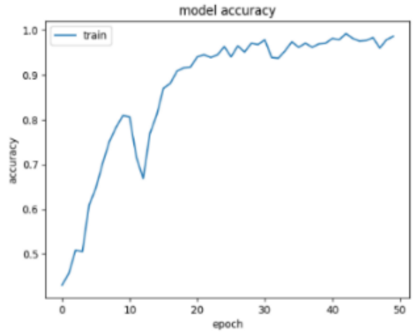
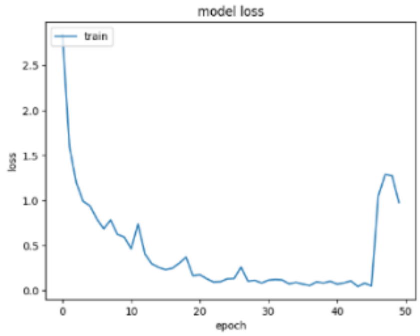
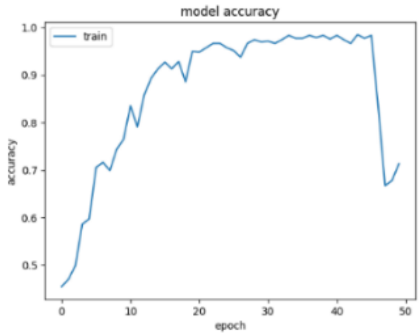
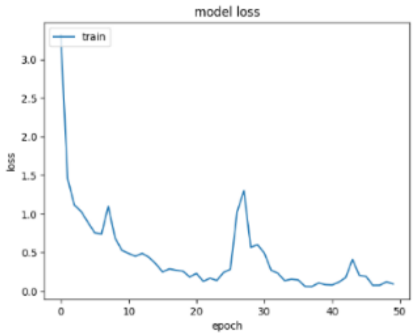
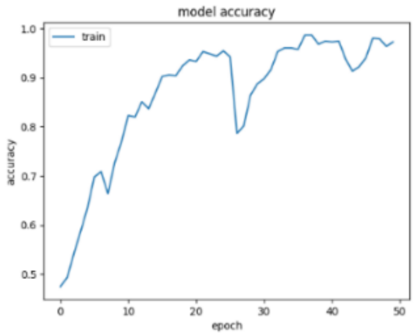
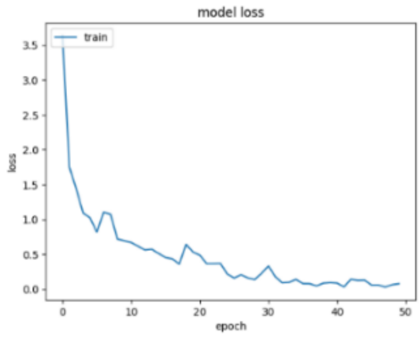
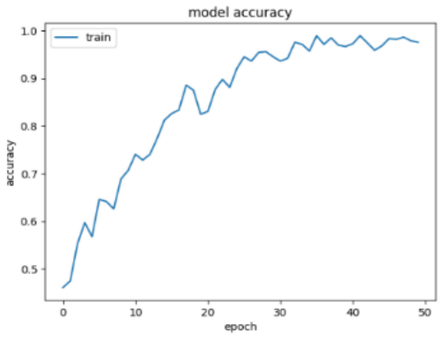
Table 2. Summary of EfficientNet model results using transfer learning

Model variant	Resolution	Epochs	Training time	Loss	Accuracy
B0	224	50	12m30s	0.79	0.71
B1	240	50	18m33s	0.77	0.51
B2	260	50	24m32s	0.80	0.77
B3	300	50	42m35s	0.72	0.77
B4	380	50	01h39m	0.73	0.80
B5	456	50	03h32m	0.84	0.57
B6	528	50	06h09m	0.69	0.74
B7	600	50	09h44m	0.77	0.65

Table 4. Custom model training results (without transfer learning)

Model variant	Resolution	Epochs	Time	Loss	Accuracy
B0	224	50	51 m 42 s	0.94	0.88
B1	240	50	01 h 11 m	1.09	0.80
B2	260	50	01 h 37 m	2542	0.40
B3	300	50	02 h 50 m	0.96	0.80
B4	380	50	06 h 36 m	1.53	0.68

Table 5. Evolution of training considering the process of transfer learning

Resolution	Loss	Accuracy
B0		
B1		
B2		
B3		
B4		

The accuracy varied according to the resolution, meaning that the higher the resolution, the higher the accuracy. Additionally, the variation in accuracy in the transfer learning process had more significant oscillation than that in the custom model, which showed a growing learning trend over the epochs. This suggests that the custom model was able to learn more effectively from the available dataset, resulting in a more stable performance over time.

3.2. Discussion

This study addresses significant gaps observed in previous research on the classification of aromatic herbs using computer vision techniques. While prior studies have explored the use of CNNs and transfer learning for plant identification, they primarily focused on distinct plant species with visually apparent differences. These studies often emphasized general plant classification rather than the nuanced differentiation required for visually similar aromatic herbs, such as rosemary, mint, and bay leaf. Consequently, they did not explicitly address the influence of subtle variations in color, shape, and texture that are critical for accurately classifying these herbs.

This research specifically investigates the application of the EfficientNet CNN architecture to tackle these challenges, examining how transfer learning can be fine-tuned to improve classification accuracy for visually similar plant species, thereby filling a key gap in the existing literature. The use of transfer learning allowed the EfficientNet model, pre-trained on the ImageNet dataset, to be successfully adapted for specific aromatic herb classification. This technique proved efficient for accelerating training, reducing computational costs, and maintaining accuracy.

Notably, lower-resolution pre-trained models, such as EfficientNet B0, showed substantially shorter training times without significantly compromising accuracy. Higher-resolution models, including EfficientNet B3 and B4, demonstrated increased accuracy, achieving values over 0.8 in certain cases. However, the computational cost associated with higher resolutions, from EfficientNet B5 to B7, was considerably higher, with training times up to five times longer, which limits their practical feasibility for resource-constrained devices. This observation underscores the importance of balancing resolution with computational efficiency, especially in field applications where resources may be limited.

Although custom-trained models (trained from scratch) achieved competitive accuracy, they required significantly longer training times and displayed greater stability in accuracy over the course of training. Transfer learning models, in contrast, exhibited some fluctuations in accuracy across different resolutions, suggesting that the pre-trained model might be sensitive to the specific characteristics of the herb dataset. The variability observed in accuracy and loss over the epoch's points to potential overfitting, particularly with higher resolutions, highlighting the importance of employing data augmentation techniques to improve model robustness and enable better generalization to new images of herbs.

The study identified that model accuracy could be impacted by the presence of images of herb bunches mixed with images of individual leaves in the dataset. Thus, a key recommendation for future research is to refine the dataset to include only isolated leaf images, which could help reduce classification confusion. Additionally, further exploration of different resolution configurations and implementing fine-grained visual classification (FGVC) techniques are recommended to improve the model's ability to differentiate subtle variations between herb species.

In comparing transfer learning-based models with custom-trained models, each approach presents distinct advantages. Transfer learning significantly reduces training time and enables rapid results, whereas custom models offer greater stability in accuracy throughout training, especially at intermediate resolutions. Although more time-consuming, custom models demonstrated higher resilience to overfitting, showing more consistent performance on unseen data a valuable characteristic in real-world applications.

In terms of accuracy, mid-resolution models, particularly EfficientNet B2 to B4, seem to offer an optimal balance between computational efficiency and generalization capability, maintaining competitive accuracy without the high computational costs associated with higher-resolution models. This study provides a balanced perspective on the strengths and limitations of each approach, offering practical insights for the development of efficient models in aromatic herb classification and similar classification contexts. Table 6 summarizes the key findings and observations from the study, highlighting model efficiency, the impact of resolution, variability in performance metrics, and recommendations for future enhancements.

The effectiveness of the aromatic herb classification model depends significantly on the selection and optimization of the neural networks employed. The techniques discussed in [37], explore feature fusion from multiple neural network architectures. This approach aims to combine the strengths of different networks to capture a broader range of discriminative features. Network selection can enhance model accuracy and optimize the learning process by leveraging information fusion to address the specific visual similarities of aromatic herbs, ensuring more reliable classification results.

Table 6. Key findings and observations

Aspect	Findings
Model efficiency	Transfer learning reduced training time but achieved comparable accuracy to custom models.
Resolution impact	Higher resolutions (B3, B4) improved accuracy, but increased computation time significantly.
Loss and accuracy variability	Transfer learning models showed greater variability in accuracy across resolutions.
Generalization issues	Custom model generalized well with lower accuracy oscillation but required extended training.
Future enhancements	Suggested refining the database with isolated leaf images and exploring different resolutions.

Accurate herb identification faces challenges due to the high visual similarity between different herb types. Studies have explored complementary strategies to address these challenges, including:

- SSL, this technique reduces reliance on labeled datasets by allowing models to learn more relevant representations of herbs from unlabeled data, enhancing classification by emphasizing the intrinsic features of the leaves [26].
- Network selection and information fusion, by optimizing the choice of neural network architectures and fusing information, this approach improves classification accuracy. According to the Zafar *et al.* [27], this strategy can be applied to determine the best-performing version of EfficientNet for herb identification.
- Feature preservation in images, advanced segmentation techniques help maintain subtle details in images, preventing the loss of crucial information necessary for distinguishing visually similar species [25].
- Image quality enhancement, methods for improving the sharpness and clarity of medical images can be adapted to ensure higher classification precision in herb analysis by providing higher-quality visual data before automated analysis [28].

Together, these techniques form a robust framework for enhancing the performance and reliability of automated herb classification models, addressing both data quality and model optimization.

4. CONCLUSION

This study successfully demonstrates the application of EfficientNet with transfer learning for aromatic herb classification, achieving notable accuracy while highlighting areas for database refinement and the importance of balancing resolution with computational efficiency for improved model generalization. This review has key limitations, including reliance on a specific image dataset of aromatic herbs, testing in a controlled environment, and a focus mainly on accuracy and computational efficiency, with limited attention to interpretability and real-time performance. Given these limitations, future research should aim to extend this investigation across diverse datasets that include a broader array of plant species and different environmental contexts to enhance the model's robustness and applicability. It is also recommended to explore higher levels of data diversity, including varying image qualities, lighting conditions, and even seasonal variations in plant appearance, to improve model generalization and adaptability. Furthermore, research should examine how FGVC and other specialized techniques can be integrated to enhance the classification accuracy for species with subtle visual distinctions. Expanding model interpretability to better understand the decision-making process of CNNs could also provide valuable insights, especially for practical applications in agriculture and botany. We propose exploring interpretability approaches such as gradient-weighted class activation mapping (Grad-CAM) and local interpretable model-agnostic explanations (LIME). These techniques would allow us to observe the regions of the image that most contribute to classification, providing transparency to the model and enabling researchers to validate whether the model is using valid botanical features, such as texture patterns, edges, and leaf veins. Expanding model interpretability to better understand the decision-making process of CNNs could provide valuable insights, especially for practical applications in agriculture and botany. The implications of this research are significant for fields that depend on accurate plant identification, such as agriculture, medicine, and the food industry. Automated herb classification could streamline processes, reduce human error, and increase efficiency in applications ranging from botanical research to commercial product development. As models continue to improve in accuracy and adaptability, they could be integrated into mobile or IoT devices, providing on-site plant identification and further democratizing access to AI-driven botanical classification tool.

FUNDING INFORMATION

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)–Finance Code 001.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Samira Nascimento Antunes	✓	✓		✓	✓				✓					
Madalena De Oliveira Barbosa Divino			✓			✓		✓	✓					
Luana dos Santos Cordeiro						✓		✓	✓					
Fernanda Pereira Leite Aguiar			✓	✓	✓			✓	✓	✓	✓			
Marcelo Tsuguio Okano	✓			✓			✓		✓	✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

This study did not involve human participants or animals.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, [MTO]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] A. Preeya V, S. P. Raja, B. K. Dhanalakshmi, H. Gururaj, V. Ravi, and P. Ravi, "Perceptual pigeon galvanized optimization of multi-objective CNN on the identification and classification of mango leaves disease," *The Open Agriculture Journal*, vol. 18, no. 1, Nov. 2024, doi: 10.2174/0118743315331264240917073431.

[2] D. B. Valdez, C. J. G. Aliac, and L. S. Feliscuzo, "Medicinal plant classification using convolutional neural network and transfer learning," in *4th IEEE International Conference on Artificial Intelligence in Engineering and Technology, IICAIET 2022*, Sep. 2022, pp. 1–6, doi: 10.1109/IICAIET55139.2022.9936868.

[3] S. Chaivivatrakul, J. Moonrinta, and S. Chaiwiwatrakul, "Convolutional neural networks for herb identification: plain background and natural environment," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 12, no. 3, pp. 1244–1252, Jun. 2022, doi: 10.18517/ijaseit.12.3.15348.

[4] R. K. Domingues, J. A. Cordioli, and D. Silva, "Development of a methodology for diagnosing faults in bearings operating under variable operating conditions based on self-supervised learning," *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM* , vol. 16, no. 1, 2024, doi: 10.36001/phmconf.2024.v16i1.4183.

[5] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOV4: optimal speed and accuracy of object detection,” *arXiv-Computer Science*, pp. 1–17, 2020.

[7] S. N. Antunes *et al.*, “Model development for identifying aromatic herbs using object detection algorithm,” *AgriEngineering*, vol. 6, no. 3, pp. 1924–1936, Jun. 2024, doi: 10.3390/agriengineering6030112.

[8] M. Tan and Q. V. Le, “EfficientNet: rethinking model scaling for convolutional neural networks,” in *36th International Conference on Machine Learning, ICML 2019*, 2019, pp. 1–11.

[9] D. Chenna, “Evolution of convolutional neural network (CNN): compute vs memory bandwidth for edge AI,” *Feedforward*, vol. 2, no. 3, pp. 2–12, 2023.

[10] S. Temsiririrkkul, P. Siritanawan, and R. Temsiririrkkul, “Which one is Kaphrao? identify Thai herbs with similar leaf structure using transfer learning of deep convolutional neural networks,” in *TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON),* Auckland, New Zealand, Dec. 2021, pp. 738–743, doi: 10.1109/TENCON54134.2021.9707324.

[11] J. Lu, L. Tan, and H. Jiang, “Review on convolutional neural network (CNN) applied to plant leaf disease classification,” *Agriculture*, vol. 11, no. 8, Jul. 2021, doi: 10.3390/agriculture11080707.

[12] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, “Review on convolutional neural networks (CNN) in vegetation remote sensing,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24–49, Mar. 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[13] M. Iman, H. R. Arabnia, and K. Rasheed, “A review of deep transfer learning and recent advancements,” *Technologies*, vol. 11, no. 2, 2023, doi: 10.3390/technologies11020040.

[14] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” in *Artificial Neural Networks and Machine Learning – ICANN 2018*, Cham: Springer, 2018, pp. 270–279, doi: 10.1007/978-3-030-01424-7_27.

[15] T. Alshalali and D. Josyula, “Fine-tuning of pre-trained deep learning models with extreme learning machine,” in *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018*, Dec. 2018, pp. 469–473, doi: 10.1109/CSCI46756.2018.00096.

[16] A. Paparella, B. Nawade, L. S.-Harpaz, and M. Ibdah, “A review of the botany, volatile composition, biochemical and molecular aspects, and traditional uses of Laurus nobilis,” *Plants*, vol. 11, no. 9, Apr. 2022, doi: 10.3390/plants11091209.

[17] C. Gonçalves, D. Fernandes, I. Silva, and V. Mateus, “Potential anti-inflammatory effect of rosmarinus officinalis in preclinical in vivo models of inflammation,” *Molecules*, vol. 27, no. 3, Jan. 2022, doi: 10.3390/molecules27030609.

[18] S. Saqib *et al.*, “Mentha: nutritional and health attributes to treat various ailments including cardiovascular diseases,” *Molecules*, vol. 27, no. 19, Oct. 2022, doi: 10.3390/molecules27196728.

[19] Z. Du, “Review of convolutional neural network,” *Science and Technology of Engineering, Chemistry and Environmental Protection*, vol. 1, no. 9, Oct. 2024, doi: 10.61173/1xfrez14.

[20] C. Post, R. Sarala, C. Gatrell, and J. E. Prescott, “Advancing theory with review articles,” *Journal of Management Studies*, vol. 57, no. 2, pp. 351–376, Mar. 2020, doi: 10.1111/joms.12549.

[21] L.-L. Ebidor and I. G. Ikhide, “Literature review in scientific research: an overview,” *East African Journal of Education Studies*, vol. 7, no. 2, pp. 179–186, 2024, doi: 10.37284/eajes.7.2.1909.

[22] E. S. Barry, J. Merkebu, and L. Varpio, “State-of-the-art literature review methodology: a six-step approach for knowledge synthesis,” *Perspectives on Medical Education*, vol. 11, no. 5, pp. 281–288, Sep. 2022, doi: 10.1007/s40037-022-00725-9.

[23] G. de B. V. Junior *et al.*, “Métricas utilizadas para avaliar a eficiência de classificadores em algoritmos inteligentes,” *Centro de Pesquisas Avançadas em Qualidade de Vida*, vol. 14, no. 2, pp. 1-14, 2022, doi: 10.36692/v14n2-01r.

[24] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning transferable architectures for scalable image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.

[25] K. Shivaprasad and A. Wadhawan, “Deep learning-based plant leaf disease detection,” *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS),* Madurai, India, pp. 360–365, 2023, doi: 10.1109/ICICCS56967.2023.10142857.

[26] M. M. Abdulrazzaq *et al.*, “Consequential advancements of self-supervised learning (SSL) in deep learning contexts,” *Mathematics*, vol. 12, no. 5, 2024, doi: 10.3390/math12050758.

[27] A. Zafar, J. Tanveer, M. U. Ali, and S. W. Lee, “BU-DLNet: breast ultrasonography-based cancer detection using deep-learning network selection and feature optimization,” *Bioengineering*, vol. 10, no. 7, pp. 1-16, Jul. 2023, doi: 10.3390/bioengineering10070825.

[28] S. M. A. Sharif, R. A. Naqvi, Z. Mehmood, J. Hussain, A. Ali, and S.-W. Lee, “MedDeblur: medical image deblurring with residual dense spatial-asymmetric attention,” *Mathematics*, vol. 11, no. 1, 2022, doi: 10.3390/math11010115.

[29] F. Zhuang *et al.*, “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021, doi: 10.1109/JPROC.2020.3004555.

[30] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.

[31] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 13001–13008, Apr. 2020, doi: 10.1609/aaai.v34i07.7000.

[32] N. Saleem *et al.*, “Class-specific data augmentation for plant stress classification,” *Plant Phenome Journal*, vol. 7, no. 1, 2024, doi: 10.1002/ppj2.20112.

[33] V. Sharma, K. Chaurasia, and A. Bansal, “Optimizing species recognition in medicinal plants: a comprehensive evaluation of deep learning models,” in *Proceedings of the 14th International Conference on Cloud Computing, Data Science and Engineering, Confluence 2024*, Jan. 2024, pp. 738–743, doi: 10.1109/Confluence60223.2024.10463198.

[34] E. Hassan, S. A. Ghazalah, N. El-Rashidy, T. A. El-Hafeez, and M. Y. Shams, “DenseNet model with attention mechanisms for robust date fruit image classification,” *International Journal of Computational Intelligence Systems*, vol. 18, no. 1, 2025, doi: 10.1007/s44196-025-00809-4.

[35] A. Howard *et al.*, “Searching for mobileNetV3,” in *Proceedings of the IEEE International Conference on Computer Vision*, Oct. 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[36] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258, doi: 10.1109/CVPR.2017. 195.

[37] M. U. Ali *et al.*, “Deep learning network selection and optimized information fusion for enhanced COVID-19 detection,” *International Journal of Imaging Systems and Technology*, vol. 34, no. 2, 2024, doi: 10.1002/ima.23001.

APPENDIX

Table 3. Evolution of training considering the transfer learning process

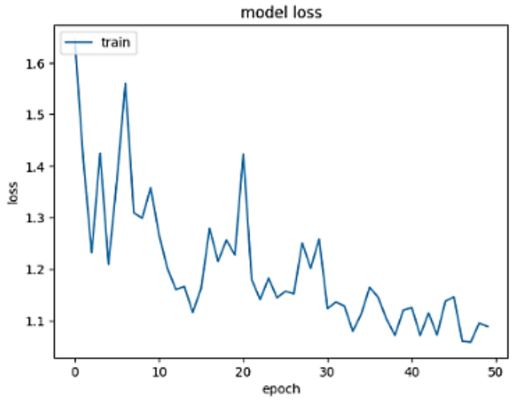
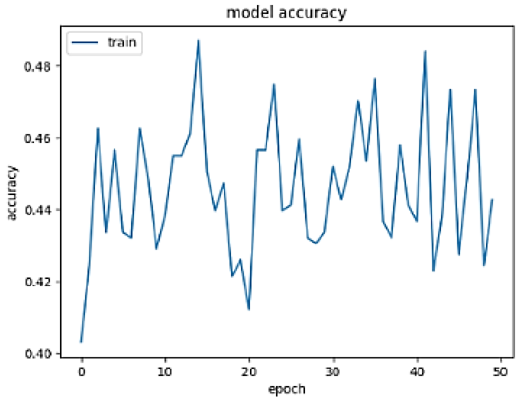
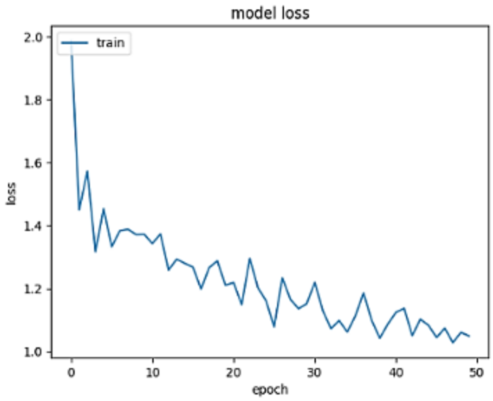
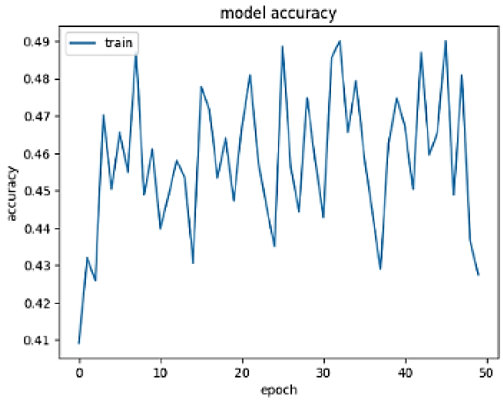
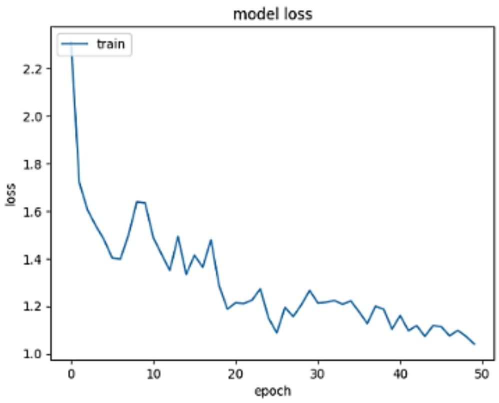
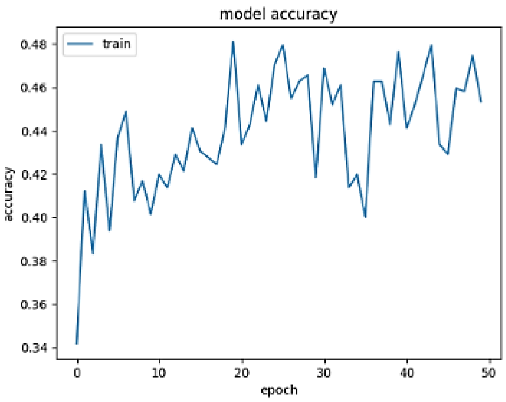
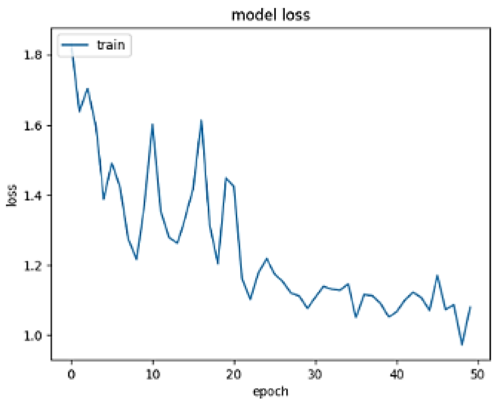
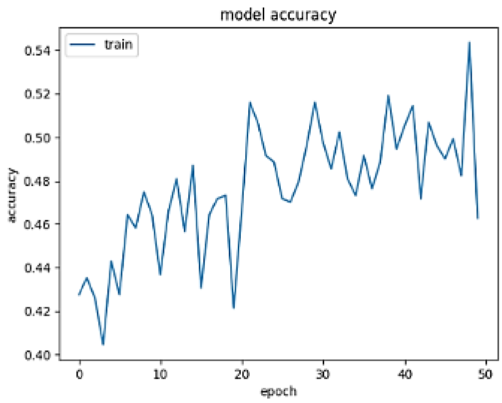
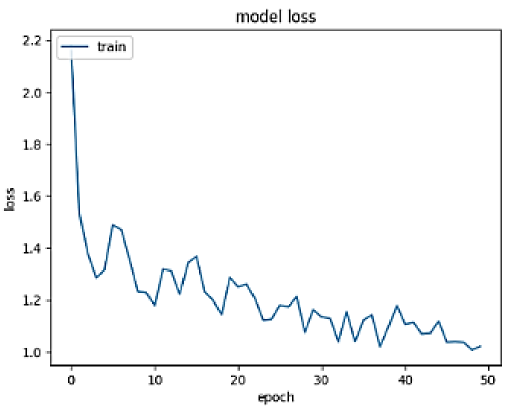
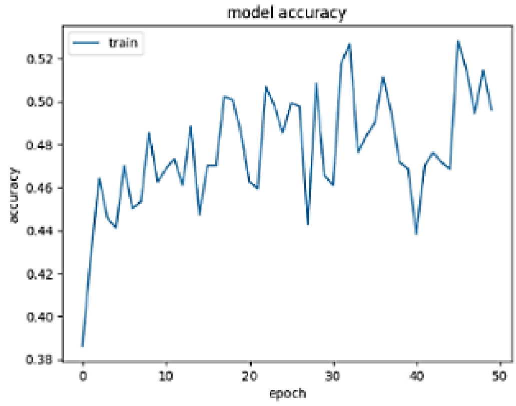
<table>
<tr><th>Resolution</th><th>Loss</th><th>Accuracy</th></tr>
<tr><td>B0</td><td>
</td><td>
</td></tr>
<tr><td>B1</td><td>
</td><td>
</td></tr>
<tr><td>B2</td><td>
</td><td>
</td></tr>
<tr><td>B3</td><td>
</td><td>
</td></tr>
</table>

Table 3. Evolution of training considering the transfer learning process (*continued*)

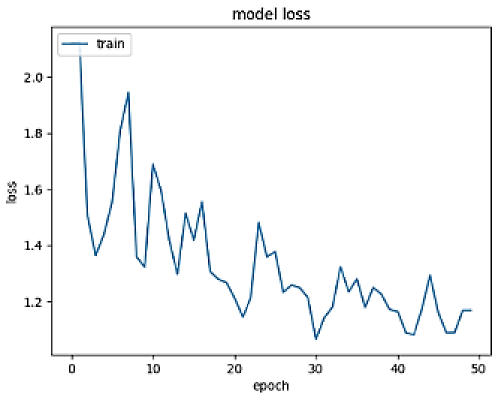
Resolution	Loss	Accuracy

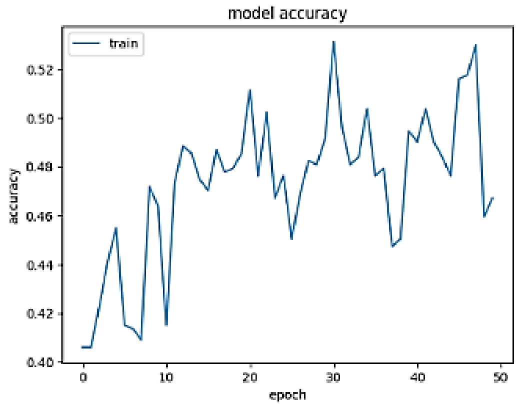
B4



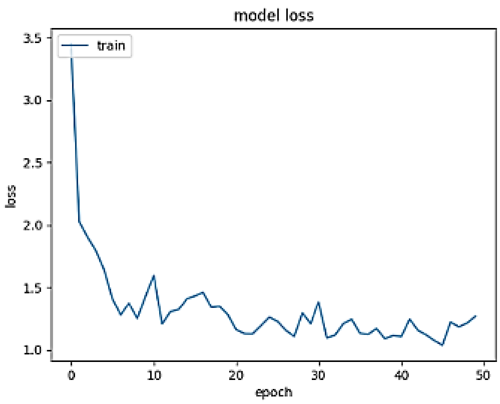


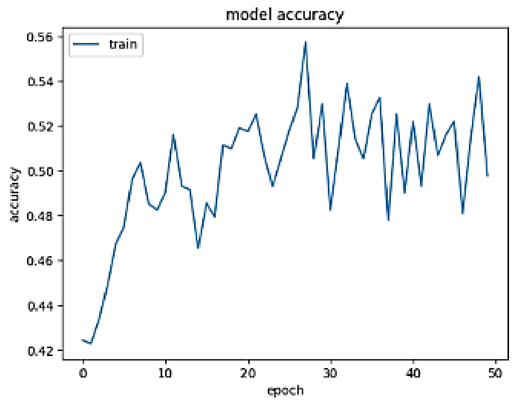
B5



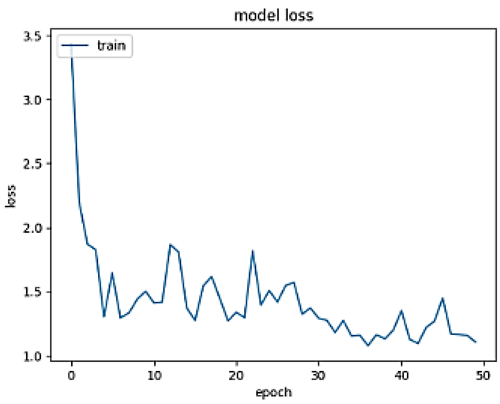


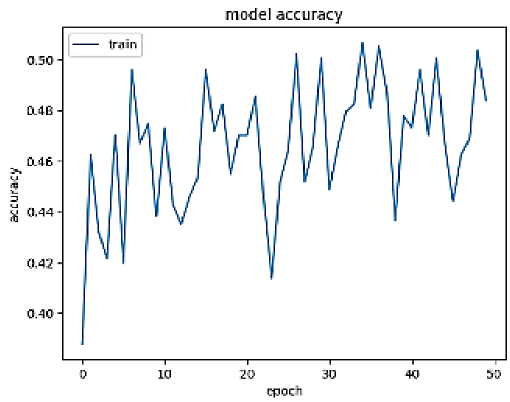
B6



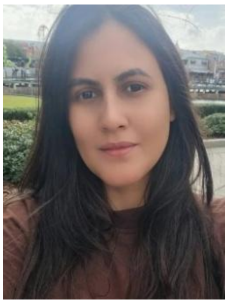


B7





BIOGRAPHIES OF AUTHORS



Samira Nascimento Antunes is Ph.D. candidate in Digital Technologies and Technological Innovation in Operation Systems. This project is part of the research project Management of Production and Innovation in Agriculture, Industry and Services with a focus on fruit and herb recognition through artificial intelligence algorithms (CNN, YOLO framework) and development of a web and mobile interface as provision of product. She can be contacted at email: samira.antunes@aluno.unip.br.

Madalena De Oliveira Barbosa Divino has a Bachelor's Degree in Systems Analysis and IT-FATEC Antônio Russo São Caetano do Sul (2012), a postgraduate degree in Teaching in Higher Education-SENAC (2015) and a degree in Pedagogical Training for Secondary Education from the State Center for Technological Education Paula Souza (2017). She currently works as a teacher-Etec Rio Grande da Serra in the technical area, since 2015. She also works in private education, as a teacher of the initial years and high school in the areas of technology, programming and robotics. She can be contacted at email: madalena.divino@etec.sp.gov.br.

Luana dos Santos Cordeiro is Pedagogy, Computer Technologist with emphasis in Business Management from FATEC-Mauá, Master's student in Production Engineering. Currently, she is a High School and Technical School teacher-ETEC Professora Maria Cristina Medeiros and at the São Paulo State Department of Education (SEDUC). She has experience in the area of Education, in different teaching modalities, working for over 13 years in the area of education permeating between early childhood education, elementary school 1 and 2, high school, technical and vocational education, in addition to inclusive education. Enthusiastic, curious and interested in the area of project development in different segments. She can be contacted at email: lsc.luana@hotmail.com.

Fernanda Pereira Leite Aguiar is Ph.D. Candidate in Production Engineering. This project seeks to develop an application, using artificial intelligence and computer vision to identify fear in beef cattle. Master in Productive Systems, at Centro Paula Souza (2023). Postgraduate MBA in Project Management from Centro Paula Souza (2018). Graduation in Information Technology Management from Faculdade de Tecnologia de Barueri (2016). Academic award for the undergraduate degree in Information Technology Management from Faculdade de Tecnologia de Barueri (2016). Participation in technology fairs (9th Technology Fair-Feteps) by the Barueri Faculty of Technology (2015). LIBRAS interpreter course by the InterLibras school (2019). Theology course by the Barueri Institute of Theology-ITEB (2019). She can be contacted at email: fernanda.p.aaguiar@gmail.com.

Prof. Dr. Marcelo Tsuguio Okano is master's in administration (2006) from USCS, Ph.D. (2010) in Production Engineering from Universidade Paulista (UNIP), and Ph.D. (2020) in Administration from the School of Business Administration of São Paulo (EAESP) of Fundação Getulio Vargas (FGV). He completed his Post-Doctorate (Production Engineering) in 2013 at FEG-UNESP and in 2015 he completed his Post-Doctorate (Production Engineering) in innovation and entrepreneurship at UNIP. Since 2022, he has been a professor in the Postgraduate Program in Production Engineering at UNIP, working in the concentration area of Operation Systems Management in projects related to the themes of information and communication technology, digital technologies, technological innovation, digital transformation, digital platforms, digital ecosystems and LGPD. Leader of the CNPq Research Group GEDITAD-Study and Development Group in Innovation, Technologies and Digital Artifacts at UNIP. Voluntary collaborating researcher at the Faculty of Technology of UNICAMP since 2020 working in the areas of IT and Telecommunications. Productivity Scholarship in Technological Development and Innovative Extension by CNPq level 2. He has several articles published in specialized journals and works presented in annals of scientific events, book chapters and published books, in addition to technological products and patented software. He can be contacted at email: marcelo.okano@unip.br.