An Intelligent Machine Learning Framework for Automated Identification of Orchid Species Using Visual Feature Fusion

Dr Govinda Rajan[1], Dr Talha Habeeb[2], Ms Sadia Kouser[2], Ms. Hoorunnisa[2]
[1]Professor, [2]Assistant Professor,[1,2]Department of Computer Science and Engineering (AIML)
[1,2]Lords Institute of Engineering and Technology (UGC-Autonomous), Himayath sagar, Hyderabad, 500091, Telangana, India.

To Cite this Article
Dr Govinda Rajan, Dr Talha Habeeb, Ms Sadia Kouser, Ms. Hoorunnisa, "An Intelligent Machine Learning Framework for Automated Identification of Orchid Species Using Visual Feature Fusion", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 01, January 2026, pp: 181-189, DOI: http://doi.org/10.64771/jsetms.2026.v03.i01.pp181-189
Submitted: 24-11-2025 *Accepted: 31-12-2025* *Published: 07-01-2025*

ABSTRACT
Image classification has become an important area of research in computer vision, with applications in agriculture, environmental monitoring, and plant species identification. Traditional methods of identifying orchid species rely on manual observation and expert knowledge, which can be time-consuming and may lead to inconsistent results, particularly when distinguishing between visually similar flower varieties. Existing classification approaches often face challenges related to insufficient feature representation, classification errors, and high computational requirements. Therefore, an efficient automated approach is required to identify orchid species by capturing their distinctive visual characteristics. To address these challenges, this study proposes an intelligent machine learning framework for automated identification of orchid species using visual feature fusion. The proposed system utilizes handcrafted feature extraction techniques, including Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Hue, Saturation, Value (HSV) histograms, to extract complementary information related to shape, texture, and color characteristics from orchid images. The extracted features are combined into a unified feature vector to provide a comprehensive visual representation of each image. The resulting feature vectors are processed using multiple machine learning classifiers, including Logistic Regression (LR), Support Vector Classifier (SVC), Quadratic Discriminant Analysis (QDA), and Stochastic Gradient Descent (SGD), to identify the corresponding orchid species. The framework also incorporates performance evaluation metrics, confusion matrices, and visualization techniques to analyse and compare classifier performance. Experimental results demonstrate that the SGD classifier achieves the highest classification accuracy of 94.29%, outperforming the other evaluated models. The proposed framework provides an effective and computationally efficient solution for automated orchid species identification using handcrafted visual feature fusion and machine learning.

Keywords: Orchid Species Identification, Machine Learning, Visual Feature Fusion, HOG, LBP, HSV, Image Classification, Computer Vision.

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
CC BY-NC-ND 4.0

1. INTRODUCTION

Chinese Cymbidium orchids are a collective term for several terrestrial orchid species native to China, which, according to plant taxonomy, belong to the genus Cymbidium within the family Orchidaceae [1]. Due to Confucius' deep fondness for orchids and his unparalleled influence, along with the numerous poems related to Chinese Cymbidium orchids that have sprung up throughout history, as well as the importance of these flowers in terms of their beautiful appearance, diverse

applications and medicinal properties, these factors collectively contribute to the Chinese "orchid culture" being a source of immense pride and cultural identity for many people in China [2] as displayed in below Fig. 1.

In this historical context, Chinese Cymbidium orchids have been artificially cultivated for centuries due to their exceptional ornamental and cultural values, resulting in the development of numerous cultivars [3]. Concurrently, this cultivation has led to a flourishing orchid trade. The prices of different cultivars exhibit substantial variation, with commonplace cultivars generally priced below US 1,000 dollars, whereas rare cultivars or natural mutations cultivars can fetch tens or even hundreds of thousands of dollars in the market [4]. However, the increasing popularity and commercialization of Chinese Cymbidium orchids have raised various issues that pose significant challenges to their long-term survival and conservation. On the one hand, long-term inadequate cultivation management and outdated technical methods, Chinese Cymbidium orchids are facing problems including mixed varieties and difficulties in cultivar identification [5].



Fig. 1: Sample Orchid flower types.

2. LITERATURE SURVEY

Wei, et al. [6] provided a comprehensive bibliometric analysis of orchid-related literature published between 1902 and 2024, based on data retrieved from the Web of Science Core Collection™ (WoS). Results: The primary goal is to assess the global research landscape of orchids by identifying key authors, institutions, and journals, as well as major research themes in the field. A thorough analysis of publication trends, citation frequencies, and keyword co-occurrence networks was conducted to uncover significant research hotspots. Łobas, et al. [7] most taxonomically challenging species complexes within the genus *Epipactis*. Because of the exceptionally high levels of morphological variability and the ability to readily cross with other species, ninety different taxa at various taxonomic ranks have already been described within its nominative subspecies, but the taxonomic status of most

of them is uncertain, widely disputed, and sometimes even irrelevant. The present review is based on results of the most recent research devoted to the *E. helleborine* group taxonomy.

Tsaballa, et al. [8] assembled using Bioedit software, and the BLAST algorithm was used on the NCBI database for species identification at the genus level. Molecular barcoding data based on genetic similarity identification was in full coherence with taxonomic classification based on morphological data. The combination of *ITS* and *matK* exhibited a greater capacity to identify a species among the Greek salep samples. Out of the 53 samples examined, 52.9% were classified as *Dactylorhiza* spp. and 33.3% as *Anacamptis* spp., whereas only 6 samples were identified as *Orchis* spp. (11.8%). Tiwari, et al. [9] facilitated the commercialization of novel varieties, opening new avenues in the orchid industry, and their global marketing as cut flowers and artificially propagated plants. Orchids constitute the first floriculture crops that revolutionized the orchid industry; however, they also hold several challenges in the natural propagation and conservation of several species that are on the verge of extinction. International organizations like CITES have come forward to address challenges associated with illegal global trade and indiscriminate use of orchid varieties, aiming for conservation and legal commercial goals.

Litvak, et al. [10] focused on a specific task of urban plants classification. The possible practical application of this work is a tool which assists people, growing plants at home, to recognized new species and to provide the relevant caring instructions. Because urban species are barely covered by the benchmark datasets, these species cannot be accurately recognized by the state-of-the-art pre-trained classification models. This paper introduces a new dataset, Urban Planter, for plant species classification with 1500 images categorized into 15 categories. Das, et al. [11] predicted models of *Dendrobium nobile*, were developed through machine learning (ML) techniques including multiple linear regression (MLR), support vector regression (SVR), gradient boosting regression (GBR), and artificial neural networks (ANNs). The best model was tested using the coefficient of determination (R^2), mean absolute errors (MAEs), and root mean square errors (RMSEs) and statistically confirmed through average rank (AR).

Batchuluun, et al. [12] conducted on image augmentation, segmentation, detection, and classification based on plant images. Specifically, previous studies on plant image classification have used various plant datasets (fruits, vegetables, flowers, trees, etc., and their leaves). However, existing plant-based image datasets are generally small. Furthermore, there are limitations in the construction of large-scale datasets. Consequently, previous research on plant classification using small training datasets encountered difficulties in achieving high accuracy. Puttemans et al. [13] investigated the possibility of using object categorization and object classification techniques in an industrial context with a very limited set of training data. As an industrial application of the proposed techniques, they investigated the case of orchid flower detection and orchid species classification in an orchid packaging plant. Due to their large variety of colors and patterns, these orchid flowers are very hard to detect with classic segmentation-based techniques but form an ideal test case for object categorization techniques.

Sarachai, et al. [14] proposed an adaptive classification model of the orchid images by using a Deep Convolutional Neural Network (D-CNN). The first part of the model improved the quality of input feature maps using an adaptive Spatial Transformer Network (STN) module by performing a spatial transformation to warp an input image which was split into different locations and scales. We applied D-CNN to extract the image features from the previous step and warp into four branches. Serrano Pereira, et al. [15] presented a generic, hierarchical identification system for automated taxonomic approximation of organisms from images. We assessed the effectiveness of this system using photographs of slipper orchids (Cypripedioideae), for which we implemented image pre-processing, segmentation, and colour and shape feature extraction algorithms to obtain digital phenotypes for 116 species. The identification system trained on these digital phenotypes uses a nested hierarchy of artificial neural networks for pattern recognition and automated classification that mirrors the Linnean

taxonomy, such that user-submitted photos can be assigned a genus, section, and species classification by traversing this hierarchy.

3. PROPOSED METHODOLOGY

The proposed methodology presents a structured analytical framework for classifying flower images using traditional computer vision techniques and machine learning algorithms. The analytical pipeline begins with image dataset organization, followed by preprocessing and handcrafted feature extraction. Instead of deep learning-based feature extraction, the study relies on combining multiple complementary feature descriptors to capture visual characteristics such as shape, texture, and color. These extracted feature vectors are then analysed using multiple ML classifiers to perform category prediction. A graphical interface is incorporated to allow user interaction for dataset visualization, model training, evaluation, and prediction tasks, as shown in fig. 2. A local storage mechanism is used to manage extracted features and trained models, ensuring efficient reuse and reduced computation time. Continuous evaluation and comparison of models enable the system to identify the most suitable classifier for accurate image classification.

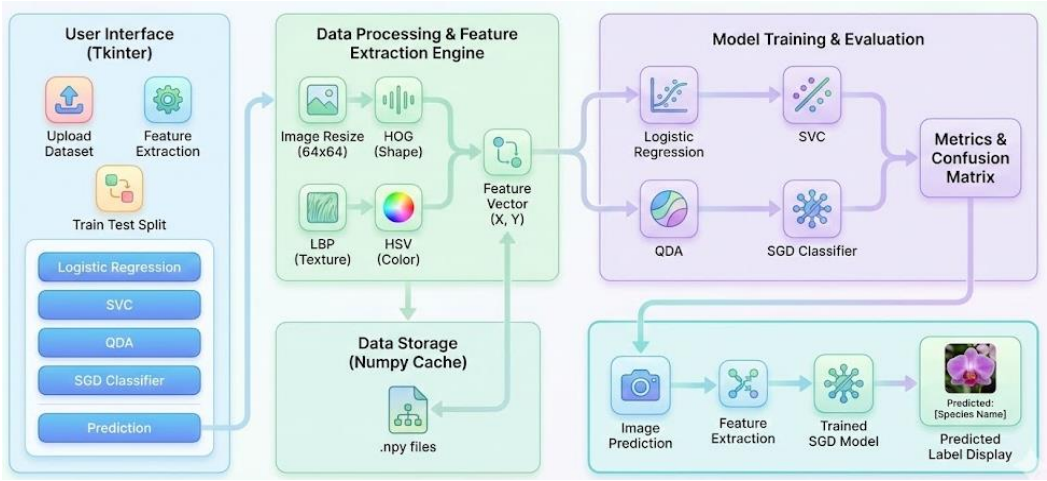

Fig. 2: System overview of orchid species classification workflow.

3.1 SGD Classifier

The SGD Classifier is an optimization-based learning algorithm suitable for large-scale and high-dimensional datasets. It updates model parameters iteratively using small batches of data, making it computationally efficient. In this study, SGD is employed to learn linear decision boundaries for orchid classification while handling high-dimensional flattened image features. Its fast convergence and scalability make it effective for real-time prediction scenarios. SGD initializes model weights randomly and updates them iteratively by minimizing a loss function using gradient descent. At each iteration, a small subset of training data is used to update weights, enabling faster learning. Final predictions are based on optimized weight vectors as shown in Fig. 3.

Internal Operations

- Initializes model weight parameters randomly.
- Processes image feature vectors in an iterative manner.
- Computes prediction error using a predefined loss function.
- Calculates gradients of the loss with respect to model weights.
- Updates weights incrementally using gradient descent optimization.
- Repeats updates until convergence or stopping criteria are met.
- Produces optimized decision boundaries for classification.
- Predicts the orchid class for unseen feature vectors.

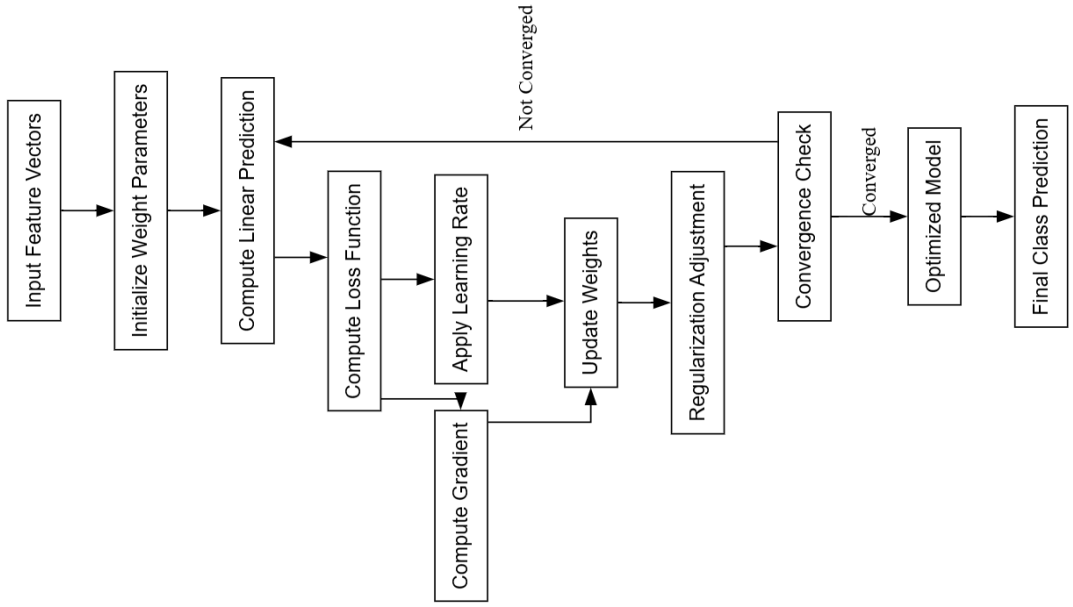

Fig. 3: Workflow of SGD Classifier.

4. RESULTS AND DISCUSSIONS

The results of this study demonstrate the effectiveness of combining handcrafted feature extraction techniques with multiple machine learning classifiers for accurate image classification. The performance of the system is evaluated using standard metrics such as accuracy, precision, recall, and F1-score, providing a comprehensive assessment of each model. A comparative analysis is conducted among LR, SVC, QDA, and SGD to identify the most suitable classifier for the given dataset. Visualization tools such as confusion matrices and performance graphs are used to interpret classification outcomes and error patterns.

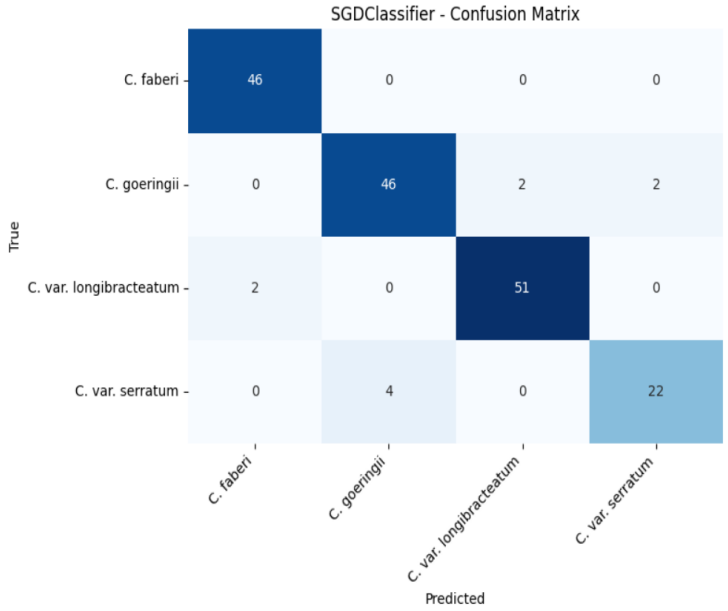

Fig. 4: SGD Confusion Matrix.

The confusion matrix presented in Fig. 4 demonstrates the performance of the SGD model. Most values lie along the diagonal, indicating high classification accuracy. C. faberi and C. var. longibracteatum are classified almost perfectly. Only minor misclassifications are observed in C. goeringii and C. var. serratum. The model efficiently handles high-dimensional feature vectors derived from HOG, LBP, and HSV. Its stochastic optimization enables faster and more effective learning. The results indicate strong generalization capability.



Fig. 5: Prediction Output – C. faberi

In Fig. 5 shows the prediction result where the system correctly identifies the orchid species as C. faberi. The model captures distinctive green petal features and structural patterns. The predicted label is displayed clearly on the image for easy interpretation. This confirms the effectiveness of feature extraction techniques. The classification demonstrates strong model confidence. The system successfully generalizes to unseen data. Visual output enhances user understanding of results. Overall, the prediction is accurate and reliable.



Fig. 6: Prediction Output C. goeringii

The prediction results in Fig. 6 shows that the system correctly classifies the image as C. goeringii. The model identifies unique color variations and petal structures of the species. Feature extraction plays a key role in distinguishing subtle differences. The predicted label is overlaid on the image. This confirms the model's ability to differentiate between similar classes. The output demonstrates robustness in classification. The system maintains consistent prediction performance.



Fig. 7: Prediction Output – C. var. longibracteatum

Fig. 7 presents the prediction result for *C. var. longibracteatum*. The model successfully recognizes the clustered floral structure and color distribution. The extracted features effectively capture both texture and shape details. The prediction label is clearly visible on the image. This indicates accurate classification by the trained model. The system handles complex visual patterns efficiently. The output confirms the reliability of feature fusion.



Fig. 8: Prediction Output – C. var. serratum.

The prediction output in Fig. 8 correctly identifies the orchid species as C. var. serratum. The model detects distinct orange-yellow color patterns and petal shapes. Feature extraction captures strong color characteristics aiding classification. The predicted label is displayed prominently on the image. This demonstrates the effectiveness of the SGD model. The system performs well even on visually distinct samples. The output is clear and interpretable.

Table 1: Performance Comparison of Machine Learning Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
LR	93.14	92.09	92.21	92.13
SVC	92.00	91.11	91.11	91.02
QDA	87.43	91.91	87.38	88.27
SGD	94.29	93.93	93.21	93.52

Table 1 presents the overall performance comparison of all machine learning models used in the classification system. The evaluation is based on key metrics including accuracy, precision, recall, and F1-score, which collectively measure the effectiveness of each model. Among all models, SGD achieves the highest accuracy of 94.29%, indicating superior performance in handling high-dimensional feature vectors. LR follows closely with 93.14% accuracy, demonstrating stable and reliable classification. SVC shows competitive performance but slightly lower than LR and SGD due to minor misclassifications. QDA records the lowest accuracy of 87.43%, highlighting its limitations with feature distribution assumptions. The table clearly shows that SGD provides the best balance between precision, recall, and F1-score.

5. CONCLUSION

This study presents an effective approach for image classification by combining handcrafted feature extraction techniques with multiple machine learning models. The use of HOG, LBP, and HSV features provides a rich representation of shape, texture, and color, which improves the ability to distinguish between different image categories. The comparative analysis of LR, SVC, QDA, and SGD highlights that each model performs differently depending on the nature of the extracted features. Performance evaluation using accuracy, precision, recall, and F1-score demonstrates that the system achieves consistent and reliable results. Confusion matrix analysis further helps in understanding classification errors and model behaviour across different classes. The system can be further enhanced by integrating

deep learning models such as CNN for improved feature learning and accuracy. Additionally, expanding the dataset with more diverse orchid species and real-world variations can improve model robustness and generalization.

REFERENCES

[1] Peng, Y., Zhou, Y., Zhang, L. et al. Orchid2024: A cultivar-level dataset and methodology for fine-grained classification of Chinese Cymbidium Orchids. Plant Methods **20**, 124 (2024). https://doi.org/10.1186/s13007-024-01252-w

[2] Ning H, Ao S, Fan Y, Fu J, Xu C. Correlation analysis between the karyotypes and phenotypic traits of Chinese cymbidium Orchid cultivars. Hortic Environ Biotechnol. 2018;59:93–103.

[3] Su S, Shao X, Zhu C, Xu J, Tang Y, Luo D, et al. An AGAMOUS-like factor is associated with the origin of two domesticated varieties in Cymbidium sinense (Orchidaceae). Hortic Res. 2018;5:48.

[4] Zhu G, Yang F, Shi S, Li D, Wang Z, Liu H, et al. Transcriptome characterization of Cymbidium sinense Orchid Dharma using 454 pyrosequencing and its application in the identification of genes Associated with Leaf Color Variation. PLoS ONE. 2015;10:e0128592.

[5] Seyler BC, Gaoue OG, Tang Y, Duffy DC, Aba E. Collapse of orchid populations altered traditional knowledge and cultural valuation in Sichuan, China. Anthropocene. 2020;29:100236.

[6] Wei, Y.; Li, J.; Jin, J.; Gao, J.; Xie, Q.; Lu, C.; Zhu, G.; Yang, F. Centenary Progress on Orchid Research: A Bibliometric Analysis. Genes **2025**, 16, 336. https://doi.org/10.3390/genes16030336

[7] Łobas, Z.; Khapugin, A.; Żołubak, E.; Jakubska-Busse, A. The Epipactis helleborine Group Orchid: An Overview of Recent Taxonomic Changes, with an Updated List of Currently Accepted Taxa. Plants **2021**, 10, 1839. https://doi.org/10.3390/plants10091839

[8] Tsaballa, A.; Kelesidis, G.; Krigas, N.; Sarropoulou, V.; Bagatzounis, P.; Grigoriadou, K. Taxonomic Identification and Molecular DNA Barcoding of Collected Wild-Growing Orchids Used Traditionally for Salep Production. Plants **2023**, 12, 3038. https://doi.org/10.3390/plants12173038

[9] Tiwari, P.; Sharma, A.; Bose, S.K.; Park, K.-I. Advances in Orchid Biology: Biotechnological Achievements, Translational Success, and Commercial Outcomes. Horticulturae **2024**, 10, 152. https://doi.org/10.3390/horticulturae10020152

[10] Litvak, M.; Divekar, S.; Rabaev, I. Urban Plants Classification Using Deep-Learning Methodology: A Case Study on a New Dataset. Signals **2022**, 3, 524-534. https://doi.org/10.3390/signals3030031

[11] Das, M.; Deb, C.K.; Pal, R.; Marwaha, S. A Machine Learning Approach for the Non-Destructive Estimation of Leaf Area in Medicinal Orchid Dendrobium nobile L. Appl. Sci. **2022**, 12, 4770. https://doi.org/10.3390/app12094770

[12] Batchuluun, G.; Nam, S.H.; Park, K.R. Deep Learning-Based Plant-Image Classification Using a Small Training Dataset. Mathematics **2022**, 10, 3091. https://doi.org/10.3390/math10173091

[13] S. Puttemans and T. Goedemé, "Visual detection and species classification of orchid flowers," 2015 14th IAPR International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 2015, pp. 505-509, doi: 10.1109/MVA.2015.7153241.

[14] Sarachai, W., Bootkrajang, J., Chaijaruwanich, J., Somhom, S. (2019). Orchids Classification Using Spatial Transformer Network with Adaptive Scaling. In: Yin, H., Camacho, D., Tino, P., Tallón-Ballesteros, A., Menezes, R., Allmendinger, R. (eds) Intelligent Data Engineering and Automated Learning – IDEAL 2019. IDEAL 2019. Lecture Notes in Computer Science(), vol 11871. Springer, Cham. https://doi.org/10.1007/978-3-030-33607-3_1

[15] Serrano Pereira, Barbara Gravendeel, Patrick Wijntjes, Rutger A. Vos bioRxiv 070904; OrchID: a Generalized Framework for Taxonomic Classification of Images Using Evolved Artificial Neural Networks doi: https://doi.org/10.1101/070904