Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France - August, 2025
Paper No. MVML 137
DOI: 10.11159/mvml25.137

Today's Technology for Today's Environment: Comparison of Convolutional Neural Networks for Invasive Plant Identification

Minghui Yao[1], Sophie Zeng[1]
[1]The Bishop's School
7607 La Jolla Blvd, La Jolla, USA
grace.yao.26@bishops.com; sophie.zeng.26@bishops.com

Abstract - Invasive plant species represent one of the most significant threats to biodiversity and ecological health worldwide, contributing to over 33% of documented animal extinctions and costing the North American economy more than $26 billion annually. In regions like San Diego, California, United States, where high biodiversity and wildfire susceptibility amplify these effects, timely and accurate identification of invasive species is essential for effective removal. However, manual identification requires expert knowledge, limiting large-scale volunteer efforts. This paper aims to develop novel artificial-intelligence-based solutions for invasive plant identification using computer vision. A custom convolutional neural network (CNN) was developed and trained to classify eleven high-priority invasive plant species in San Diego. Transfer learning models based on MobileNetV2 and ResNet50V2 were implemented and evaluated to compare performance, scalability, and suitability for field deployment. All models achieved high top-N accuracies, with the custom CNN and ResNet50V2 and MobileNetV2 transfer models reaching top-5 accuracies of over 94%. These results demonstrate the value of top-N accuracy in real-world ecological contexts, where presenting multiple likely classifications can effectively guide user decisions. This paper highlights the potential of AI-driven tools to reduce barriers to environmental action by democratizing species identification and enabling more consistent monitoring. By combining locally-collected ecological data with lightweight, field-deployable deep learning architectures, these novel models can support real-time, community-powered species identification. This approach contributes to a scalable and cost-effective framework to support biodiversity conservation and sustainable ecosystem management through accessible machine learning applications.

***Keywords*:** convolutional neural networks, transfer learning, top-N accuracy, invasive plants, invasive species

1. Introduction

Environmental protection has become a rising topic in modern science, technology, and social debates, but there are still many areas that have been left virtually untouched, such as the battle against invasive plants. Invasive species are currently the biggest threat to global biodiversity, second only to habitat loss. They are a factor in over 33% of animal extinctions and over 25% of plant extinctions [1]. Especially in dry areas like much of California, invasive plants accelerate erosion, create abnormal buildup on forest floors, and serve as fuel for more frequent and intense wildfires [2]. They also greatly harm the global economy, affecting property values, agricultural productivity, public utility operations, native fisheries, tourism, and outdoor recreation. Invasive plants and animals cost North America $2 billion per year in the early 1960s to over $26 billion per year since 2010, with their global economic cost estimated to be $1.288 trillion over the past 50 years [3]. However, a substantial workforce is required to reduce or eradicate dangerous invasive plant species, and to ensure that they do not reinvade a region, long-term and consistent check-ins on the affected area are necessary. This type of removal requires a high level of knowledge of all the plant species in the area, presenting a large time and effort barrier for volunteers and disincentivizing individual removal. Through breaking down this knowledge barrier, more proactive and widespread management could substantially reduce future costs at the trillion-dollar scale [4].

Thus, this paper presents a novel approach that leverages artificial intelligence (AI) to address the ecological and economic challenges posed by invasive plants. Artificial intelligence, particularly in the form of computer vision, has shown immense potential in environmental protection, from identifying endangered species, tracking deforestation, or monitoring ocean water quality [5]. By automating complex identification tasks, AI can bridge the gap between citizen scientists and the specialized knowledge typically required for effective conservation work. In this paper, we develop and implement a convolutional neural network (CNN) to recognize the most invasive plant species in San Diego County, California, United States, an area particularly vulnerable due to its high biodiversity and risk of wildfires. Furthermore, the custom CNN was compared to several leading transfer learning models, including MobileNetV2 and ResNet50V2 to evaluate the tradeoffs in accuracy and generalizability. Through this comparative analysis, we highlight the strengths of different machine learning models in achieving high performance in invasive plant identification as viable tools for integration into mobile applications for citizen scientists and local land management efforts.

2. Methodology

2.1. Data Collection

Government databases, including the San Diego County Invasive Weed Watch and the California Department of Food and Agriculture's *Encycloweedia*, were utilized to identify priority species in San Diego County for identification, focusing on legally mandated A-rated plants, defined as "a pest of known economic or environmental detriment and is not known to be established in California. A-rated pests are prohibited from entering the state… If found entering or established in the state, A-rated pests are subject to state (or commissioner when acting as a state agent) enforced action involving eradication, quarantine regulation, containment, rejection, or other holding action" [6]. As such, a list of eleven high-priority invasive plant species was determined:

1. *Aegilops triuncialis*
2. *Arundo donax*
3. *Centaurea calcitrapa*
4. *Centaurea stoebe*
5. *Cortaderia selloana*
6. *Cynara cardunculus*
7. *Euphorbia terracina*
8. *Hypericum canariense*
9. *Lepidium latifolium*
10. *Lythrum salicaria*
11. *Volutaria tubuliflora*

Image data were manually downloaded from the online Calflora and the Bugwood Image Databases, ensuring variation in lighting, focus, plant parts, and growth stages, with about 150 images per plant in the final dataset.

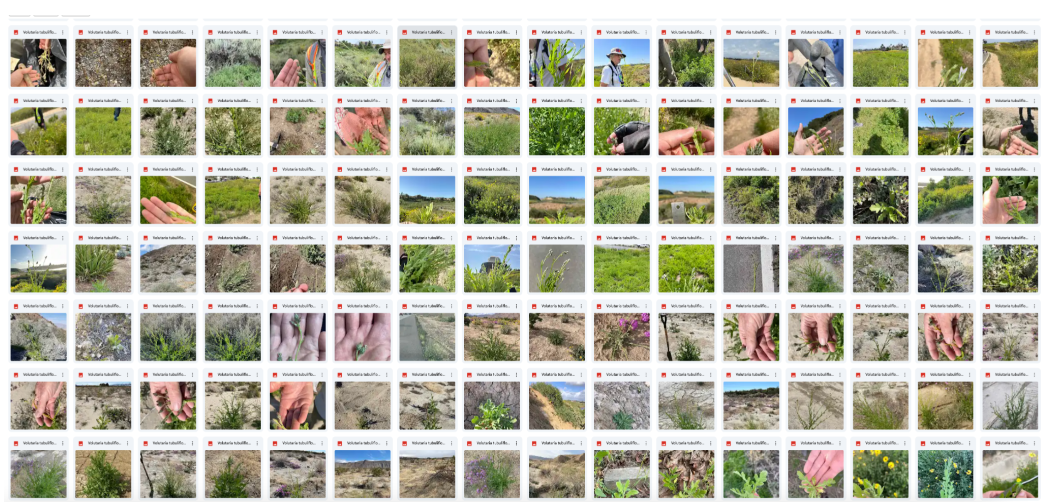

Fig. 1: Example of collected dataset for *Volutaria tubuliflora* (imported to Google Drive)

2.2. Convolutional Neural Network

Convolutional neural networks (CNNs) are a class of deep learning models particularly effective for image classification tasks due to their ability to automatically learn spatial hierarchies of features through convolutional layers. For this study, a custom CNN architecture was implemented using the Keras deep learning framework to classify the 11 high-priority invasive plant species determined above. The model was built using a sequential approach and consisted of six convolutional layers, each followed by ReLU activations, synchronized batch normalization, and max pooling layers to progressively reduce spatial dimensions. Dropout layers were interspersed to mitigate overfitting by randomly deactivating neurons during training. The extracted feature maps were flattened into a one-dimensional vector and passed through a dense layer of 512 neurons before producing final predictions through a softmax-activated dense layer of 11 outputs, corresponding to the target plant species. L1 regularization was applied to the dense layers to further prevent overfitting. The model was compiled with the RMSprop optimizer, using categorical cross-entropy as the loss function and accuracy as the evaluation metric.

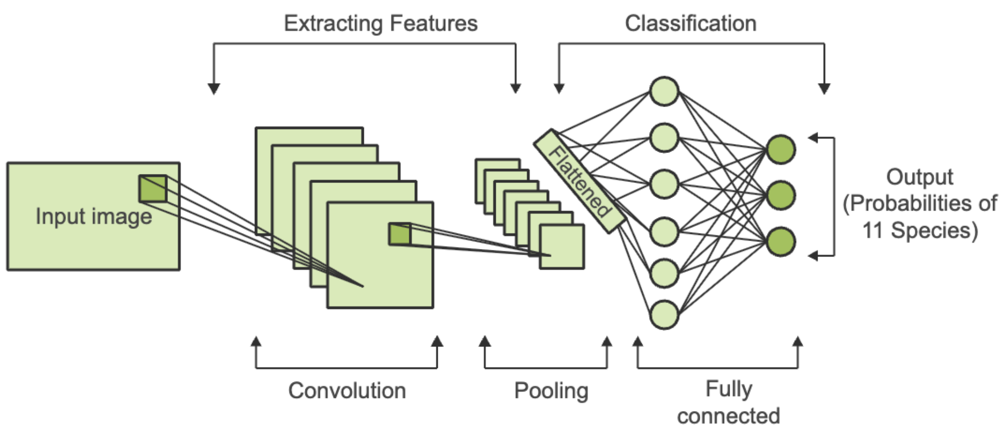

Fig. 2: Flowchart of our CNN architecture

2.3. Transfer Learning Models

Two additional transfer learning approaches were implemented using pre-trained convolutional neural networks. The first model employed ResNet50V2, leveraging its deep residual architecture to extract high-level visual features from input images resized to 224×224 [7]. The pretrained ResNet50V2 base was frozen to preserve ImageNet-learned weights and followed by a Flatten layer and a dense softmax classifier tailored to the invasive plant categories. The model was compiled using categorical cross entropy loss and the Adam optimizer, and training performance was tracked using accuracy metrics and checkpoint callbacks. A second model incorporated MobileNetV2, a lightweight and efficient architecture ideal for deployment on resource-constrained devices [8]. Input images were resized to 300×300 and normalized to a [0,1] scale. The MobileNetV2 base was frozen and followed by a global average pooling layer, a dense ReLU-activated layer with L2 regularization, batch normalization, dropout, and a final softmax output layer. This model was trained with a learning rate scheduler and checkpointing based on validation accuracy. Both transfer learning models demonstrated effective feature reuse from large-scale datasets, significantly reducing training time while maintaining classification performance on this specific set of plant species.

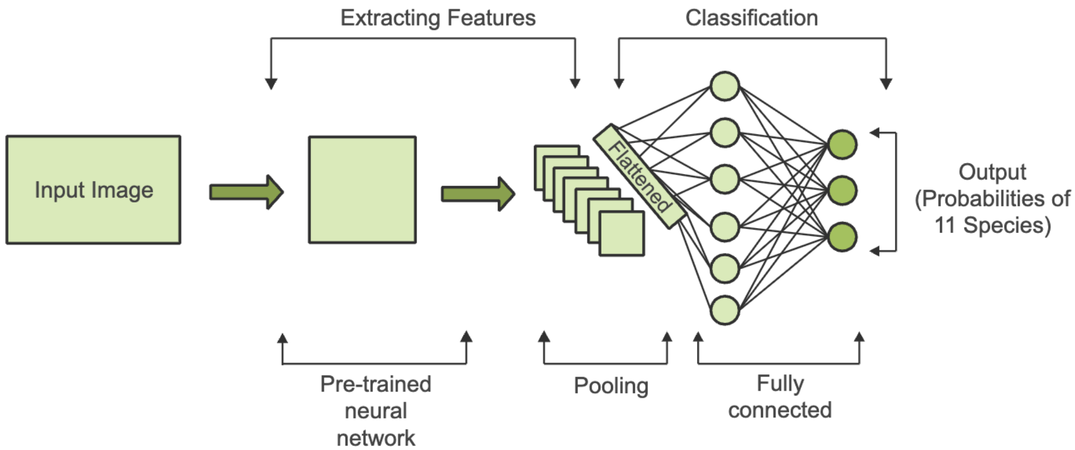

Fig. 3: Flowchart of our transfer model architecture

3. Results & Discussion

3.1. Convolutional Neural Network

The best-performing CNN's performance across epochs was visualized by plotting accuracy for the training and validation datasets.

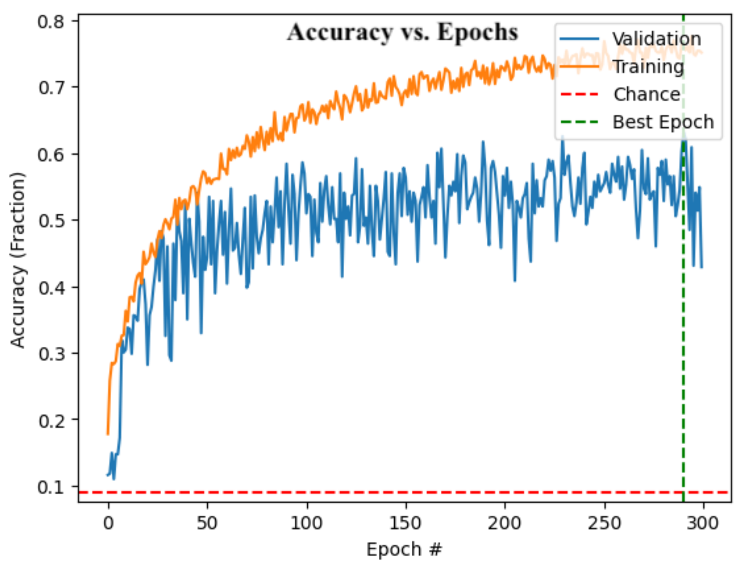

Fig. 4: Accuracy plot of best-performing CNN

To evaluate CNN performance beyond standard accuracy metrics, top-N accuracy and confusion matrix analyses were conducted. Top-N accuracy assessed the model's ability to include the correct class within its top N predicted probabilities, offering a more useful metric for multi-class classification. This accuracy metric is valuable in real-world applications of these models as it reflects scenarios where presenting multiple likely predictions of harmful invasive plants for removal can still guide effective decision-making even if the top prediction is incorrect.

Table 1: Top 1/2/3/4/5 accuracy of best-performing CNN

Top 1 Accuracy	Top 2 Accuracy	Top 3 Accuracy	Top 4 Accuracy	Top 5 Accuracy
0.6605	0.7971	0.8634	0.9275	**0.9440**

Additionally, confusion matrices were generated for the best-performing CNN iteration. Confusion matrices provide a detailed breakdown of a classification model's performance by showing not only overall accuracy but also how individual classes are predicted, helping to identify specific types of errors, such as which plant classes are commonly confused with others, enabling targeted improvements. In real-world applications, especially with multiple categories like invasive plant species, this detailed insight is crucial for understanding model strengths and weaknesses beyond simple accuracy metrics.

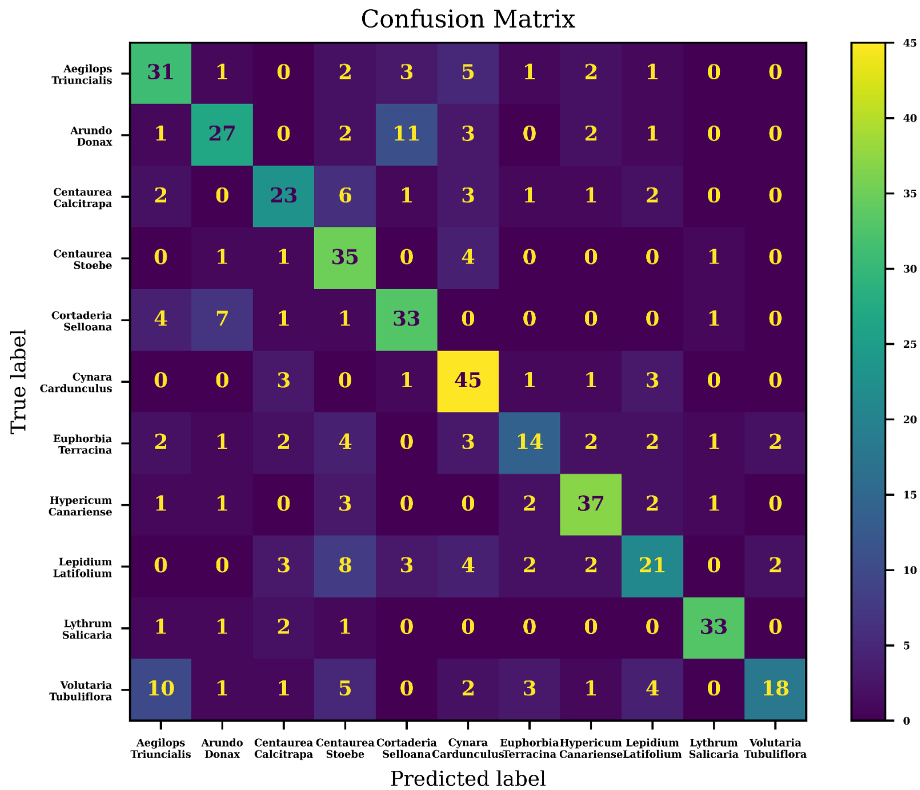

Fig. 5: Confusion matrix of best-performing CNN

The highest-performing CNN reached a top-5 accuracy of 94.40% (Table 1) and generally predicted labels correctly as seen in the diagonal in the confusion matrix (Fig. 5).

3.2. Transfer Learning Models

To assess the effectiveness of transfer learning approaches, two pre-trained convolutional neural networks, ResNet50V2 and MobileNetV2, were fine-tuned on the collected San Diegan invasive plant species dataset. The ResNet50V2 model was evaluated using standard accuracy and loss, in addition to the top-N accuracy metric specified in Section 3.1.

Table 2: Top 1/2/3/4/5 accuracy of ResNet50V2 Model

Top 1 Accuracy	Top 2 Accuracy	Top 3 Accuracy	Top 4 Accuracy	Top 5 Accuracy
0.631	0.7876	0.8743	0.9097	**0.9451**

The ResNet50V2 model demonstrated top-N accuracies useful for field removal. MobileNetV2 outperformed both ResNet50V2 and the custom CNN in terms of top-1 classification precision on the test dataset. MobileNetV2's lighter architecture makes it an effective choice for deployment in field or mobile environments where computational resources may be limited. It demonstrated a similarly useful top-5 accuracy to the other models.

Table 3: Top 1/2/3/4/5 accuracy of MobileNetV2 Model

Top 1 Accuracy	Top 2 Accuracy	Top 3 Accuracy	Top 4 Accuracy	Top 5 Accuracy
0.69381	0.8301	0.8850	0.9257	**0.9416**

All models evaluated, including ResNet50V2, MobileNetV2, and the custom CNN, demonstrated strong top-5 accuracy, demonstrating their utility in real-world invasive species identification and removal efforts. MobileNetV2 may be the most effective model overall due to its balance of precision and lightweight architecture suited for mobile deployment; however, both ResNet50V2 and the custom CNN still showed competitive performances. These results suggest that top-N accuracy is a valuable metric for ecological field applications, where suggesting multiple likely classifications is useful when deciding whether or not to remove a certain plant, especially when the baseline user in this application does not have experience recognizing invasive plant species.

4. Conclusion

Invasive plant species pose a critical threat to global biodiversity and ecosystems, contributing to over a third of documented extinctions and causing extensive environmental degradation, especially in vulnerable regions like California. Beyond ecological damage, they impose staggering economic costs, highlighting the urgent need for scalable, intelligent solutions to aid in their detection and removal. This study presents an AI-driven approach to support ecological conservation through the automated identification of invasive plant species in San Diego County. By developing a custom CNN and evaluating it alongside two advanced transfer learning models, ResNet50V2 and MobileNetV2, we demonstrated the feasibility and effectiveness of computer vision in real-world environmental applications. All models achieved strong top-5 accuracies of about 94%, indicating their robustness in guiding effective decision-making for removal in the field, even when the top prediction is incorrect. Among the models, MobileNetV2 stood out as the most suitable for field deployment, achieving the highest top-1 accuracy while maintaining a lightweight architecture compatible with mobile and low-resource settings, but the custom CNN and ResNet50V2 also performed competitively.

This paper demonstrates the potential of AI-powered tools to democratize access to specialized environmental knowledge, lower the barrier for individual action, and support more proactive and cost-effective invasive species management. By integrating these models into mobile applications for citizen scientists, conservation volunteers, or land management agencies, this technology can enable more widespread and consistent monitoring, reducing long-term ecological and economic damage. Future work may focus on expanding species coverage, improving generalization across regions, and integrating real-time geolocation and feedback systems to create a comprehensive, intelligent framework for invasive plant detection and removal.

Acknowledgements

We sincerely thank all of the professionals that we interviewed from the San Diego River Park Foundation (SDRPF), the California Invasive Plant Council (Cal-IPC), and San Diego Park and Recreation for their valuable input and feedback: Mr. Kristofer Gonzalez, Ms. Chandra Kinney, Ms. Nicole Valentine, Ms. Constance Taylor, and Ms. Melanie Fontana. We also extend our gratitude to Mrs. Ritsema, The Bishop's School environmental science teacher, for connecting us with San Diego Park rangers. We would also like to warmly thank Dr. Marcus Jaiclin of The Bishop's School for his continued support and mentorship throughout our research, programming, and redesigning process.

References

[1] O. Rosane, "Invasive Species Have Led to a Third of Animal Extinctions Since 1500," *EcoWatch*, Mar. 04, 2019. http://www.ecowatch.com/invasive-species-animal-extinctions-2630614032.html (accessed May 24, 2025).

[2] E. Arnett, "This Invasive Species Is Fueling Western Wildfires | Theodore Roosevelt Conservation Partnership," *Theodore Roosevelt Conservation Partnership*, Aug. 28, 2020. http://www.trcp.org/2020/08/28/invasive-species-fueling-western-wildfires/ (accessed May 24, 2025).

[3] "Economic and Social Impacts | National Invasive Species Information Center," *Invasivespeciesinfo.gov*, 2021. http://www.invasivespeciesinfo.gov/subject/economic-and-social-impacts (accessed May 24, 2025).

[4] R. N. Cuthbert, C. Diagne, E. J. Hudgins, A. Turbelin, D. A. Ahmed, C. Albert, T. W. Bodey, E. Briski, F. Essl, P. J. Haubrock, R. E. Gozlan, N. Kirichenko, M. Kourantidou, A. M. Kramer, F. Courchamp, "Biological invasion costs reveal insufficient proactive management worldwide," *Science of The Total Environment*, vol. 819, p. 153404, May 2022, doi: https://doi.org/10.1016/j.scitotenv.2022.153404.

[5] D. B. Olawade, O. Z. Wada, A. O. Ige, B. I. Egbewole, A. Olojo, and B. I. Oladapo, "Artificial Intelligence in Environmental Monitoring: Advancements, Challenges, and Future Directions," *Hygiene and Environmental Health Advances*, vol. 12, pp. 100114–100114, Oct. 2024, doi: https://doi.org/10.1016/j.heha.2024.100114.

[6] "CDFA - Plant Health - Integrated Pest Control - Encycloweedia: Weed Ratings," *Ca.gov*, 2025. http://www.cdfa.ca.gov/plant/ipc/encycloweedia/winfo_weedratings.html (accessed May 24, 2025).

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv*, Dec. 2015, doi: https://doi.org/10.48550/arxiv.1512.03385.

[8] M. Sandler, A. W. Howard, M. Zhu, Andrey Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *arXiv (Cornell University)*, Jan. 2018, doi: https://doi.org/10.48550/arxiv.1801.04381.