Ethnobotanical risk assessment of visually similar medicinal plants: a machine learning approach to identifying species prone to misidentification in traditional medicine

I Gusti Ngurah Sentana Putra, Muhammad Syafiq, Utami Dyah Syafitri

Correspondence

I Gusti Ngurah Sentana Putra[1*], Muhammad Syafiq[1], Utami Dyah Syafitri[1]

[1] School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

*Corresponding Author: ngurahsentana@apps.ipb.ac.id

Ethnobotany Research and Applications 35:44 (2026) - http://dx.doi.org/10.32859/era.35.44.1-28
Manuscript received: 08/06/2026 - Revised manuscript received: 30/07/2026 - Published: 02/08/2026

Research

Abstract

Background: Plant misidentification in traditional medicine is an underreported public health risk because visually similar medicinal species are often confused, leading to therapeutic failure or toxicity. Previous work addressed medicinal plant image classification or documented misidentification separately, but none linked machine-learning confusion patterns with pharmacological risk profiling.

Methods: This study trained a Convolutional Neural Network (CNN) and a Siamese Network with triplet loss on 20,000 leaf images from 20 medicinal plant species, sourced from two Bangladeshi repositories. The Siamese Network extracted 128-dimensional embeddings to compute Euclidean distances for all 190 species pairs. Misclassification rates were used to assign risk categories, and Pearson correlation tested the link between embedding proximity and error rates.

Results: The CNN achieved 88.45% overall accuracy. The Siamese Network identified 14 critical pairs (≥6% misclassification) and 7 high-risk pairs (4-5%). The most confusable pair was *Justicia adhatoda* misidentified as *Terminalia arjuna* (14%), followed by *J. adhatoda* as *Centella asiatica* (12%), *Kalanchoe pinnata* as T. arjuna (11%), and K. pinnata as *Mikania micrantha* (10%). A significant negative correlation ($r = -0.304$, $p < 0.0001$) confirmed that closer embeddings correspond to higher error rates.

Conclusions: The 21 high-risk pairs involve pharmacologically incompatible compounds: vasicine from *J. adhatoda* acts as a uterotonic and abortifacient; bufadienolides from K. pinnata cause cardiac glycoside-like toxicity; cardenolides from *Calotropis gigantea* are cardiotoxic; and Hibiscus rosa-sinensis has antifertility effects. These findings support prioritizing field training, visual guides, and digital tools to reduce plant misidentification in herbal medicine.

Keywords: Medicinal plant misidentification; Ethnobotany; Deep learning; Siamese Network; Traditional medicine safety; Pharmacological risk; Plant morphology; Visual Similar

Background

Traditional plant-based medicine remains the primary healthcare system for an estimated 80% of the global population in low- and middle-income countries, according to World Health Organization data. (Eddouks *et al.* 2023; Sofowora *et al.* 2013; WHO, 2019). This reliance reflects not only economic accessibility but also centuries of accumulated ethnobotanical knowledge transmitted through apprenticeship, oral traditions, and direct community practice (Pan *et al.* 2013). However, this knowledge transmission system carries an inherent vulnerability: the absence of standardized morphological keys for field identification in the hands of practitioners, community harvesters, and household consumers (Biswas *et al.* 2021; Ekor, 2014; Gairola *et al.* 2021; Kyei *et al.* 2022).

Plant misidentification has historically produced some of the most devastating phytomedicine safety disasters on record. The substitution of Stephania tetrandra S.Moore with Aristolochia fangchi Y.C.Wu ex L.D.Chow & S.M.Hwang two species that share superficially similar leaf morphology, caused an epidemic of aristolochic acid nephropathy affecting hundreds of patients in Belgium and subsequently documented across multiple countries, with a high proportion progressing to end-stage renal failure and urothelial malignancy (Teschke & Eickhoff, 2023). In traditional medicine systems of South and Southeast Asia, similar substitution events, often undocumented and thus statistically invisible, are believed to contribute substantially to adverse events attributed broadly to "herbal toxicity" (Ekor, 2014). The WHO strategy document on traditional and complementary medicine (WHO 2019) explicitly identifies incorrect species identification as a primary quality-control failure point in herbal product supply chains.

The pharmacological consequences of consuming one plant in place of another depend entirely on the phytochemical divergence between the two species. When morphologically similar species have convergent pharmacological profiles, the clinical impact of misidentification may be minimal. However, many cases involve species that have undergone morphological convergence without pharmacological convergence, a particularly dangerous combination (Hazekamp & Fischedick, 2012). Uterotonic alkaloids, cardiac glycosides, phorbol ester irritants, and antifertility flavonoids do not announce their presence visually; they are detectable only through chemical analysis or tragically through adverse clinical outcomes (Kyei *et al.* 2022; Teschke & Eickhoff, 2023).

Deep learning, particularly Convolutional Neural Networks, has emerged as a powerful paradigm for automating plant species identification from leaf images, achieving accuracies that equal or surpass non-specialist human observers under controlled conditions (Bouakkaz *et al.* 2025; Dhaka *et al.* 2021; Mulugeta *et al.* 2024). However, the value of these models in the ethnobotanical safety context lies not only in their classification accuracy but in their capacity to fail in informative ways. When a neural network systematically confuses two species at high rates, this constitutes quantitative evidence that the two species share feature-space proximity, a computational proxy for the visual similarity that confounds human observers in the field (Lankasena *et al.* 2024). This reframing is directly relevant to ethnobotanical practice: kaviraj practitioners and community harvesters in South Asia rely primarily on morphological cues, leaf shape, venation, texture, and colour, for species recognition in the absence of botanical keys or laboratory support. A machine learning model trained on the same leaf images these practitioners would encounter provides, through its systematic error patterns, a quantitative map of the visual ambiguities that pose risk to those practitioners and their patients.

The decision to combine a CNN classifier with a Siamese embedding network, rather than adopting a single-tool approach as several comparable studies have done, was deliberate rather than incidental. A CNN alone yields a categorical misclassification rate for each species pair, but this scalar count offers no continuous measure of how close two species sit in feature space, nor does it generalise beyond the fixed set of pairs observed during testing. A Siamese network alone, conversely, yields a continuous similarity metric but does not by itself establish which of these similarities translate into practical identification errors under the classifier a practitioner-facing tool would actually deploy. By pairing the two, this study obtains both a direct, empirically observed error rate (from the CNN) and an independent, continuous geometric explanation for why that error occurs (from the Siamese embedding space), and the correlation between the two (Pearson $r = -0.304$, $p < 0.0001$) provides internal validation that is unavailable to single-tool designs. This combined-tool strategy therefore directly serves the study's stated aim of producing a risk atlas that is both behaviourally grounded and mechanistically interpretable, which single-instrument designs in prior CNN-only or Siamese-only medicinal plant studies could not achieve.

The ethnobotanical context of Bangladesh is particularly relevant here. Kaviraj practitioners typically acquire their knowledge through apprenticeship and oral transmission rather than written botanical keys; species are identified by a combination of leaf feel, aroma, habitat, associated plants, and seasonal availability (Islam & Uddin, 2022; Rudra *et al.* 2022). Of these cues, leaf morphology observed in the hand or from a photograph is often the only one available at the point of market purchase or household collection, particularly when plants are sold as dried or detached leaves stripped of their flowers, fruits, and stems. It is precisely this reduced-information identification context that the machine learning models in this study simulate: both the CNN classifier and the Siamese Network operate exclusively on leaf image data, mirroring the information a practitioner or household consumer has available when identifying a leaf presented without its full botanical context.

Furthermore, the species in this study are not merely botanical curiosities: they represent the everyday pharmacopoeia of millions of households across Bangladesh, India, Sri Lanka, and beyond. *Centella asiatica* (thankuni), *Ocimum tenuiflorum* (tulsi), and *Azadirachta indica* (neem) are consumed daily in teas, salads, and tonics without any formal identification step. The trust placed in these widely familiar species creates a specific vulnerability: when a morphologically similar but pharmacologically divergent species is mistaken for one of these trusted plants, the consumer has no reason to expect harm, there is no unusual taste or smell to trigger caution, and there is no reporting mechanism to link the adverse event to the botanical substitution. The pair catalog produced in this study is therefore not merely a classification benchmark; it is a map of the specific points in the traditional medicine supply chain where silent pharmacological substitution is most likely to occur.

Siamese Networks with triplet loss, originally developed for biometric identification tasks (Schroff *et al.* 2015), extend this capability by learning a metric embedding space in which intra-species distances are minimized and inter-species distances are maximized. The resulting Euclidean distances between class centroids provide a continuous quantitative measure of morphological similarity that can be correlated with real-world misidentification risk (Figueroa-Mata & Mata-Montero, 2020). To our knowledge, no prior study has integrated Siamese Network embedding analysis with systematic pharmacological risk profiling to produce a comprehensive misidentification risk map across a diverse, multi-species collection of medicinal plants.

The present study employs two publicly available image datasets from Bangladesh, a country with a rich and well-documented tradition of kaviraj (traditional healer) practice and a pharmacopoeially diverse flora (Islam & Uddin, 2022; Rudra *et al.* 2022; Uddin & Zidorn, 2020) as a case study for a broadly applicable methodology. Bangladesh was selected not merely for convenience of data availability but because it is one of the few settings where two independently curated, peer-reviewed leaf-image datasets covering pharmacologically significant species already existed in the public domain, allowing this analysis to be built on validated data rather than newly collected images. This addresses a genuine research gap: despite Bangladesh's dense reliance on kaviraj practice and its diverse medicinal flora, no prior study had combined these two resources into a misidentification risk atlas, and the ten overlapping and ten complementary species across the two datasets provide the paired-source structure this analysis requires. That said, because Bangladesh is the only geographic source represented, the resulting risk pairs should be read as evidence from this setting rather than as a global inventory of confusable species, a limitation discussed further below. The 20 species analyzed span both datasets and include many of the most widely used medicinal plants across South Asia, Southeast Asia, and Africa (Kutal *et al.* 2021). The objectives are: (1) to train and evaluate a Convolutional Neural Network for 20-species leaf classification; (2) to construct a Siamese Network embedding space and compute quantitative inter-species similarity metrics; (3) to identify and categorize all critical and high-risk species pairs by misclassification rate and embedding proximity; and (4) to systematically review the pharmacological profiles of species in all 21 risk pairs and characterize the potential human health consequences of each specific confusion event, thereby producing the first evidence-based misidentification risk atlas for these species.

Materials and Methods

Note on the computational tools used

Note for readers unfamiliar with machine learning methods: this study uses two complementary computational tools to measure how visually similar medicinal plant leaves are to each other. The first (a Convolutional Neural Network, CNN) is trained to classify leaf images by species; its errors reveal which species pairs a visual recognition system confuses most often. The second (a Siamese Network) produces a numerical "distance" score between every pair of species: a smaller distance means the leaves look more alike. Neither tool requires the reader to have a background in computing; the outputs, misclassification rates and distance scores, can be read like any other quantitative similarity measure in ecology or taxonomy. The pharmacological significance of the pairs identified is discussed fully in plain language throughout the Results and Discussion.

Dataset sources and species composition

Two publicly available and peer-reviewed image datasets were combined. The first, the Bangladeshi Medicinal Plant Dataset (Borkatulla *et al.* 2023), is available on Kaggle and comprises leaf images of ten species, including *Terminalia bellirica* (Gaertn.). Roxb. (Bohera), *Euphorbia tithymaloides* L. (Devilbackbone), *Terminalia chebula* Retz. (Haritoki), *Cymbopogon citratus* (DC.) Stapf (Lemongrass), *Catharanthus roseus* (L.) G.Don (Nayontara), *Azadirachta indica* A. Juss. (Neem), *Bryophyllum pinnatum* (Lam.) Oken (Pathorkuchi), *Centella asiatica* (L.) Urb. (Thankuni), *Ocimum tenuiflorum* L. (Tulsi), and *Merremia vitifolia* (Burm.f.) Hallier f. (Zenora). Each class contains 500 original images augmented to 1,000 images per class, with a 70:20:10 train-validation-test split.

The second dataset, BDMediLeaves (Ahmed *et al.* 2023), is deposited on Mendeley Data and includes leaf images of ten additional species: *Azadirachta indica* A. Juss. (AI), *Calotropis gigantea* (L.) Dryand. (CG), *Centella asiatica* (L.) Urb. (CA), *Hibiscus rosa-sinensis* L. (HRS), *Justicia adhatoda* L. (JA), *Kalanchoe pinnata* (Lam.) Pers. (KP), *Mikania micrantha* Kunth (MM), *Ocimum tenuiflorum* L. (OT), *Terminalia arjuna* (Roxb. ex DC.) Wight & Arn. (TA), and *Phyllanthus emblica* L. (PE). Original image counts per class range from 151 to 229, augmented to 1,000 per class following the same 70:20:10 split. Botanical readers should note that *Bryophyllum pinnatum* (Lam.) Oken (Pathorkuchi, Dataset 1) and *Kalanchoe pinnata* (Lam.) Pers. (KP, Dataset 2) are accepted as synonyms in the current taxonomy; similarly, Neem (Dataset 1) and AI (Dataset 2) represent the same species, *Azadirachta indica* A. Juss. captured under different imaging conditions. Both are treated as distinct classification targets in this study because they originate from different collection protocols and geographic sub-regions, and because their intra-class variability is of independent analytical interest.

All images were resized to 224 × 224 pixels. Data augmentation included random horizontal and vertical flips, rotation at 30°, 60°, and 90°, random affine transformation (rotation ±10°, translation 10% of image dimensions, scale 90-110%), and pixel intensity normalization (Shorten & Khoshgoftaar, 2019). The combined dataset comprised 20,000 images across 20 classes (14,000 training, 4,000 validation, 2,000 test).

Dataset scope and limitations

Several dataset limitations are important for readers to consider when interpreting the results. First, both source datasets were collected in Bangladesh under standardized photographic conditions; all images depict isolated, flattened leaves against uniform backgrounds, and do not include flowers, fruits, bark, or whole-plant architecture. Real-world field identification by kaviraj practitioners and community harvesters typically involves multi-organ assessment under variable lighting, and the trained models may not generalize directly to those conditions without further validation. Second, although data augmentation was applied to introduce geometric and photometric variation, this does not substitute for geographically diverse field collections; intra-species variation in leaf morphology across different growing regions, seasons, and developmental stages is not captured in the present dataset. Third, the 20-species set covers plants documented in Bangladeshi ethnomedicinal practice and cannot be assumed to represent species composition or visual confusion patterns in other traditional medicine systems. Readers should treat the misidentification risk pairs identified here as ethnobotanically relevant case studies within this geographic and dataset scope rather than exhaustive maps of global confusion risk. Fourth, image augmentation was applied uniformly and does not capture the full range of natural morphological variation within a species caused by age, soil type, water stress, or seasonal growth stage; a mature *Justicia adhatoda* leaf from a monsoon-season plant may differ substantially from a juvenile leaf photographed under drought conditions, and such within-species variation is not represented in the augmented training set. Fifth, the study design treats the two source datasets as a combined pool; however, the datasets were collected under different protocols and geographic sub-regions of Bangladesh, meaning that some inter-dataset species pairs (e.g., Neem from Dataset 1 and AI from Dataset 2, both *Azadirachta indica*) may carry methodological variation unrelated to genuine biological similarity. Readers designing replication studies should consider using a single harmonized dataset or explicitly modelling dataset of origin as a covariate. Future work should also assess model performance on images collected in actual field conditions by non-specialist collectors, as this more closely represents the deployment context for community-based plant identification tools. Taken together, these five limitations affect the external validity and interpretive scope of the findings, but they do not undermine the minimum data requirements needed for the two analyses performed here. For CNN classification, the smaller class (151-229 original images, augmented to 1,000) still exceeds commonly cited minimum thresholds of roughly 100-150 images per class for transfer-learning-based leaf classifiers, which is consistent with the 88.45% test accuracy achieved. For the Siamese Network, triplet loss requires only that enough same-class pairs exist to mine informative positives and negatives; with 100 test images per class and 14,000 valid training triplets, this requirement was comfortably met, as reflected in the converged triplet loss of 0.198. The five limitations above therefore constrain how far the resulting risk pairs can be generalized geographically and ecologically, rather than calling into question whether the underlying classification and embedding analyses were adequately powered.

Dataset scope and limitations

Several dataset limitations are important for readers to consider when interpreting the results. First, both source datasets were collected in Bangladesh under standardized photographic conditions; all images depict isolated, flattened leaves against uniform backgrounds, and do not include flowers, fruits, bark, or whole-plant architecture. Real-world field identification by kaviraj practitioners and community harvesters typically involves multi-organ assessment under variable lighting, and the trained models may not generalize directly to those conditions without further validation. Second, although data augmentation was applied to introduce geometric and photometric variation, this does not substitute for geographically diverse field collections; intra-species variation in leaf morphology across different growing regions, seasons, and developmental stages is not captured in the present dataset. Third, the 20-species set covers plants documented in Bangladeshi ethnomedicinal practice and cannot be assumed to represent species composition or visual confusion patterns in other traditional medicine systems. Readers should treat the misidentification risk pairs identified here as ethnobotanically relevant case studies within this geographic and dataset scope rather than exhaustive maps of global confusion risk. Fourth, image augmentation was applied uniformly and does not capture the full range of natural morphological variation within a species caused by age, soil type, water stress, or seasonal growth stage; a mature *Justicia adhatoda* leaf from a monsoon-season plant may differ substantially from a juvenile leaf photographed under drought conditions, and such within-species variation is not represented in the augmented training set. Fifth, the study design treats the two source datasets as a combined pool; however, the datasets were collected under different protocols and geographic sub-regions of Bangladesh, meaning that some inter-dataset species pairs (e.g., Neem from Dataset 1 and AI from Dataset 2, both *Azadirachta indica*) may carry methodological variation unrelated to genuine biological similarity. Readers designing replication studies should consider using a single harmonized dataset or explicitly modelling dataset of origin as a covariate. Future work should also assess model performance on images collected in actual field conditions by non-specialist collectors, as this more closely represents the deployment context for community-based plant identification tools. Taken together, these five limitations affect the external validity and interpretive scope of the findings, but they do not undermine the minimum data requirements needed for the two analyses performed here. For CNN classification, the smaller class (151-229 original images, augmented to 1,000) still exceeds commonly cited minimum thresholds of roughly 100-150 images per class for transfer-learning-based leaf classifiers, which is consistent with the 88.45% test accuracy achieved. For the Siamese Network, triplet loss requires only that enough same-class pairs exist to mine informative positives and negatives; with 100 test images per class and 14,000 valid training triplets, this requirement was comfortably met, as reflected in the converged triplet loss of 0.198. The five limitations above therefore constrain how far the resulting risk pairs can be generalized geographically and ecologically, rather than calling into question whether the underlying classification and embedding analyses were adequately powered.

Convolutional Neural Network Architecture and Training

In brief, a CNN is a type of deep learning model that automatically learns to detect visual features in images (such as leaf edge shape, vein patterns, and texture) and uses those features to classify which species a leaf belongs to. The Convolutional Neural Network (CNN) was designed as a 20-class image classifier, structured using sequential convolutional blocks (Lecun *et al.* 2015; LeCun *et al.* 1998). Each block consisted of a convolutional layer with a 3×3 kernel and a stride of 1, followed by batch normalization, ReLU activation, and a 2×2 max-pooling layer (Bouakkaz *et al.* 2025). The convolution operation for an input feature map X and filter W is defined as:

$$Y[i,j] = (X * W)[i,j] = \sum_{m} \sum_{n} X[i+m, j+n] \cdot W[m,n] \tag{1}$$

Batch normalization was applied to accelerate training and reduce internal covariate shift. For a given mini-batch $\mathcal{B} = x_1, \ldots, x_m$, the normalized output is:

$$\hat{x}_i = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}}, y_i = \gamma \hat{x}_i + \beta \tag{2}$$

where $\mu_{\mathcal{B}}$ and $\sigma_{\mathcal{B}}^2$ are the batch mean and variance, ϵ is a small constant for numerical stability, and γ, β are learnable parameters. The ReLU activation introduces non-linearity:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{3}$$

The final convolutional block was followed by a flattening layer and two fully-connected (dense) layers, terminating in a 20-neuron softmax output layer. The softmax function converts logits z_j into class probabilities:

$$p(y = j \mid \mathbf{z}) = \frac{e^{z_j}}{\sum_{k=1}^{20} e^{z_k}} \quad (4)$$

The network was optimized using the Adam optimizer to minimize categorical cross-entropy loss. For a true label distribution y and predicted probabilities $\hat{y}$, the loss is:

$$\mathcal{L}_{\text{CE}} = -\sum_{j=1}^{20} y_j \log(\hat{y}_j) \quad (5)$$

Training proceeded for 11 epochs with a batch size of 32. The same backbone architecture was later reused as the shared-weight encoder branch in the Siamese Network.

Siamese Network with Triplet Loss

In brief, a Siamese Network is a pair of identical neural networks trained together to measure how visually similar two images are, producing a numerical distance score rather than a categorical label. The Siamese Network employed the trained CNN as a shared encoder $f_\theta(\cdot)$, producing 128-dimensional embedding vectors from input images. To ensure meaningful similarity comparisons, the embeddings were $L2$-normalized:

$$\mathbf{e} = \frac{f_\theta(x)}{\| f_\theta(x) \|_2}, \| \mathbf{e} \|_2 = 1 \quad (6)$$

The network was trained using triplet loss, which operates on triplets consisting of an anchor a, a positive sample p (same class as anchor), and a negative sample n (different class). The triplet loss function is:

$$\mathcal{L}_{\text{triplet}}(a, p, n) = \max\left(\| \mathbf{e}_a - \mathbf{e}_p \|_2^2 - \| \mathbf{e}_a - \mathbf{e}_n \|_2^2 + \alpha, 0\right) \quad (7)$$

Here, $\alpha = 0.2$ is the margin parameter ensuring a minimum separation between positive and negative distances. The squared Euclidean distance is defined as:

$$\| \mathbf{e}_a - \mathbf{e}_p \|_2^2 = \sum_{k=1}^{128} (e_{a,k} - e_{p,k})^2 \quad (8)$$

A total of 14,000 valid triplets were generated from the training set, where for each anchor, the hardest positive and hardest negative within a batch were selected (semi-hard negative mining). The final triplet loss after convergence reached 0.198, indicating effective embedding space separation. To prevent data leakage, the train, validation, and test split was performed at the level of the original, pre-augmentation specimen images, so that augmented copies of the same source image were assigned to only one split and never appeared in more than one of the training, validation, and test sets. Full implementation details, including software and hardware versions, dropout and regularization settings, random seeds, and the early-stopping criterion used for model selection, are reported in the Supplementary Information.

After training, the centroid embedding for each class c was computed by averaging embeddings of all 100 test samples from that class:

$$\mu_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \mathbf{e}_i^{(c)} \quad (9)$$

where $N_c = 100$ for all $c = 1, \ldots, 20$. This formed a 20×128 centroid matrix. Because these centroids, and the resulting distance-based risk map, are computed from the held-out test set, they are reported here as a post-hoc characterization of the trained model's error structure rather than as a deployment-ready risk map; extending this map to field use would require external validation on an independent image set not used in training, validation, or centroid computation.

Quantification of Inter-Species Visual Similarity

To quantify visual similarity between species, the Euclidean distance between each pair of class centroids i and j was calculated:

$$d(i,j) = \| \mu_i - \mu_j \|_2 = \sqrt{\sum_{k=1}^{128} (\mu_{i,k} - \mu_{j,k})^2} \tag{10}$$

This was computed for all $\binom{20}{2} = 190$ unique pairs, generating a symmetric 20×20 distance matrix. To express proximity on a normalized scale, a similarity score $S(i,j) \in [0,1]$ was derived:

$$S(i,j) = \frac{1}{1 + d(i,j)} \tag{12}$$

The rationale is that as distance $d(i,j) \to 0$, similarity approaches 1; as $d(i,j) \to \infty$, similarity approaches 0. Because the L2-normalized embedding distances observed in this dataset are small, S(i,j) may compress toward 1 across most pairs and so should be interpreted as a rank-preserving transformation rather than a calibrated probability-like score; ranking pairs by raw distance d(i,j), or rescaling S(i,j) to the observed minimum-maximum range of distances, is likely to be more interpretable for readers than the untransformed 1/(1+d) values reported here.

The strength of association between embedding distance and per-pair misclassification rate $M(i,j)$ (derived from the original classifier confusion matrix) was assessed using the Pearson product-moment correlation coefficient: Because misclassification counts are bounded count data and the same species recurs across many of the 380 directed comparisons (violating independence), the Pearson coefficient reported below should be read as a descriptive association rather than an independent-samples test; a rank-based Spearman correlation and a species-clustered permutation or bootstrap procedure, together with confidence intervals rather than a p-value alone, would provide more defensible inference and are recommended before this correlation is presented as confirmatory evidence.

$$r = \frac{\sum_{i<j} (d(i,j) - \bar{d})(M(i,j) - \bar{M})}{\sqrt{\sum_{i<j} (d(i,j) - \bar{d})^2 \sum_{i<j} (M(i,j) - \bar{M})^2}}$$

where $\bar{d}$ and $\bar{M}$ are the means over all 190 pairs.

For two-dimensional visualization of the embedding space, t-distributed Stochastic Neighbor Embedding (der Maaten & Hinton, 2008) was applied to the 128-dimensional centroids. t-SNE minimizes the Kullback-Leibler divergence between joint probabilities in high-dimensional and low-dimensional spaces. The similarity in high dimension is defined as:

$$p_{j|i} = \frac{\exp(-\| \mu_i - \mu_j \|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\| \mu_i - \mu_k \|^2 / 2\sigma_i^2)}$$

and symmetrized as $p_{ij} = (p_{j|i} + p_{i|j})/2n$. In the 2D map with points y_i, y_j, the similarity is modeled by a Student-t distribution with one degree of freedom:

$$q_{ij} = \frac{\left(1 + \| y_i - y_j \|^2\right)^{-1}}{\sum_{k \neq l} (1 + \| y_k - y_l \|^2)^{-1}}$$

The optimization minimizes $\mathrm{KL}(P \parallel Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$. The Barnes-Hut approximation with perplexity = 30 was used to improve computational efficiency. The t-SNE projection was applied to the full set of 2,000 individual test-sample embeddings (100 per class), not to the 20 class centroids, consistent with Figure 6; a perplexity of 30 is appropriate at this sample size. Initialization, learning rate, and the number of iterations followed the Barnes-Hut t-SNE defaults, and are reported in the Supplementary Information together with the random seed used.

Risk categorization

Each of the 190 directed pairs was assigned a risk category based on the misclassification count across 100 test samples: critical (≥6 errors), high (4-5 errors), or medium (<4 errors). For all critical and high-risk pairs, a structured pharmacological risk profile was compiled from the peer-reviewed literature, detailing primary bioactive compounds, known therapeutic uses, documented toxicity or contraindications, and the specific pharmacological consequence of consuming species A in place of species B. Note on pair terminology: with 20 species there are 190 unordered (undirected) species pairs; because misclassification is asymmetric (the rate at which species A is misidentified as B need not equal the rate at which B is misidentified as A), each unordered pair yields two directed comparisons, for 380 directed comparisons overall. Table 2 reports directional misclassification (true species → predicted species), so risk categories and denominators throughout this study are defined at the level of the 380 directed comparisons rather than the 190 unordered pairs.

The critical/high/medium thresholds were evaluated using Wilson score 95% confidence intervals for a binomial proportion with n = 100 test images per class. At the high/critical boundary, an error count of 4/100 corresponds to a rate of 4% (95% CI: 1.6-9.8%) and 6/100 to 6% (95% CI: 2.8-12.5%); these intervals overlap substantially, confirming that a single-pair distinction between, for example, 4/100 and 6/100 errors is not statistically sharp at this sample size. The critical/high/medium categories are therefore best interpreted as an ordinal, practically motivated ranking of relative risk within this dataset rather than as statistically distinct strata, and Table 2 has been read alongside these confidence intervals rather than the point estimates alone.

Results

Convolutional Neural Network training performance

Training dynamics are shown in Figure 1. Validation loss decreased monotonically from 2.73 (epoch 0) to 0.47 (epoch 10), and validation accuracy increased from 22% to 85%, with the training and validation curves tracking closely, indicating adequate generalization without substantial overfitting. The overall test accuracy on 2,000 unseen images was 88.45% (1,769 correct predictions). Macro-averaged precision was 0.888, recall 0.884, and F1-score 0.882.

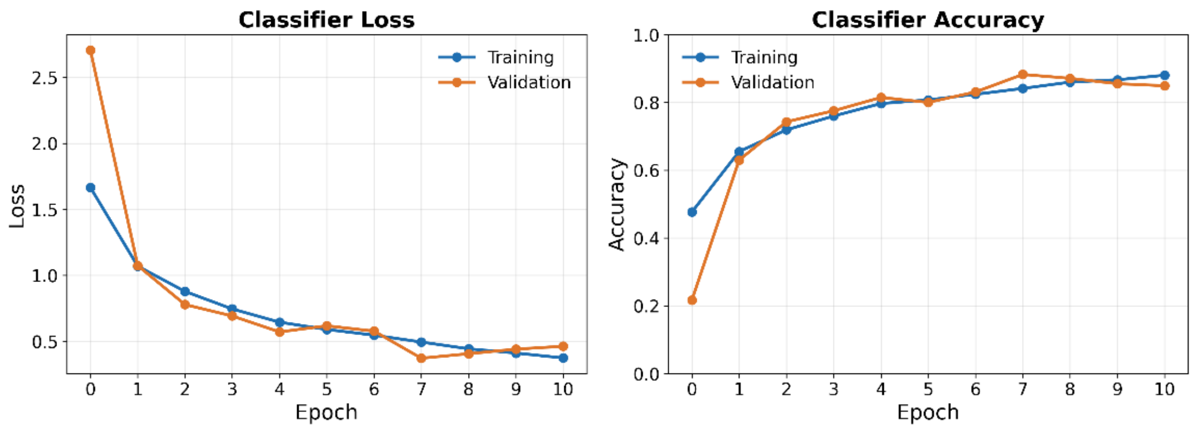

Figure 1. Convolutional Neural Network training history: classifier loss (left) and accuracy (right) across 11 training epochs. Training curves (blue) and validation curves (orange) are shown.

Per-class classification performance and embedding properties

Per-class results are summarized in Table 1 and Figure 2. Five species achieved accuracy ≥98%: Neem *Azadirachta indica* A. Juss. (100%), *Phyllanthus emblica* L. (100%), Bohera *Terminalia bellirica* (Gaertn.) Roxb. (99%), Zenora *Merremia vitifolia* (Burm.f.) Hallier f. (99%), and Nayontara *Catharanthus roseus* (L.) G.Don (98%). The four lowest-performing species were *Justicia adhatoda* L. (62%) and *Kalanchoe pinnata* (Lam.). Pers. (66%), *Ocimum tenuiflorum* L. from BDMediLeaves (68%), and *Mikania micrantha* Kunth (72%). Embedding separation analysis (average Euclidean distance of each class centroid to all other class centroids) showed that *Cymbopogon citratus* (DC.) Stapf (0.087) was the most isolated species in the embedding space, while *Catharanthus roseus* (L.) G.Don (0.042) was the least separated yet retained high accuracy (98%), indicating that the model found discriminative features beyond centroid distance alone.

Confusion matrix: systematic misclassification patterns

The full confusion matrix (Figure 3) reveals highly non-random error patterns concentrated in three overlapping confusion clusters. The *Justicia adhatoda* L. cluster exhibited the most severe errors: 14 samples were predicted as *Terminalia arjuna* Wight & Arn., and 12 as *Centella asiatica* (L.). Urb., and 6 as *Azadirachta indica* A. Juss. The *Kalanchoe pinnata* (Lam.) Pers. cluster showed 11 errors as *Terminalia arjuna* Wight & Arn. and 10 as *Mikania micrantha* Kunth. The *Ocimum tenuiflorum* L. cluster generated 8 errors as *Centella asiatica* (L.) Urb., 8 as *Hibiscus rosa-sinensis* L., and 7 as *Justicia adhatoda* L. By contrast, Neem *Azadirachta indica* A. Juss. and *Phyllanthus emblica* L. produced zero classification errors across 100 test samples each.

To contextualize these error rates for ethnobotanical readers: a misclassification rate of 14% for *Justicia adhatoda* L. as *Terminalia arjuna* means that in a dataset of 100 standardized photographs under controlled conditions, 14 were assigned the wrong label by an optimized deep learning classifier. In real field conditions, with variable lighting, partial leaf views, and morphological variation from season and growing region, the confusion rate between these species would reasonably be expected to be higher. Equally important is the asymmetry of harm: a practitioner who consistently identifies *Centella asiatica* correctly but occasionally supplies *Justicia adhatoda* to a pregnant patient has committed an error with potentially serious reproductive consequences, even if that error occurs in only a small fraction of identifications. The misclassification rates reported here should therefore be read not as abstract performance metrics but as probability estimates of risk events within a traditional medicine supply chain.

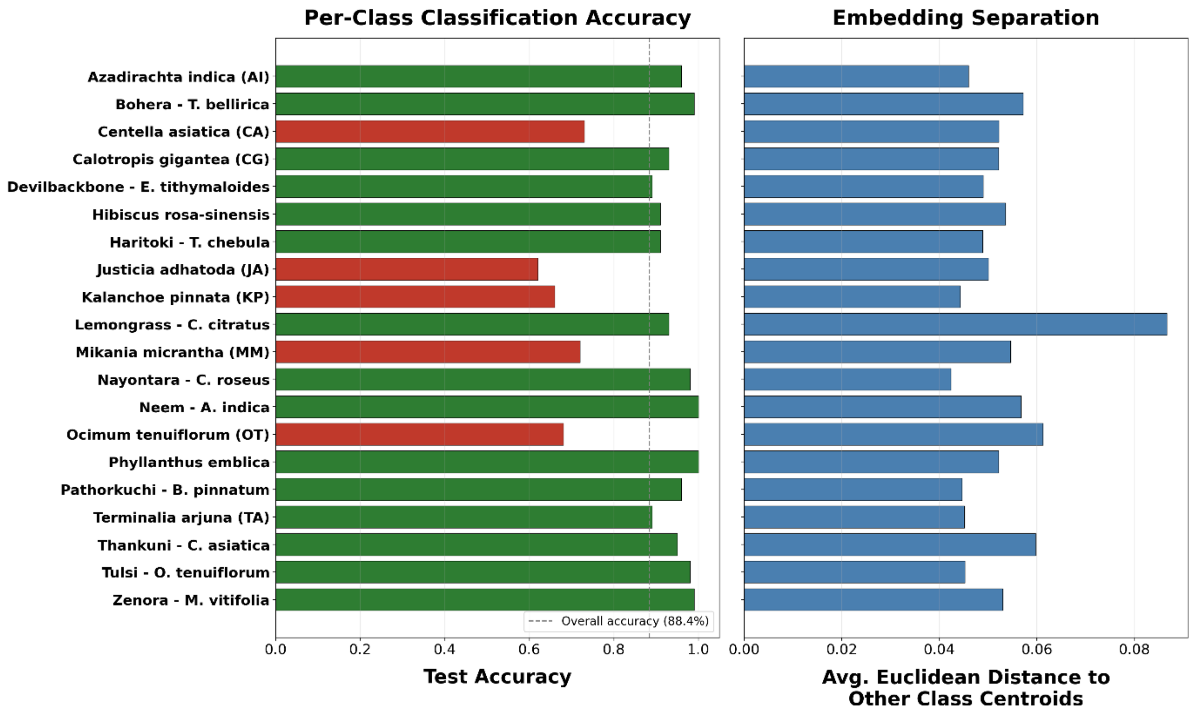

Figure 2. Per-class classification accuracy (left) and Siamese Network embedding separation, measured as average Euclidean distance to other class centroids (right), for 20 medicinal plant species on 2,000 test samples (100 per species).

Table 1. Per-class classification performance and Siamese Network embedding properties on 2,000 test samples (100 per species).

Species	**Acc.(%)**	**Prec.**	**Recall**	**F1**	**Intra-spread**	**Avg. dist. to other classes**
Azadirachta indica A. Juss. (AI)	96	0.850	0.96	0.901	0.029	0.046
Bohera *Terminalia bellirica* Roxb.	99	0.990	0.99	0.990	0.026	0.057
Centella asiatica (L.) Urb. (CA)	73	0.702	0.73	0.716	0.035	0.052
Calotropis gigantea (L.) Dryand. (CG)	93	0.921	0.93	0.925	0.027	0.052
Devilbackbone *Euphorbia tithymaloides* L.	89	0.957	0.89	0.922	0.042	0.049
Hibiscus rosa-sinensis L.	91	0.843	0.91	0.875	0.037	0.054
Haritoki *Terminalia chebula* Retz.	91	0.968	0.91	0.938	0.029	0.049
Justicia adhatoda L.	62	0.765	0.62	0.685	0.038	0.050
Kalanchoe pinnata (Lam.) Pers.	66	0.985	0.66	0.790	0.046	0.044
Lemongrass *Cymbopogon citratus* Stapf	93	0.969	0.93	0.949	0.034	0.087
Mikania micrantha Kunth	72	0.809	0.72	0.762	0.041	0.055
Nayontara *Catharanthus roseus* G.Don	98	0.952	0.98	0.966	0.026	0.042
Neem *Azadirachta indica* A. Juss.	100	0.962	1.00	0.980	0.021	0.057
Ocimum tenuiflorum L. (OT)	68	0.782	0.68	0.727	0.028	0.061
Phyllanthus emblica L.	100	0.781	1.00	0.877	0.041	0.052

Pathorkuchi *Bryophyllum pinnatum* Oken	96	0.951	0.96	0.955	0.022	0.045
Terminalia arjuna Wight & Arn.	89	0.730	0.89	0.802	0.032	0.045
Thankuni *Centella asiatica* Urb.	95	0.960	0.95	0.955	0.031	0.060
Tulsi *Ocimum tenuiflorum* L.	98	0.942	0.98	0.961	0.038	0.045
Zenora *Merremia vitifolia* Hallier f.	99	0.934	0.99	0.961	0.021	0.053

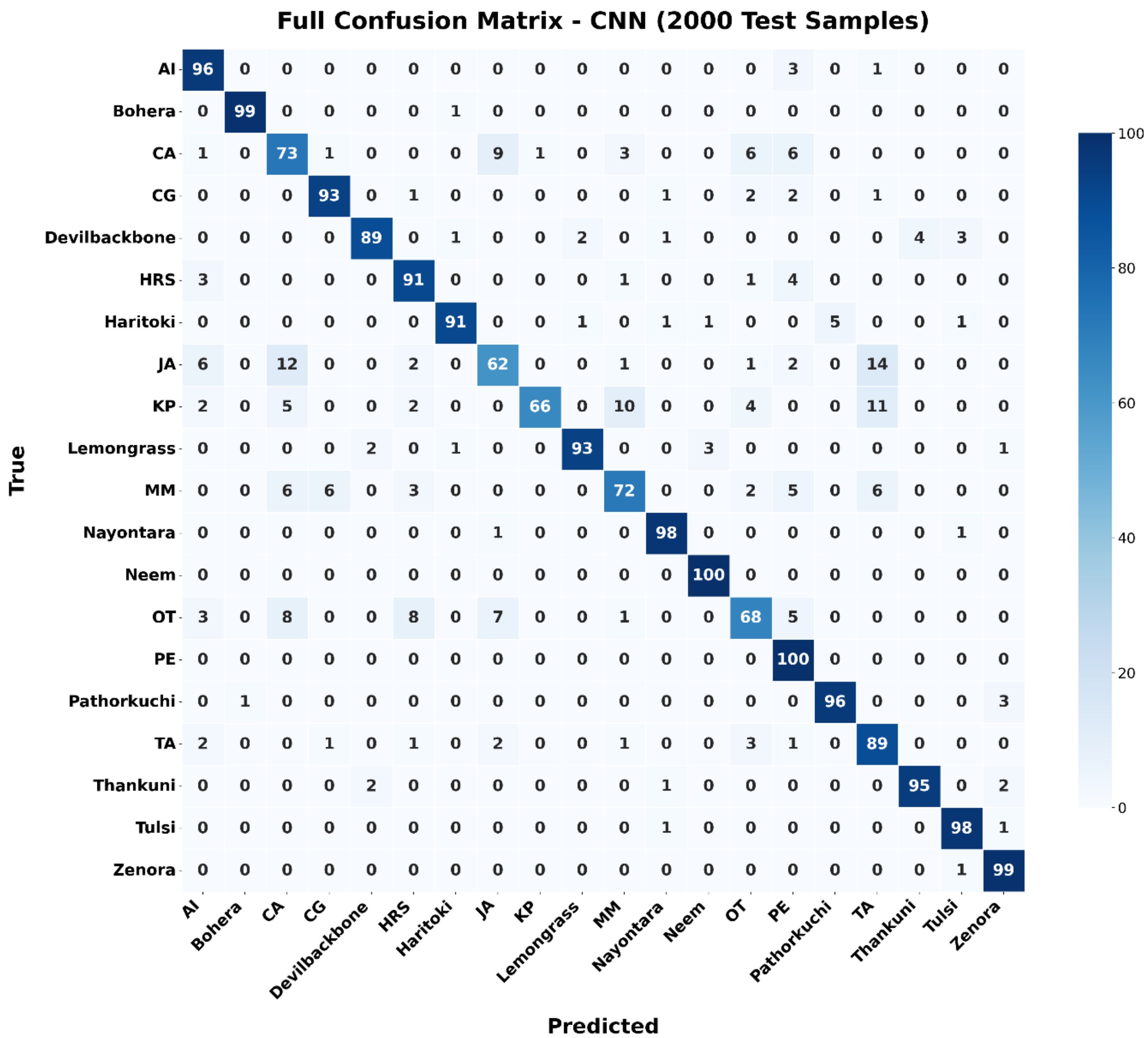

Figure 3. Full confusion matrix of the Convolutional Neural Network on 2,000 test samples (100 per species). Diagonal values indicate correct predictions; off-diagonal values indicate misclassification events.

Siamese Network embedding: distance and similarity matrices

The Euclidean distance matrix (Figure 4) identifies the five closest species pairs in embedding space: *Justicia adhatoda* L. - *Mikania micrantha* Kunth (d = 0.011), *Justicia adhatoda* L. - *Centella asiatica* (L.) Urb. (d = 0.012), *Centella asiatica* (L.) Urb. - *Mikania micrantha* Kunth (d = 0.014), *Centella asiatica* (L.) Urb. - *Calotropis gigantea* (L.) Dryand. (d = 0.018), and *Hibiscus rosa-sinensis* L. - *Ocimum tenuiflorum* L. (d = 0.019). The similarity matrix (Figure 5) shows that 87% of pairs have similarity scores above 0.93, underscoring the challenging nature of visual discrimination within this species set. Pearson correlation between pairwise Euclidean distance and misclassification rate confirmed a significant negative relationship ($r = -0.304$; $p < 0.0001$): pairs closer in embedding space were significantly more likely to be confused by the classifier.

Figure 4 presents pairwise Euclidean distances between class centroids in the Siamese Network embedding space; smaller values indicate greater visual similarity and higher potential for field misidentification. The five closest pairs are: *Justicia adhatoda* L. - *Mikania micrantha* Kunth (d = 0.011), *Justicia adhatoda* L. - *Centella asiatica* (L.) Urb. (d = 0.012), *Centella asiatica* - *Mikania micrantha* (d = 0.014), *Centella asiatica* - *Calotropis gigantea* (L.) Dryand. (d = 0.018), and *Hibiscus rosa-sinensis* L. - *Ocimum tenuiflorum* L. (d = 0.019). Notably, these visually confusable pairs do not correspond to phylogenetic proximity: *Justicia adhatoda* (Acanthaceae) and *Mikania micrantha* (Asteraceae) belong to different families, illustrating convergent leaf morphology across taxonomic boundaries. *Centella asiatica* appears in multiple high-proximity pairs, suggesting it shares a broadly distributed leaf form that presents identification challenges across the dataset.

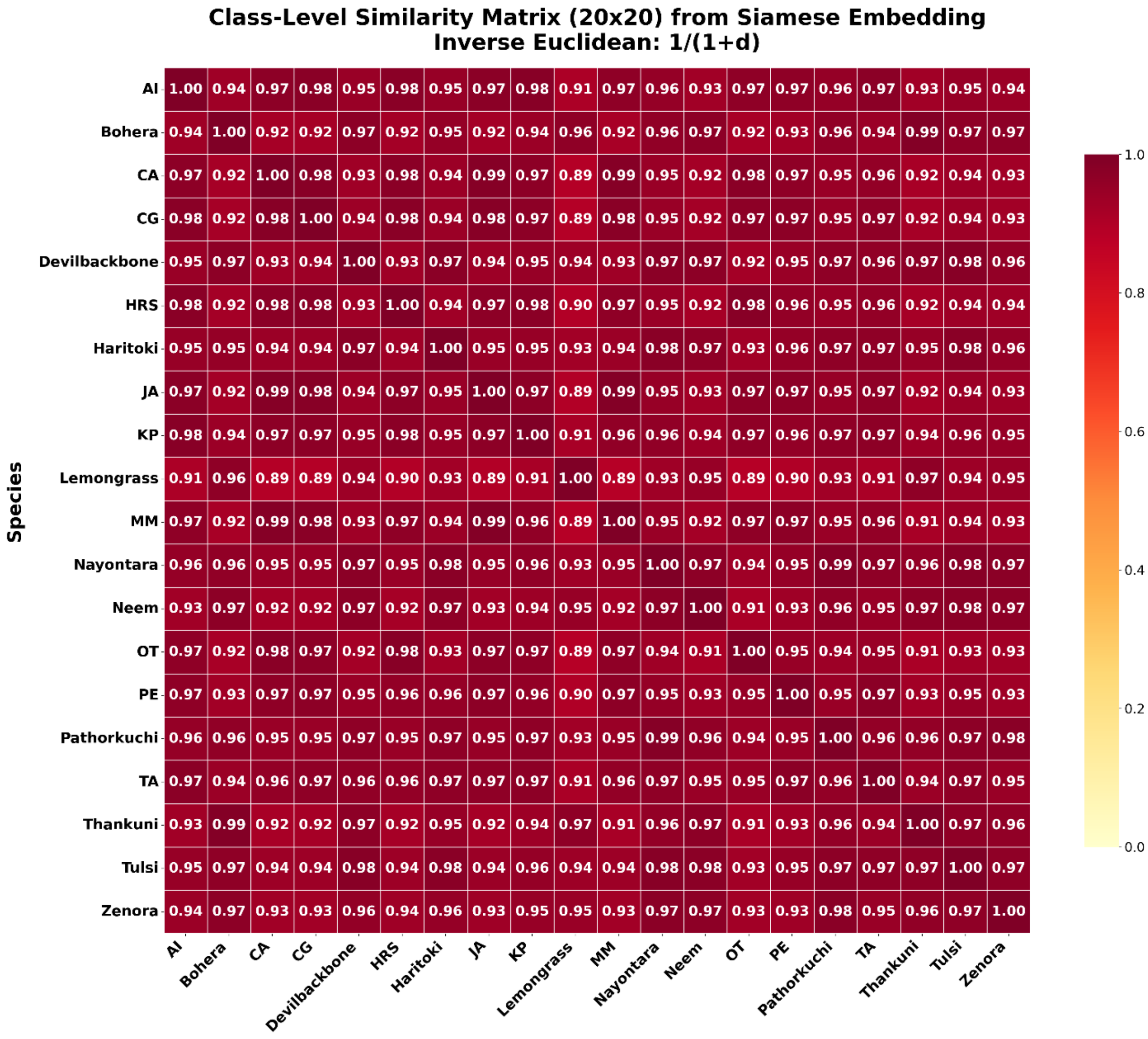

Figure 4. 20×20 Euclidean distance matrix between class centroids in the Siamese Network embedding space. Lighter cells indicate species pairs with lower visual distinctiveness and a higher risk of misidentification.

Figure 5 converts Euclidean distances into similarity scores (S = 1/(1+d)), mapping all 190 pairwise comparisons onto a 0-1 scale where values near 1 indicate near-identical visual appearance. The most striking finding is that 87% of all species pairs exceed a similarity score of 0.93, revealing a broadly challenging dataset for visual discrimination. The darkest cells correspond to the five closest pairs from Figure 4; notably, the *Justicia adhatoda* - *Mikania micrantha* pair approaches S = 0.99. From an ethnobotanical perspective, this result underscores that visual similarity as detected by machine learning does not align with taxonomic or pharmacological classification: two species used for entirely different purposes may look nearly identical in a photograph, while closely related species may appear visually distinct. Identification tools relying on leaf images alone will require supplementary morphological characters for reliable species-level discrimination among these plants.

To validate that the embedding distances reflect genuine visual confusion, Pearson correlation was computed between pairwise embedding distance and the observed CNN misclassification rate. The significant negative result ($r = -0.304$; $p < 0.0001$) confirms that species pairs placed closer in the embedding space were significantly more often confused during classification. For ethnobotanical interpretation, this means Figure 4 is not merely a technical output, it is a statistically validated map of visual confusion risk grounded in the same morphological cues a field practitioner would use when identifying plants by sight.

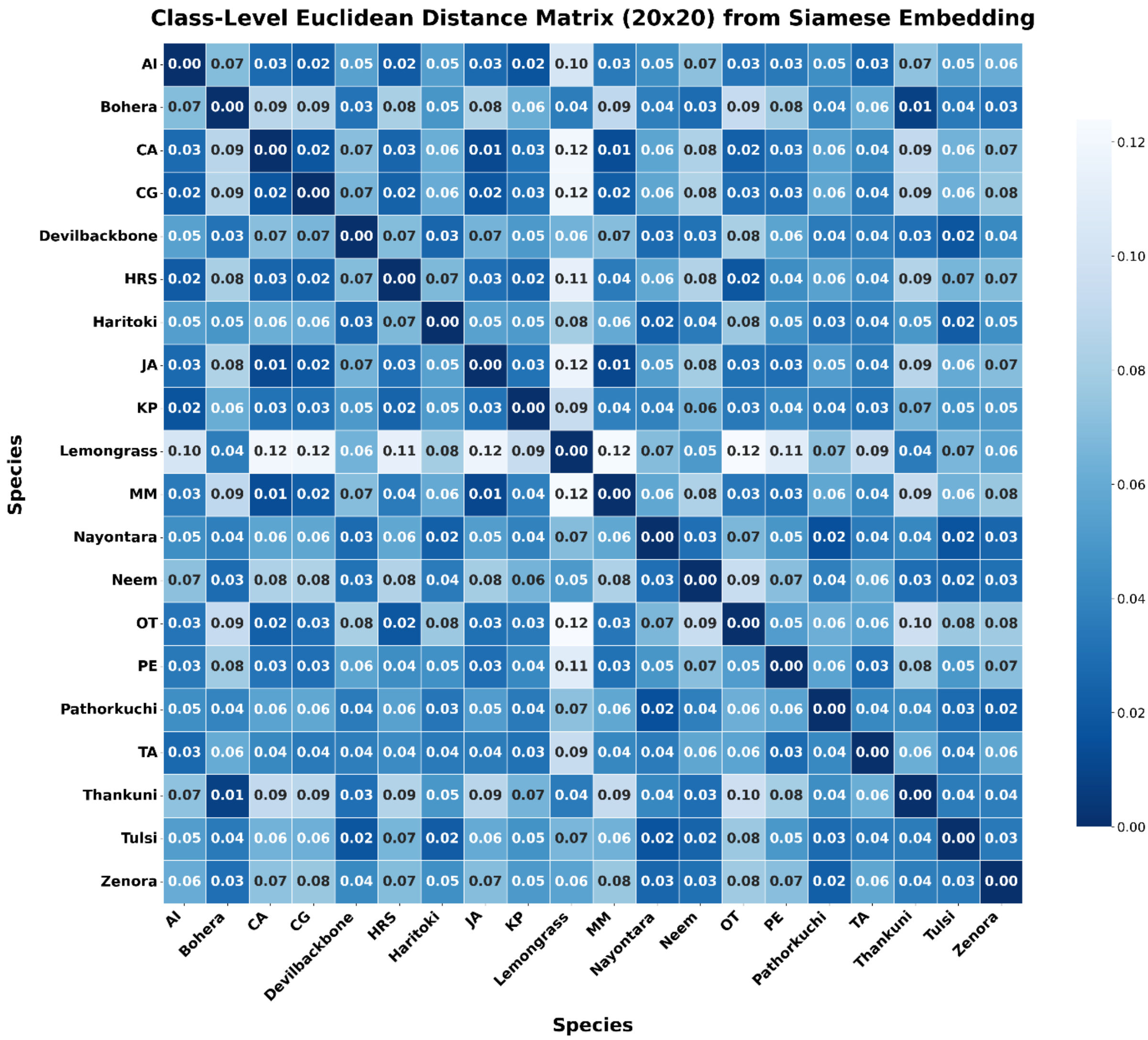

Figure 5. 20×20 similarity matrix between class centroids (score = 1/(1+d)). Darker red indicates greater visual similarity; pairs approaching 1.0 are most prone to confusion between humans and algorithms.

For researchers in ethnobotany, this correlation offers a practical tool: one could use the embedding distance matrix to predict which species pairs are most prone to identification errors before conducting large-scale field trials. Such predictions can guide the selection of additional distinguishing features such as fruit morphology, venation patterns, or even chemical markers to improve identification accuracy.

Specifically for ethnobotanical documentation practice, the distance matrix in Figure 4 could be operationalized as follows. When a medicinal plant survey is conducted in a community where the species in this dataset are present, the five highest-proximity pairs ($d < 0.020$) should be treated as mandatory double-check pairs: whenever a collector identifies a specimen as one of these species, a second morphological character should be recorded to confirm, inflorescence structure, stem cross-section, latex presence, or aroma. Such a protocol requires no laboratory equipment and could be integrated into standard ethnobotanical survey forms. It would reduce the probability of misidentification propagating through a documented dataset or herbal supply chain and would generate valuable comparative data on how often these species are confused in actual field conditions, filling a gap that currently makes herbal adverse event attribution unreliable.

t-SNE visualization of embedding space

The t-SNE projection (Figure 6) reveals that *Cymbopogon citratus* (DC.) Stapf forms the most spatially separated cluster, consistent with its unique elongated strap-like leaf morphology. A dense central overlap region contains *Justicia adhatoda*

L. and *Centella asiatica* (L.). Urb., *Terminalia arjuna* Wight & Arn., *Mikania micrantha* Kunth, and *Ocimum tenuiflorum* L., confirming that these species occupy an indistinct region of feature space. Species forming tight, isolated clusters Pathorkuchi, Bohera, and Neem correspond precisely to those with the highest per-class accuracy.

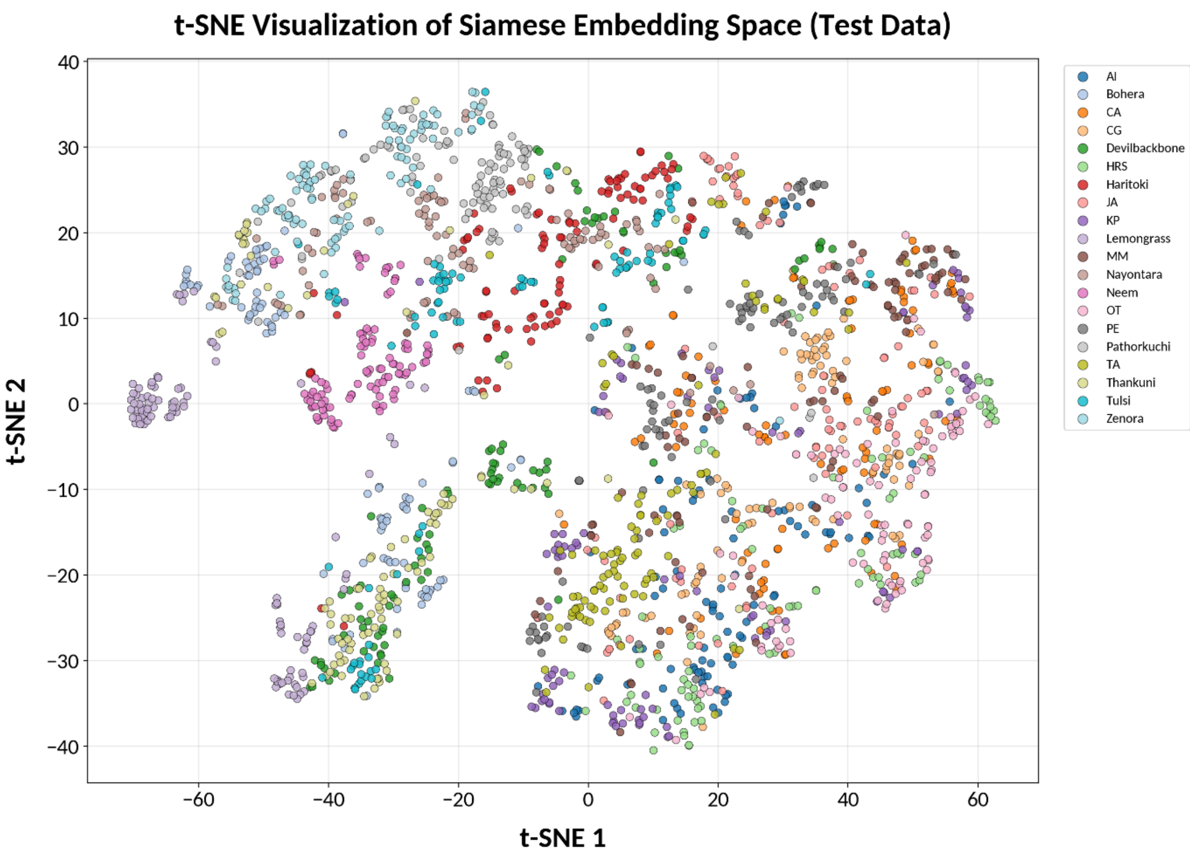

Figure 6. t-SNE visualization of the Siamese Network embedding space on 2,000 test samples. Each point represents one leaf image; colors indicate species identity. Overlapping clusters indicate species prone to visual confusion.

Figure 6 presents a two-dimensional t-SNE projection of the 128-dimensional Siamese Network embedding space. Points close together in this plot share similar visual features as perceived by the model. The plot confirms the distance and similarity matrix findings: most species do not form well-separated clusters, with the majority of points concentrated in a dense central region. The clearest outlier is *Cymbopogon citratus* (lemongrass), whose distinctive strap-like leaf morphology places it far from all other species. Within the dense central cluster, *Justicia adhatoda*, *Centella asiatica*, *Mikania micrantha*, and *Calotropis gigantea* are tightly intermingled, consistent with their low pairwise distances. *Hibiscus rosa-sinensis* and *Ocimum tenuiflorum* share overlapping neighbourhood positions, while species such as *Terminalia bellirica* and *Terminalia chebula* occupy slightly more peripheral positions reflecting their relatively higher classification accuracy. The diffuse cluster structure explains why even an 11-epoch CNN produced measurable misclassification rates and visually validates the use of a Siamese metric learning approach to quantify relative visual proximity among these medicinal species.

Table 2 presents all 21 critical and high-risk species pairs with full botanical names, misclassification counts, Euclidean distances, and similarity scores. Table 3 provides the comprehensive pharmacological risk profiles for all species involved, detailing key bioactive compounds, primary ethnomedicinal uses, known toxicities or contraindications, and the specific consequence of confusion. The following section discusses each pair group in detail.

Table 2. Complete catalog of all 21 critical and high-risk misclassification pairs. Pairs are ordered by risk category and misclassification count within category.

Species pair (predicted as →)	**Errors/100**	**Euclidean dist.**	**Similarity score**	**Category**
Justicia adhatoda L. → *Terminalia arjuna* Wight & Arn.	14	0.036	0.965	Critical

Justicia adhatoda L. → *Centella asiatica* (L.) Urb.	12	0.012	0.988	Critical
Kalanchoe pinnata (Lam.) Pers. → *Terminalia arjuna* Wight & Arn.	11	0.031	0.970	Critical
Kalanchoe pinnata (Lam.) Pers. → *Mikania micrantha* Kunth	10	0.038	0.963	Critical
Centella asiatica (L.) Urb. → *Justicia adhatoda* L.	9	0.012	0.988	Critical
Ocimum tenuiflorum L. → *Centella asiatica* (L.) Urb.	8	0.023	0.977	Critical
Ocimum tenuiflorum L. → *Hibiscus rosa-sinensis* L.	8	0.019	0.982	Critical
Ocimum tenuiflorum L. → *Justicia adhatoda* L.	7	0.031	0.970	Critical
Mikania micrantha Kunth → *Centella asiatica* (L.) Urb.	6	0.014	0.986	Critical
Mikania micrantha Kunth → *Calotropis gigantea* (L.) Dryand.	6	0.022	0.978	Critical
Mikania micrantha Kunth → *Terminalia arjuna* Wight & Arn.	6	0.042	0.960	Critical
Centella asiatica (L.) Urb. → *Ocimum tenuiflorum* L.	6	0.023	0.977	Critical
Centella asiatica (L.) Urb. → *Phyllanthus emblica* L.	6	0.033	0.968	Critical
Justicia adhatoda L. → *Azadirachta indica* A. Juss.	6	0.027	0.973	Critical
Kalanchoe pinnata (Lam.) Pers. → *Centella asiatica* (L.) Urb.	5	0.032	0.969	High
Haritoki *Terminalia chebula* Retz. → Pathorkuchi *Bryophyllum pinnatum* Oken	5	0.033	0.967	High
Mikania micrantha Kunth → *Phyllanthus emblica* L.	5	0.031	0.970	High
Ocimum tenuiflorum L. → *Phyllanthus emblica* L.	5	0.052	0.951	High
Hibiscus rosa-sinensis L. → *Phyllanthus emblica* L.	4	0.045	0.957	High
Devilbackbone *Euphorbia tithymaloides* L. → Thankuni *Centella asiatica* Urb.	4	0.030	0.970	High
Kalanchoe pinnata (Lam.) Pers. → *Ocimum tenuiflorum* L.	4	0.034	0.967	High

Critical (≥6 errors/100); High (4-5 errors/100). Arrow (→) indicates direction of misclassification: true species on left predicted as species on right.

The catalog presented in Table 2 identifies 21 species pairs that pose the greatest risk for visual misclassification by the convolutional neural network. Among these, 14 pairs are classified as critical, defined as those with six or more misclassifications per 100 test samples (red shading). The remaining seven pairs are classified as high-risk, with four to five misclassifications per 100 samples (yellow shading). Notably, the most severe confusion occurs when *Justicia adhatoda* L. is misidentified as *Terminalia arjuna* Wight & Arn., with 14 errors per 100 test images. This is followed by *Justicia adhatoda* misclassified as *Centella asiatica* (L.) Urb. (12 errors), *Kalanchoe pinnata* (Lam.) Pers. as *Terminalia arjuna* (11 errors), and *Kalanchoe pinnata* as *Mikania micrantha* Kunth (10 errors). These critical pairs consistently show very low Euclidean distances (0.012-0.042) and correspondingly high similarity scores (0.960-0.988), confirming that the network's embedding space is statistically associated with the visual overlap that leads to frequent misidentifications; this association has not yet been benchmarked against independent expert-botanist similarity ratings or practitioner confusion data, and such a comparison would be needed before the embedding space could be described as an accurate or validated representation of human-perceived visual overlap.

An examination of the species involved reveals that a small number of plants act as "hubs" of confusion. *Centella asiatica* (L.) Urb., *Justicia adhatoda* L., *Ocimum tenuiflorum* L., *Mikania micrantha* Kunth, and *Kalanchoe pinnata* (Lam.) Pers. appear repeatedly across multiple critical pairs. For example, *Centella asiatica* is involved in six distinct critical or high-risk pairs: confused with *Justicia adhatoda* (12 errors), with *Ocimum tenuiflorum* (6 errors), with *Phyllanthus emblica* L. (6 errors), and with *Kalanchoe pinnata* (5 errors), among others. This pattern suggests that these species share morphological traits such as leaf shape, venation pattern, or growth habit that are particularly difficult for a standard image classifier to separate. From an ethnobotanical perspective, the frequent confusion between *Ocimum tenuiflorum* (holy basil, widely used for respiratory and stress-related ailments) and *Hibiscus rosa-sinensis* L. (used for hair and skin preparations) is especially noteworthy, as these plants serve distinct therapeutic roles despite their visual similarity.

The practical implications of this catalog are straightforward for field ethnobotany and automated plant identification. Any image-based identification tool deployed using this model would require additional safeguards for these 21 high-risk pairs. Possible mitigation strategies include requiring multiple images per plant (e.g., leaf, flower, and bark), incorporating geographic location as an auxiliary input feature, or implementing a human verification step for any prediction involving the hub species identified above. Furthermore, the pharmacological risk profiles in Table 3 (not shown here in full) would allow developers to prioritize which confusions pose the greatest danger to users. For instance, confusing a non-toxic medicinal plant with a species containing cardiac glycosides or hepatotoxic compounds could have serious health consequences. Therefore, this pair catalog serves not only as a performance metric but also as a safety roadmap to improve plant recognition systems for traditional medicine documentation and community-based identification tools. Unless a clinical case report is explicitly cited, the toxicity and pharmacological evidence summarized in Table 3 is drawn primarily from in vitro, animal, and traditional-use literature rather than from confirmed clinical cases of substitution; readers should treat these entries as pharmacologically plausible risk indicators grounded in the cited primary literature, not as documented clinical outcomes, unless the underlying source states otherwise.

Table 3. Comprehensive pharmacological risk profiles for all species implicated in critical and high-risk pairs.

Species	**Primary ethnomedicinal use**	**Key bioactive compounds**	**Risk profile & consequence of misidentification**
Justicia adhatoda L. (Malabar Nut)	Expectorant; bronchodilator; antitubercular; anti-asthmatic; uterotonic (traditional abortifacient)	Vasicine (~0.5-1% dry wt), vasicinone, deoxyvasicine; quinazoline alkaloids (~233 identified phytochemicals)	Vasicine is uterotonic/abortifacient via prostaglandin release; use in pregnancy requires caution given documented uterotonic activity; potential hepatotoxicity at high doses; interacts with respiratory medications (Musa *et al.* 2022)
Centella asiatica (L.) Urb. (Gotu Kola)	Neuroprotective agent; wound healing; anti-anxiety; cognitive enhancer; anti-inflammatory	Asiaticoside, madecassoside, asiaticosides, brahmoside; triterpene saponins up to 8% in leaves	Dose-dependent hepatotoxicity on chronic high-dose use; skin sensitization; caution in hepatic disease. Generally safe at conventional doses (Biswas *et al.* 2021; He *et al.* 2023)
Terminalia arjuna (Roxb.) Wight & Arn.	Cardioprotective; antihypertensive; anti-atherogenic; anti-ischemic; diuretic	Arjunic acid, arjunolic acid, arjunetin, arjunone; flavonoid glycosides; tannins (12-16%); minerals	Herb-drug interactions with antihypertensives and cardiac glycosides; may cause hypotension at high doses; not a substitute for prescribed cardiac therapy (Ramesh & Palaniappan, 2023)
Kalanchoe pinnata (Lam.) Pers. (Pathorkuchi/Air Plant)	Anti-ulcer; immunomodulator; antileishmanial; analgesic; anti-inflammatory; traditional abortifacient	Bufadienolides (bryophyllin A/B structural analogs of cardiac glycosides); quercetin, kaempferol; flavonoids, phenols	Cardiac glycoside-like toxicity (Na^+/K^+-ATPase inhibition) from bufadienolides at elevated doses; dangerous in patients on digoxin; documented toxicity in livestock; NOT for unsupervised cardiac use (Faundes-Gandolfo *et al.* 2024; Hernandez-Caballero *et al.* 2022)
Mikania micrantha Kunth (Bitter Vine)	Antipruritic (folk); antiseptic wound dressing (traditional); antimicrobial; antipyretic (traditional)	Mikanolide, dihydromikanolide, deoxymikanolide; sesquiterpene lactones; germacranolide-type compounds	No established safe oral dosing profile; sesquiterpene lactones are contact allergens and cytotoxic; primarily an invasive weed, systemic oral safety is undocumented; consuming this species instead of a conventional medicinal plant carries undefined but potentially serious risk (Cheng *et al.* 2024)
Ocimum tenuiflorum L. (Holy Basil / Tulsi)	Adaptogen; antimicrobial; anti-inflammatory; antioxidant; immunomodulator; anxiolytic	Eugenol (up to 70% of essential oil), rosmarinic acid, ursolic acid, luteolin, orientin, vicenin	Generally safe at conventional doses; eugenol acts as an antiplatelet agent caution with anticoagulants; some laboratory extracts have shown effects on reproductive parameters in rodent models; eugenol has uterine stimulant properties at high concentrations in

			experimental settings; caution advised in pregnancy (Bhattarai *et al.* 2024; Pradhan *et al.* 2022)
Hibiscus rosa-sinensis L. (Chinese Hibiscus)	Antihypertensive; antidiabetic; anti-inflammatory; emmenagogue; antifertility	Quercetin, kaempferol, gossypetin; cyanidin; anthocyanins; gossypol-related compounds in leaves	Proven anti-implantation and antispermatogenic activity in animal models; considered a potent emmenagogue strictly contraindicated in pregnancy or in those attempting to conceive; interacts with warfarin via flavonoid-CYP pathway modulation (Vasudeva & Sharma, 2008)
Azadirachta indica A. Juss. (Neem)	Antimicrobial; antifungal; antiparasitic; antidiabetic; anti-inflammatory; dermatological	Azadirachtin, nimbin, nimbidin, nimbidol, gedunin, nimbolide; limonoids; quercetin	Hepatotoxic in neonates and young children at high doses; nimbolide cytotoxic; Reye-like syndrome reported in infants; not appropriate as a substitute for other species consumed by vulnerable populations (Ekor, 2014)
Calotropis gigantea (L.) Dryand. (Crown Flower)	Traditional antiseptic; anti-inflammatory; antileprotic; some traditional antipyretic use; latex as topical agent	Calotropin, calotoxin, calactin, uscharin (cardenolide cardiac glycosides; high experimental cardiotoxic potency); resin glycosides; taraxeryl acetate	Cardenolides exhibit high cardiotoxic potency in experimental studies, acting on Na^+/K^+-ATPase; consumption at toxic doses may cause severe bradycardia and ventricular arrhythmia in animal models. Latex is an irritant to mucosae and skin. Inadvertent consumption in place of a conventional medicinal herb poses serious safety risks (Kyei *et al.* 2022)
Phyllanthus emblica L. (Amla / Indian Gooseberry)	Antioxidant; immunomodulatory; hepatoprotective; anti-aging; anti-diabetic; anti-inflammatory	Emblicanin A, emblicanin B; gallic acid, ellagic acid, chebulic acid; ascorbic acid (highest natural concentration); tannins	Relatively safe at conventional doses; high oxalic acid content may contribute to renal calculi with excessive long-term consumption; mildly hypoglycemic; adjust in patients on antidiabetics (Prananda *et al.* 2023)
Haritoki *Terminalia chebula* Retz. (Chebulic Myrobalan)	Laxative; antioxidant; antimicrobial; anti-inflammatory; component of Triphala formulation	Chebulic acid, chebulinic acid, corilagin; tannins (30-40%); gallic acid, ellagic acid; terchebin	Generally safe; excessive tannin intake causes gastrointestinal constipation and nausea; high tannin loads may reduce iron absorption; safe profile compared to *Bryophyllum pinnatum*, which it is confused with (Kyei *et al.* 2022)
Pathorkuchi *Bryophyllum pinnatum* (Lam.) Oken [syn. *K. pinnata*]	Anti-ulcer; kidney stone dissolution (traditional); antimicrobial; anti-inflammatory; antifertility	Bufadienolides (bryophyllin A bryotoxin; bryophyllin B); flavonoids; quercetin; malic acid; amino acids	Same bufadienolide toxicity profile as *Kalanchoe pinnata* (these are synonymous species); cardiac glycoside-like risks at elevated doses; antifertility activity documented; not for unsupervised use (Faundes-Gandolfo *et al.* 2024)
Devilbackbone *Euphorbia tithymaloides* L.	Traditional wound antiseptic (topical); dermatological applications; some traditional analgesic use	Diterpene esters (5-deoxyingenol, phorbol ester analogs); ingenane-type diterpenes; rubber compounds; resins; latex	Potentially toxic if consumed internally: phorbol ester-type diterpenes are documented mucosal and gastrointestinal irritants in experimental studies; tumor-promoting co-carcinogens in animal models; latex causes dermatitis and eye injury. Inadvertent consumption in place of *Centella asiatica* may cause gastrointestinal toxicity proportional to dose (Ekor, 2014; Kyei *et al.* 2022)

Discussion

Visual similarity as a quantifiable predictor of misidentification risk

The negative Pearson correlation between Euclidean embedding distance and misclassification rate ($r = -0.304$; $p < 0.0001$) indicates that machine-learning feature proximity is a statistically valid proxy for morphological confusability from a human observer's perspective. This finding has a practical corollary: species pairs that occupy adjacent regions of the Convolutional Neural Network feature space designed to extract the most discriminative visual features a deep neural network can find are those that would most likely confuse a practitioner relying on visual inspection alone, particularly under suboptimal field conditions (dried specimens, variable illumination, damaged leaf margins, intra-species ontogenetic variation). This validates the use of machine learning embedding analysis as an ethnobotanical risk assessment tool, complementing rather than replacing expert botanical judgment (Lankasena *et al.* 2024; Mulugeta *et al.* 2024).

The 88.45% overall test accuracy on 20 species with 100 samples each is broadly consistent with the literature on Convolutional Neural Network performance for medicinal plant leaf classification under multi-class conditions (Al-Qurran *et al.* 2022; Bouakkaz *et al.* 2025; Deshmukh, 2024). However, the distribution of errors is highly non-uniform, confirming that overall accuracy figures mask clinically significant variation at the pair level. This point is critical for ethnobotanical applications: a model that is 90% accurate but fails systematically on the specific pairs with the most dangerous pharmacological divergence offers less safety value than the accuracy figure suggests.

This point deserves elaboration for an ethnobotanical audience unfamiliar with machine learning benchmarking. Traditional medicine safety research has historically focused on the intrinsic toxicity of individual plant species, examining dose-response relationships and adverse event profiles in isolation. What this study demonstrates is that a parallel and largely invisible risk axis exists: the probability that the species being consumed is not the species the practitioner or consumer intended. This misidentification axis is structurally invisible to conventional pharmacovigilance because adverse event reports record only the species name the practitioner believed they were using, not the species that was actually consumed. A systematic approach to misidentification risk, grounded in quantitative visual similarity analysis as presented here, offers a framework for making this invisible risk axis visible and actionable within ethnobotanical safety research.

The *Justicia adhatoda* L. confusion cluster - uterotonic alkaloid risk

Justicia adhatoda L. (family Acanthaceae), known locally as "basak" in Bangladesh and "malabar nut" or "adusa" across South Asia, is among the most extensively used respiratory medicinal plants in the region, employed for the treatment of cough, asthma, bronchitis, and tuberculosis (Musa *et al.* 2022). Its remarkable therapeutic value is well-supported pharmacologically: vasicine (peganine), the principal quinazoline alkaloid constituting approximately 0.5-1% of dry leaf weight, is the precursor of the widely prescribed semi-synthetic mucolytic agent bromhexine (Musa *et al.* 2022). However, vasicine also carries the most critical contraindication among compounds in this species set: documented uterotonic and abortifacient activity in experimental studies, mediated by stimulation of prostaglandin biosynthesis in uterine tissue. Studies on myometrial preparations and animal models indicate that vasicine induces uterine contractions, and dose-dependent effects have been reported in preclinical models (Musa *et al.* 2022). The uterotonic properties of vasicine have been studied in preclinical contexts, further underscoring the reproductive safety concern.

The pair *Justicia adhatoda* L. - *Terminalia arjuna* (Roxb. ex DC.) Wight & Arn. was the most error-prone in the entire dataset (14/100 test samples of *Justicia adhatoda* L. predicted as *Terminalia arjuna*). These two species share superficially similar leaf morphology: both possess lanceolate-elliptic laminae, prominently pinnate venation, smooth or slightly wavy margins, glossy adaxial surface, and medium-to-large size (15-25 cm length). Their Euclidean embedding distance is 0.036, which is moderate but within the critical range. The pharmacological consequence of this confusion operates in two directions. A patient using *Terminalia arjuna* (Roxb. ex DC.) Wight & Arn. for evidence-supported cardiovascular protection, this species contains arjunic acid, arjunone, and multiple flavonoid glycosides with proven cardiotonic and anti-atherogenic activity (Bhumika *et al.* 2022; Ramesh & Palaniappan, 2023) who inadvertently consumes *Justicia adhatoda* L. instead will receive none of the desired cardiac benefit. More critically, a pregnant individual who consumes *Justicia adhatoda* L. believing it to be a safer species faces risk of uterotonic complications including preterm labor or miscarriage through vasicine-mediated prostaglandin release (Musa *et al.* 2022).

The bidirectional confusion between *Justicia adhatoda* L. and *Centella asiatica* (L.) Urb. (12/100 samples of *Justicia adhatoda* misclassified as *Centella asiatica*; 9/100 of *Centella asiatica* misclassified as *Justicia adhatoda*) carries distinct but equally serious consequences for each direction. *Centella asiatica* (L.) Urb. known as "thankuni" in Bangladesh and "gotu kola" internationally, is one of the safest and most widely consumed medicinal plants in Asia, used as both food and medicine for wound healing, cognitive enhancement, anti-anxiety, and neuroprotection. Its principal bioactive compounds, the triterpene saponins asiaticoside and madecassoside, have well-documented neurotropic and wound-healing mechanisms (Biswas *et al.* 2021; He *et al.* 2023). A pregnant woman or person with respiratory illness who intentionally consumes *Centella asiatica* (L.) Urb. but receives *Justicia adhatoda* L. due to an identification error is exposed to vasicine without any expectation of its uterotonic risk, with no warning signs before consumption. This scenario is compounded by the extremely low Euclidean distance between the two species (d = 0.012, the second-smallest of all 190 pairs), confirming that these leaves are among the most visually indistinguishable pairs in the entire dataset.

The confusion of *Justicia adhatoda* L. with *Azadirachta indica* A. Juss. (6/100 errors), while a smaller proportion, is notable because *Azadirachta indica* A. Juss. (neem) is commonly consumed in Bangladesh and South Asia as a health tonic, antimicrobial, and antidiabetic agent. Substitution with *Justicia adhatoda* L., which has a distinct alkaloid profile and documented dose-dependent hepatotoxicity, again poses risk to those unaware of the substitution (Ekor, 2014).

The *Kalanchoe pinnata* (Lam.) Pers. confusion cluster - cardiac glycoside risk

Kalanchoe pinnata (Lam.) Pers. (synonym: *Bryophyllum pinnatum* (Lam.) Oken; family Crassulaceae) is known as "pathorkuchi" in Bangladesh and is used widely across South Asia for kidney stone dissolution, wound healing, and anti-inflammatory purposes. Botanists reading this study should note the taxonomic synonymy: the Pathorkuchi class from Dataset 1 (*Bryophyllum pinnatum* (Lam.) Oken) and the KP class from Dataset 2 (*Kalanchoe pinnata* (Lam.) Pers.) represent the same species captured under different imaging protocols, treated as distinct classification targets herein.

The defining pharmacological concern of this species is its content of bufadienolides, specifically bryophyllin A (also termed bryotoxin) and bryophyllin B, which are steroidal compounds structurally analogous to cardiac glycosides in the digitalis family. Like digoxin, bufadienolides inhibit the Na^+/K^+-ATPase pump in cardiac myocytes, potentially causing bradycardia, AV block, and ventricular arrhythmia at elevated doses (Hernández-Caballero *et al.* 2022; Faundes-Gandolfo *et al.* 2024). While *Kalanchoe pinnata* (Lam.) Pers. is not universally classified as acutely toxic at conventional traditional use doses, cardiac glycoside accumulation is dose-dependent and particularly dangerous in patients already receiving digitalis-based cardiac medications, where additive glycoside effects can precipitate toxicity at otherwise sub-toxic doses.

The pair *Kalanchoe pinnata* (Lam.) Pers. - *Terminalia arjuna* (Roxb. ex DC.) Wight & Arn. (11/100 errors) is particularly alarming because both species are used in cardiovascular contexts. A patient taking *Terminalia arjuna* (Roxb. ex DC.) Wight & Arn. as an herbal adjunct for heart failure, hypertension, or ischaemia use supported by clinical and pre-clinical evidence (Ramesh & Palaniappan 2023) who inadvertently receives *Kalanchoe pinnata* (Lam.) Pers. instead, is exposed to bufadienolide cardiac glycosides. In a patient simultaneously prescribed conventional cardiac medications (digoxin, beta-blockers, ACE inhibitors), this substitution could produce additive or synergistic cardiotoxic effects that are indistinguishable from drug-drug interactions without knowledge of the botanical substitution.

The confusion of *Kalanchoe pinnata* (Lam.) Pers. with *Mikania micrantha* Kunth (10/100 errors) involves a species substitution of a completely different character. *Mikania micrantha* Kunth (family Asteraceae) is classified globally as one of the world's worst invasive weeds(Rai *et al.* 2022); its traditional uses are primarily in the communities of origin in tropical South America (antipyretic, antimicrobial wound dressing, antipruritic) and it has limited ethnomedical documentation as an orally consumed plant in South Asian contexts (Cheng *et al.* 2024). Its principal bioactive compounds are sesquiterpene lactones (mikanolide, dihydromikanolide), which are documented contact allergens and have demonstrated in vitro cytotoxicity. A person intending to consume *Kalanchoe pinnata* (Lam.) Pers.) for its traditional uses who receives *Mikania micrantha* Kunth encounters a plant with essentially no established oral safety profile, with compounds of uncertain systemic toxicity a definitionally hazardous substitution.

The *Ocimum tenuiflorum* L. confusion cluster - antifertility and uterotonic risk

Ocimum tenuiflorum L. (Holy Basil / Tulsi, family Lamiaceae) is perhaps the best-known and most widely trusted medicinal plant in South and Southeast Asian traditions. Consumed daily as a herbal tea or tonic in hundreds of millions of households, it carries a perception of absolute safety that may paradoxically exacerbate risk when it is mistaken for more dangerous species. The pharmacological concern specific to *Ocimum tenuiflorum* L. itself is relatively minor at conventional doses: eugenol (its principal constituent at up to 70% of essential oil) has antiplatelet activity that may prolong bleeding time in patients on anticoagulant therapy (Bhattarai *et al.* 2024). More critically, the confusion identified by our model runs in the opposite direction: *Ocimum tenuiflorum* L. is predicted to be a species that poses substantially greater pharmacological hazard.

The confusion of *Ocimum tenuiflorum* L. with *Hibiscus rosa-sinensis* L. (8/100 errors) represents one of the most consequential pairs in the dataset from a reproductive health perspective. *Hibiscus rosa-sinensis* L. possesses well-documented antifertility activity: flower extracts have produced near-complete inhibition of post-coital implantation in rodent models at effective doses, and root extracts demonstrate anti-implantation and uterotropic effects in animal studies (Vasudeva & Sharma 2008). The active antifertility compounds include quercetin, kaempferol, gossypetin, and cyanidin-class anthocyanins that modulate estrogenic pathways and endometrial receptivity. A pregnant woman or one attempting to conceive who habitually consumes what she believes to be her daily *Ocimum tenuiflorum* L. tonic but has been supplied with *Hibiscus rosa-sinensis* L. leaf through identification error, faces cumulative antifertility compound exposure without any awareness of the substitution. The embedding distance between these species is only 0.019, the fifth-smallest pair overall, confirming that visual features alone provide insufficient discrimination.

The confusion of *Ocimum tenuiflorum* L. with *Justicia adhatoda* L. (7/100 errors) reintroduces the uterotonic vasicine risk described above. Still, in an even more insidious context: *Ocimum tenuiflorum* L. is consumed daily and is considered universally safe, meaning a consumer would have no reason to suspect pharmacological harm. The confusion with *Centella asiatica* (L.) Urb. (8/100 errors) involves substitution of the adaptogen with a neurotropic/wound-healing species. At the same time, both are relatively safe; therapeutic failure occurs when the expected adaptogenic benefit of *Ocimum tenuiflorum* L. is not received, which is particularly concerning in contexts where the plant is being used to manage chronic stress or anxiety (Hasan *et al.* 2023; Lopresti *et al.* 2022; Muszalska & Wieckanowska, 2024; Pradhan *et al.* 2022).

The *Mikania micrantha* Kunth confusion cluster - cardiotoxic and undocumented risk

The misclassification of *Mikania micrantha* Kunth as *Calotropis gigantea* (L.) Dryand. (6/100 errors) requires special emphasis because it represents the most acutely toxic substitution in this dataset (Sivaraman *et al.* 2023). *Calotropis gigantea* (L.) Dryand. (family Apocynaceae) contains a suite of cardenolide cardiac glycosides, calotropin, calotoxin, calactin, and uscharin, reported to exhibit high cardiotoxic potency in experimental studies, comparable to or exceeding that of digitalis derivatives. These compounds inhibit the cardiac Na^+/K^+-ATPase with documented binding affinity, and consumption at toxic doses may produce bradycardia, atrioventricular block, and ventricular arrhythmia in animal models. The latex of this species additionally causes mucosal and gastrointestinal irritation. In the traditional medicine context, *Calotropis gigantea* (L.) Dryand. is used primarily for topical applications (antiseptic, antileprotic) rather than oral consumption; however, the identification error risk operates in reverse: a person intending to collect an oral herbal medicine could collect *Calotropis gigantea* (L.) Dryand. believing it to be a benign leafy medicinal species.

The further confusions of *Mikania micrantha* Kunth with *Centella asiatica* (L.) Urb. (6/100) and *Terminalia arjuna* Wight & Arn. (6/100) follow the same principle: *Mikania micrantha* Kunth has no established safe oral consumption profile, so its substitution for conventionally consumed species introduces uncharacterized phytochemical exposure. The sesquiterpene lactones in *Mikania micrantha* Kunth (mikanolide, dihydromikanolide) have demonstrated cytotoxicity and are classified as contact sensitizers in occupational exposure assessments (Cheng *et al.* 2024).

High-risk pairs - targeted guidance needs

Among the seven high-risk pairs (4-5 errors per 100), two warrant detailed discussion for their pharmacological implications. The confusion of *Terminalia chebula* Retz. (Haritoki) with *Bryophyllum pinnatum* (Lam.) Oken (Pathorkuchi; 5/100 errors) brings together two species with fundamentally different safety profiles. *Terminalia chebula* Retz. is a tannin-rich tree fruit component of the classical Triphala formulation, long regarded as one of the safest herbal preparations in Ayurveda. Its leaf is essentially non-toxic at conventional use levels. *Bryophyllum pinnatum* (Lam.) Oken, as discussed above, contains bufadienolides. The substitution of a safe tannin-rich species with a cardiac glycoside-class plant in preparations intended for elderly patients with gastrointestinal complaints a primary use-case for Triphala introduces cardiac risk in a population already vulnerable to cardiovascular complications.

The confusion of *Euphorbia tithymaloides* L. (Devilbackbone) with *Centella asiatica* (L.) Urb. (Thankuni; 4/100 errors) is among the most pharmacologically extreme substitutions in the high-risk category. *Euphorbia tithymaloides* L. (family Euphorbiaceae) contains diterpene esters of the ingenane and phorbol ester structural classes with documented irritant and co-carcinogenic properties in experimental studies. Phorbol ester-type diterpenes have been shown to cause gastrointestinal mucosal irritation, induce inflammatory responses, and act as co-carcinogens through protein kinase C activation in laboratory models. The substitution of this species for *Centella asiatica* (L.) Urb. a plant frequently consumed in large quantities as a food vegetable in salads and herbal teas could produce gastrointestinal toxicity proportional to dose consumed ((Ekor, 2014; Kyei *et al.* 2022). Remarkably, the embedding distance between these species was only 0.030, suggesting that despite their entirely different family origins, their leaf morphology generates very similar feature-space representations.

Cross-cutting patterns: implications for ethnobotanical practice and documentation

Several cross-cutting observations emerge from the analysis of all 21 pairs. First, the three highest-risk species in terms of appearing as the "wrong" target in misidentification events are *Centella asiatica* (L.) Urb. (appearing as a misclassification destination in 5 critical pairs), *Terminalia arjuna* Wight & Arn. (3 critical pairs), and *Phyllanthus emblica* L. (3 high pairs). These three species are all widely trusted and broadly consumed, which explains why they serve as substitution attractors: practitioners who encounter an unknown leaf may unconsciously default to a familiar, trusted species, and their high visual similarity to more dangerous neighbors reinforce this cognitive misattribution.

Second, the confusion cluster centered on *Justicia adhatoda* L. - *Centella asiatica* (L.) Urb. - *Mikania micrantha* Kunth (d = 0.011-0.014, the three smallest pairwise distances in the dataset) suggests that these three phylogenetically diverse species have converged on a leaf morphological archetype smooth-margined, oval-elliptic laminae of medium size with pinnate venation that cannot be reliably distinguished from leaf image alone by either human observers or current state-of-the-art deep learning models. This finding argues for the development of multi-organ identification protocols that incorporate additional discriminating features: the strongly aromatic scent of *Justicia adhatoda* L., the umbellate inflorescence of *Centella asiatica* (L.) Urb., and the invasive climbing habit of *Mikania micrantha* Kunth are all features invisible from a single leaf image but immediately apparent in the field.

Third, the geographic universality of the risk should be emphasized. All critical species in this study are distributed well beyond Bangladesh: it is the underlying species pairs, not the specific risk magnitudes reported here, that are geographically transferable: actual risk in another setting also depends on local co-occurrence, vernacular naming, harvesting and preparation practices, and practitioner knowledge, so the framework should be considered transferable while the specific risk map requires local validation wherever it is applied. *Justicia adhatoda* L. is native to the Indian subcontinent and is naturalized across tropical Asia and Africa; *Centella asiatica* (L.) Urb. is pantropical; *Ocimum tenuiflorum* L. is cultivated throughout South and Southeast Asia and Africa; *Kalanchoe pinnata* (Lam.) Pers. is cosmopolitan in the tropics; and *Calotropis gigantea* (L.) Dryand. occurs throughout South and Southeast Asia. The risk assessment produced here therefore has applicability beyond the Bangladeshi context, wherever these species co-occur and where morphological identification without chemical verification is practiced (Islam & Uddin, 2022; Uddin & Zidorn, 2020).

The computational findings reported here should be understood primarily as inputs to ethnobotanical practice and documentation, rather than as stand-alone technological outputs. The 21 high-risk and critical species pairs constitute an evidence-based priority list for the development of targeted field guides: visual comparison plates highlighting the distinguishing morphological features of each pair (leaf margin detail, venation density, petiole structure, growth habit) would directly address the identification challenge quantified by the model. Similarly, practitioners' training programmes in community-based health contexts could prioritize these specific pairs in morphological identification exercises. The embedding distance matrix itself, translated into accessible visual formats, could serve as a practical tool for ethnobotanical fieldwork planning and curriculum design for traditional medicine practitioners in Bangladesh and across the co-occurrence range of these species.

A further application that the present findings suggest is the integration of this risk information into traditional knowledge documentation protocols. Ethnobotanical surveys routinely record species names, uses, and preparation methods; the addition of a "confusion species" field, documenting which other species a practitioner acknowledges as visually similar or occasionally confused, would generate community-level data on misidentification risk that could be cross-referenced with the computational pairs identified here. Where community-acknowledged confusion pairs overlap with computationally identified high-risk pairs, the risk signal is reinforced; where they diverge, the divergence itself is informative, suggesting that local ecological or morphological knowledge provides disambiguation cues not captured in the leaf image alone. Such an integrated approach would align computational ethnobotany with community-based participatory research methods that are increasingly central to responsible ethnobotanical practice.

There is also potential value in connecting these computational findings to the growing body of ethnopharmacological research documenting adverse events from herbal medicines in South Asian clinical settings. Case reports of unexplained hepatotoxicity, uterotonic complications, or cardiac irregularities following herbal remedy consumption rarely record the botanical identity of the plant consumed beyond the vernacular name provided by the patient. The species pairs identified here provide a hypothesis-generating framework: when a patient presents with symptoms consistent with vasicine exposure after consuming what they describe as a respiratory herb, the *Justicia adhatoda - Centella asiatica* confusion pair identified in this study is a biologically plausible mechanism. Systematic documentation of vernacular name usage alongside verified botanical identity in clinical settings, informed by high-risk pair lists such as the one produced here, would strengthen the evidence base for herbal medicine pharmacovigilance in ways that laboratory toxicology alone cannot achieve.

Implications for traditional knowledge documentation and pharmacovigilance

The findings of this study have direct and practical implications for how ethnobotanical surveys are conducted and how their data are interpreted. Current best-practice ethnobotanical survey methods in South Asia document species by vernacular name, primary use, preparation method, and collector identity, but rarely include a systematic record of the morphological features used to distinguish the collected specimen from other visually similar species in the same habitat. The 21 high-risk pairs identified here provide an evidence base for the development of differentiation checklists: simple, field-usable instruments that guide practitioners and collectors to record one or two key discriminating characters when handling species known to be prone to confusion. For the *Justicia adhatoda - Centella asiatica* pair, the relevant discriminating character is growth habit: *Justicia adhatoda* is an erect shrub with opposite lanceolate leaves on a woody stem, while *Centella asiatica* is a creeping herb with rounded leaves on slender petioles arising from stolons at ground level. These differences are immediately apparent from a whole-plant view but disappear when only a detached leaf is presented. Documenting this distinction in field protocols and training materials directly addresses the identification context in which confusion is most likely to occur.

For the *Kalanchoe pinnata - Terminalia arjuna* pair, discrimination relies on scale and growth form: *Terminalia arjuna* is a large riparian tree with bark that is grey-green and exfoliating, while *Kalanchoe pinnata* is a succulent herb of rocky or disturbed ground with fleshy notched leaves that produce plantlets along their margins. Again, the whole-plant view eliminates confusion, but isolated leaf photographs, the format in which identification errors most often propagate through markets and digital identification tools, present a genuinely ambiguous morphological signal. For the *Ocimum tenuiflorum - Hibiscus rosa-sinensis* pair, the discriminating character available from leaf image is margin complexity: Hibiscus leaves typically show coarser, more irregular serration and a broader ovate form compared to the slightly more lanceolate, finely toothed Ocimum leaf. Incorporating these specific differential characters into field identification training for the high-risk pairs identified in this study would constitute a targeted, evidence-based safety intervention with minimal cost and immediate applicability.

At the pharmacovigilance level, this study argues for a structural change in how adverse events attributed to herbal medicines are recorded and investigated. Currently, when a patient presents with an adverse event following consumption of a herbal remedy, the species identity of the plant consumed is rarely verified; the report is filed under the vernacular name provided by the patient, which may itself reflect a misidentification. The WHO Uppsala Monitoring Centre database and national pharmacovigilance systems therefore systematically undercount the contribution of plant misidentification to herbal adverse events, attributing outcomes to the "intended" species rather than the "consumed" species. The pair catalog produced in this study could serve as a hypothesis-generating reference for clinical pharmacovigilance investigators: when an adverse event attributed to a low-risk species produces a clinical profile consistent with a higher-risk species in the same confusion cluster (for example, uterotonic effects in a patient who consumed a plant identified as *Centella asiatica*), the misidentification hypothesis should be investigated rather than discarded. Systematic specimen collection from patients at the point of adverse event reporting, combined with DNA barcoding or chromatographic verification of species identity, would transform the evidence base for herbal medicine safety in communities where the risk pairs identified here are co-distributed.

The machine learning approach applied here has important limitations that ethnobotanists should factor into any translation of these findings to practice. The models were trained on standardized, single-leaf images under controlled photographic conditions; real-world field identification involves partial views, overlapping foliage, different growth stages, seasonal color variation, and lighting conditions ranging from forest shade to direct tropical sunlight. Transfer of model accuracy to field deployment without extensive real-world validation would be premature (Lankasena *et al.* 2024). Furthermore, no model, however accurate, can replace the holistic multi-sensory botanical expertise of a trained pharmacognosist or ethnobotanist. The appropriate role of these tools is as a field verification aid, a second-check mechanism, not as a standalone identification authority. Smartphone-based deployment of Convolutional Neural Network classifiers with explicit uncertainty messaging for high-risk pairs offers a practical near-term application pathway (Mulugeta *et al.* 2024).

Conclusion

This study provides the first systematic machine learning-based misidentification risk atlas for 20 medicinal plant species, combining Convolutional Neural Network classification error analysis with Siamese Network embedding similarity quantification, and linking computational findings to structured pharmacological risk profiles. The 14 critical and 7 high-risk pairs identified are not pharmacologically equivalent exchanges: they involve species containing uterotonic alkaloids, cardiac glycoside-class bufadienolides and cardenolides, phorbol ester-type diterpene irritants, and antifertility flavonoids. The consequences of misidentification in these cases span from therapeutic failure to acute toxicity, miscarriage, or severe cardiac arrhythmia in susceptible individuals.

The three pairs posing the highest combined visual-confusion and pharmacological-incompatibility risk are: (1) *Justicia adhatoda* L. confused with *Centella asiatica* (L.) Urb. and *Terminalia arjuna* Wight & Arn., primarily threatening to pregnant individuals and cardiac patients; (2) *Kalanchoe pinnata* (Lam.) Pers. confused with *Terminalia arjuna* Wight & Arn. and *Mikania micrantha* Kunth, threatening cardiac patients and introducing undefined oral toxicity risk; and (3) *Ocimum tenuiflorum* L. confused with *Hibiscus rosa-sinensis* L., threatening reproductive health of women using tulsi as a daily wellness tonic.

From the perspective of ethnobotany as a discipline, this study demonstrates that machine learning error analysis is not merely an engineering tool but a novel form of ethnobotanical evidence. The systematic confusion patterns identified by a deep learning model trained on leaf images reveal the botanical ambiguities that exist at the interface between morphological form and pharmacological function, precisely the interface that traditional medicine practitioners navigate daily without the benefit of chemical analysis or laboratory confirmation. The 21 risk pairs identified here represent, in quantitative terms, the most vulnerable nodes in the identification chain linking medicinal plant collection to patient consumption. Addressing these nodes through targeted botanical training, improved field guides, and community-based verification protocols represents a practical, low-cost, and immediately implementable strategy for reducing the human cost of plant misidentification in traditional medicine systems.

The evidence base produced here should serve as a foundation for: (i) development of field identification guides specifically designed to highlight discriminating morphological features between the 21 identified risk pairs; (ii) targeted training programs for traditional medicine practitioners prioritizing correct identification of species in high-risk confusion clusters; (iii) smartphone-based AI identification tools with explicit risk warnings for critical pairs; and (iv) pharmacovigilance initiatives that explicitly account for species misidentification as a confounding variable in adverse event reporting for herbal medicines. The Siamese Network embedding framework described is directly applicable to any multi-species medicinal flora worldwide, offering a scalable methodology for ethnobotanical safety assessment wherever quantitative plant image data are available.

Declarations

List of abbreviations: All species names are written in full throughout the manuscript to avoid ambiguity. Technical terms: Convolutional Neural Network: multi-layer feature-extraction architecture for image classification; Siamese Network: shared-weight neural network for similarity learning; Triplet loss: training objective minimizing anchor-positive distance while maximizing anchor-negative distance; t-SNE: t-distributed Stochastic Neighbor Embedding; Euclidean distance: L2 norm between embedding vectors.

Ethics approval and consent to participate: This study used two publicly available, anonymized plant image datasets. No human or animal subjects were involved. Ethics approval was not required. Access and Benefit-Sharing (Nagoya Protocol) statement: this study did not involve new collection of genetic resources or plant material; it relied exclusively on secondary, previously published leaf-image datasets (Borkatulla et al. 2023; Ahmed et al. 2023) released by their original authors under open data-sharing terms. No unpublished traditional knowledge was accessed, and no direct engagement with knowledge-holding communities took place. The dataset citations in this manuscript constitute the attribution owed to the original data providers; no additional benefit-sharing obligations were triggered under this study design.

Consent for publication: Not applicable.

Availability of data and materials: Bangladeshi Medicinal Plant Dataset) is available at Kaggle (doi: 10.1016/j.dib.2023.109211). The trained model weights, the exact train/validation/test split indices, the random seeds used for training and triplet mining, and the risk-table generation scripts are deposited in a public Zenodo repository and are citable by permanent DOI, rather than being available only on request.

Competing interests: The authors declare no conflicts of interest.

Funding: This research was funded by the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan / LPDP), Ministry of Finance, Republic of Indonesia.

Authors' contributions: I Gusti Ngurah Sentana Putra designed the study, built and trained computational models, and produced all figures and tables. Muhammad Syafiq conducted the pharmacological literature review and risk profile compilation. Utami Dyah Syafitri supervised the study and approved the final version. All authors read and approved the final manuscript.

Acknowledgements

This research was supported by graduate scholarship funding from the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan/LPDP) for I Gusti Ngurah Sentana Putra (NIB: 202404110201596) and Muhammad Syafiq (NIB: 20240711020517). The authors sincerely thank the School of Data Science, Mathematics, and Informatics (SSMI), IPB University, for providing institutional facilities and technical support, and the curation team of the Bangladesh Medicinal Plants Dataset for providing open-access image data.

Literature cited

Ahmed MR, Fahim MAI, Islam AKMM, Islam S, Shatabda S. 2023. BDMediLeaves: A leaf images dataset for Bangladeshi medicinal plant identification. Data in Brief 50:109488. doi: 10.1016/j.dib.2023.109488

Al-Qurran R, Al-Ayyoub M, Shatnawi A. 2022. Plant classification in the wild: energy evaluation for deep learning models. Multimedia Tools and Applications 81(21):30143-30167. doi: 10.1007/s11042-022-12695-5

Bhattarai K, Bhattarai R, Pandey RD, Paudel B, Bhattarai HD. 2024. A comprehensive review of the phytochemical constituents and bioactivities of *Ocimum tenuiflorum*. BioMed Research International 8895039. doi: 10.1155/2024/8895039

Bhumika S, Yadav P, Singh D. 2022. A review on pharmacological action of *Terminalia arjuna* (Roxb.) Wt. and Arn. World Journal of Pharmaceutical Research 11(9):1760-1778.

Biswas D, Mandal S, Chatterjee Saha S, Tudu CK, Nandy S, Batiha GE, Shekhawat MS, Pandey DK, Dey A. 2021. Ethnobotany, phytochemistry, pharmacology, and toxicity of *Centella asiatica* (L.) Urban: A comprehensive review. Phytotherapy Research 35(12):6624-6654. doi: 10.1002/ptr.7248

Borkatulla B, Ferdous J, Uddin AH, Mahmud P. 2023. Bangladeshi medicinal plant dataset. Data in Brief 48:109211. doi: 10.1016/j.dib.2023.109211

Bouakkaz H, Bouakkaz M, Kerrache CA, Dhelim S. 2025. Enhanced classification of medicinal plants using deep learning and optimized CNN architectures. Heliyon 1: e42385. doi: 10.1016/j.heliyon.2025.e42385

Cheng Q, Lyu B, Hu J, Zhang Z, Huang Y, Wang Z. 2024. Research on the antipruritic active ingredients of *Mikania micrantha*. Fitoterapia 174,:105837. doi: 10.1016/j.fitote.2024.105837

der Maaten L, Hinton G. 2008. Visualizing data using t-SNE . Journal of Machine Learning Research 9:2579-2605.

Deshmukh M. 2024. Deep learning for the classification and recognition of medicinal plant species. Indian Journal of Science and Technology 17(11):1070-1077. doi: 10.17485/IJST/v17i11.3002

Dhaka VS, Meena SV, Rani G, Sinwar D, Kavita, Ijaz MF, Wozniak M. 2021. A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. Sensors 21(14):4749. doi: 10.3390/s21144749

Eddouks M, Ajebli M, Hebi M. 2023. Ethnopharmacological survey of plants used to treat diabetes in Tafilalet region of Morocco . Journal of Ethnopharmacology 307:116208. doi: 10.1016/j.jep.2023.116208

Ekor M. 2014. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. Frontiers in Pharmacology 4:177. doi: 10.3389/fphar.2013.00177

Faundes-Gandolfo N, Jara-Gutierrez C, Parraga M, Montenegro I, Vera W, Escobar M, Madrid A, Valenzuela-Valderrama M, Villena J. 2024. *Kalanchoe pinnata* (Lam.) Pers. leaf ethanolic extract exerts selective anticancer activity through ROS -induced apoptotic cell death in human cancer cell lines. BMC Complementary Medicine and Therapies 24:270. doi: 10.1186/s12906-024-04570-7

Figueroa-Mata G, Mata-Montero E. 2020. Using a convolutional Siamese network for image-based plant species identification with small datasets. Biomimetics 5(1): 8. doi: 10.3390/biomimetics5010008

Gairola S, Sharma J, Singh Gaur D. 2021. Traditional uses of medicinal plants and validation of identification through DNA barcoding in Uttarakhand Himalaya, India Journal of Ethnopharmacology 269:113743. doi: 10.1016/j.jep.2020.113743

Hasan MR, Alotaibi BS, Althafar ZM, Mujamammi AH, Jameela J. 2023. An update on the therapeutic anticancer potential of

Ocimum sanctum L .: `` Elixir of life''. Molecules 28(3):1193. doi: 10.3390/molecules28031193

Hazekamp A, Fischedick JT. 2012. *Cannabis*-from cultivar to chemovar. Drug Testing and Analysis 4(7-8):660-667. doi: 10.1002/dta.407

He Z, Hu Y, Niu Z, Zhong K, Liu T, Yang M, Ji L, Hu W. 2023. A review of pharmacokinetic and pharmacological properties of asiaticoside, a major active constituent of *Centella asiatica* (L .) Urb. Journal of Ethnopharmacology 302(Pt A):115865. doi: 10.1016/j.jep.2022.115865

Hernandez-Caballero ME, Sierra-Ramire JA, Villalobos-Valencia R, Sesena-Mendez E. 2022. Potential of *Kalanchoe pinnata* as a cancer treatment adjuvant and an epigenetic regulator. Molecules 27(19):6425. doi: 10.3390/molecules27196425

Islam S, Uddin MZ. 2022. Study of ethnomedicinal plants used by the local people of Raipura Upazila of Narsingdi District. Bangladesh Journal of Plant Taxonomy 29(1):137-156. doi: 10.3329/bjpt.v29i1.60452

Kutal DH, Kunwar RM, Uprety Y, Adhikari YP, Bhattarai S, Adhikari B, Kunwar LM, Bhatt MD, Bussmann RW. 2021. Selection of medicinal plants for traditional medicines in Nepal. Journal of Ethnobiology and Ethnomedicine 17(1):73. doi: 10.1186/s13002-021-00497-0

Kyei S, Amoateng P, Tagoe TA, Mensah NY. 2022. Toxicity and safety implications of medicinal plant use. In Garg N, Manoharadas S (eds) Recent Advances in Medicinal Plants. IntechOpen. doi: 10.5772/intechopen.101804

Lankasena N, Nugara RN, Wisumperuma D, Seneviratne B, Chandranimal D, Perera K. 2024. Misidentifications in ayurvedic medicinal plants: Convolutional Neural Network (CNN) to overcome identification confusions. Computers in Biology and Medicine 183:109349. doi: 10.1016/j.compbiomed.2024.109349

Lecun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature 521(7553):436-444. doi: 10.1038/nature14539

LeCun Y, Bottou L, Bengio Y, Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86(11):2278-2324. doi: 10.1109/5.726791

Lopresti AL, Smith SJ, Metse AP, Drummond PD. 2022. A randomized, double-blind, placebo-controlled trial investigating the effects of an *Ocimum tenuiflorum* (Holy Basil) extract (Holixer TM) on stress, mood, and sleep in adults experiencing stress. Frontiers in Nutrition 9:965130. doi: 10.3389/fnut.2022.965130

Mulugeta AK, Sharma DP, Mesfin AH. 2024. Deep learning for medicinal plant species classification and recognition: a systematic review. Frontiers in Plant Science 14:1286088. doi: 10.3389/fpls.2023.1286088

Musa M, Jan G, Jan FG, Hamayun M, Irfan M, Rauf A, Alsahammari A, Alharbi M, Suleria HAR, Ali N. 2022. Pharmacological activities and gas chromatography-mass spectrometry analysis for the identification of bioactive compounds from *Justicia adhatoda* L . Frontiers in Pharmacology 13:922388. doi: 10.3389/fphar.2022.922388

Muszalska A, Wieckanowska J. 2024. *Ocimum tenuiflorum* L. (Holy Basil) in modern medicine: broad-spectrum therapeutic potential. Herba Polonica 70(2):39-55. doi: 10.5604/01.3001.0054.6597

Pan SY, Zhou SF, Gao SH, Yu ZL, Zhang SF, Tang MK, Sun JN, Ma DL., Han YF, Fong WF, Ko KM. 2013. New perspectives on how to discover drugs from herbal medicines: CAM 's outstanding contribution to modern therapeutics. Evidence-Based Complementary and Alternative Medicine 627375. doi: 10.1155/2013/627375

Pradhan D, Biswasroy P, Halder J, Cheruvanachari P, Dubey D, Kumar V, Kar B, Kar DM, Rath G, Ghosh G. 2022. A comprehensive review on phytochemistry, molecular pharmacology, clinical and translational outfit of *Ocimum sanctum* L. South African Journal of Botany 150:342-360. doi: 10.1016/j.sajb.2022.07.037

Prananda AT, Dalimunthe A, Harahap U, Simanjuntak Y, Peronika E, Karosekali NE, Hasibuan PAZ, Syahputra RA, Situmorang PC, Nurkolis F. 2023. *Phyllanthus emblica*: a comprehensive review of its phytochemical composition and pharmacological properties. Frontiers in Pharmacology 14:1288618. doi: 10.3389/fphar.2023.1288618

Rai PK, Chauhan NP, Datta A, Singh SK. 2022. Invasive alien weeds: implications for biodiversity, ecosystem services and human health. Botany Letters 169(1):55-66. doi: 10.1080/23818107.2021.1957018

Ramesh P, Palaniappan A. 2023. *Terminalia arjuna* , a cardioprotective herbal medicine---relevancy in the modern era of pharmaceuticals and green nanomedicine: a review. Pharmaceuticals 16(1):126. doi: 10.3390/ph16010126

Rudra MS, Chowdhury H, Faruque M, Uddin S. 2022. Diversity and ethnobotanical use of traditional medicinal plants in Badolchori Vadi Sora Village Common forests, Rangamati , Bangladesh . Bangladesh Journal of Plant Taxonomy 29(1):109-128. doi: 10.3329/bjpt.v29i1.60449

Schroff F, Kalenichenko D, Philbin J. 2015. FaceNet: a unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR):815-823. doi: 10.1109/CVPR.2015.7298682

Shorten C, Khoshgoftaar TM. 2019. A survey on image data augmentation for deep learning. Journal of Big Data 6(1): 1-48.

doi: 10.1186/s40537-019-0197-0

Sivaraman G, Srinivasan V, Dhurai B. 2023. Antibacterial activity of silver nanoparticles synthesized from Calotropis gigantea leaf extract. Biomass Conversion and Biorefinery. doi: 10.1007/s13399-023-04091-0

Sofowora A, Ogunbodede E, Onayade A. 2013. The role and place of medicinal plants in the strategies for disease prevention*.* African Journal of Traditional, Complementary and Alternative Medicines 10(5):210-229. doi: 10.4314/ajtcam.v10i5.2

Teschke R, Eickhoff A. 2023. Herbal hepatotoxicity in traditional and modern medicine: actual key issues and new encouraging steps. Frontiers in Pharmacology 14:1173378. doi: 10.3389/fphar.2023.1173378

Uddin MJ, Zidorn C. 2020. Traditional herbal medicines against CNS disorders from Bangladesh . Natural Products and Bioprospecting 11(2):107-159. doi: 10.1007/s13659-020-00269-7

Vasudeva N, Sharma SK. 2008. Post-coital antifertility activity of *Hibiscus rosa-sinensis* L inn. roots. *Evidence-Based* Complementary and Alternative Medicine 5(1):91-94. doi: 10.1093/ecam/nel044

WHO. 2019. WHO global report on traditional and complementary medicine 2019.

Appendix

Appendix A. Complete list of species-pair misclassification events

Table 2 in the main text lists only the 21 critical and high-risk species pairs. Table A1 below reports the complete set of 75 directed species-pair comparisons (of the 380 total directed comparisons across 20 species) for which at least one misclassification occurred on the 2,000-image test set, ordered by descending error count. Pairs not listed here had zero observed errors between the two species.

Table A1. Complete list of all 75 directed species-pair misclassification events observed on the 2,000-image test set (100 images per species), ordered by descending error count.

True species (code)	Predicted species (code)	Errors / 100	Error rate (%)
JA	TA	14	14
JA	CA	12	12
KP	TA	11	11
KP	MM	10	10
CA	JA	9	9
OT	CA	8	8
OT	HRS	8	8
OT	JA	7	7
CA	OT	6	6
CA	PE	6	6
JA	AI	6	6
MM	CA	6	6
MM	CG	6	6
MM	TA	6	6
Haritoki	Pathorkuchi	5	5
KP	CA	5	5
MM	PE	5	5
OT	PE	5	5
Devilbackbone	Thankuni	4	4
HRS	PE	4	4
KP	OT	4	4
AI	PE	3	3
CA	MM	3	3
Devilbackbone	Tulsi	3	3
HRS	AI	3	3
Lemongrass	Neem	3	3
MM	HRS	3	3
OT	AI	3	3
Pathorkuchi	Zenora	3	3
TA	OT	3	3
CG	OT	2	2
CG	PE	2	2
Devilbackbone	Lemongrass	2	2
JA	HRS	2	2
JA	PE	2	2
KP	AI	2	2
KP	HRS	2	2
Lemongrass	Devilbackbone	2	2
MM	OT	2	2
TA	AI	2	2
TA	JA	2	2
Thankuni	Devilbackbone	2	2
Thankuni	Zenora	2	2
AI	TA	1	1

Bohera	Haritoki	1	1
CA	AI	1	1
CA	CG	1	1
CA	KP	1	1
CG	HRS	1	1
CG	Nayontara	1	1
CG	TA	1	1
Devilbackbone	Haritoki	1	1
Devilbackbone	Nayontara	1	1
HRS	MM	1	1
HRS	OT	1	1
Haritoki	Lemongrass	1	1
Haritoki	Nayontara	1	1
Haritoki	Neem	1	1
Haritoki	Tulsi	1	1
JA	MM	1	1
JA	OT	1	1
Lemongrass	Haritoki	1	1
Lemongrass	Zenora	1	1
Nayontara	JA	1	1
Nayontara	Tulsi	1	1
OT	MM	1	1
Pathorkuchi	Bohera	1	1
TA	CG	1	1
TA	HRS	1	1
TA	MM	1	1
TA	PE	1	1
Thankuni	Nayontara	1	1
Tulsi	Nayontara	1	1
Tulsi	Zenora	1	1
Zenora	Tulsi	1	1

Appendix B. ROC analysis for optimal embedding-similarity threshold

To evaluate whether the Siamese Network's 128-dimensional embedding similarity score could itself serve as a binary same-species / different-species classifier, a receiver operating characteristic (ROC) analysis was performed by treating the similarity score between every pair of test-set embeddings as a continuous predictor of same-species membership. The optimal decision threshold was selected using Youden's J statistic (J = sensitivity + specificity − 1). Table A2 and Figure A1 summarize this analysis.

Table A2. Optimal similarity-threshold analysis for same-species discrimination using the Siamese embedding space.

Optimal threshold	Sensitivity	Specificity	Youden's J	AUC
0.998	0.654	0.725	0.379	0.757

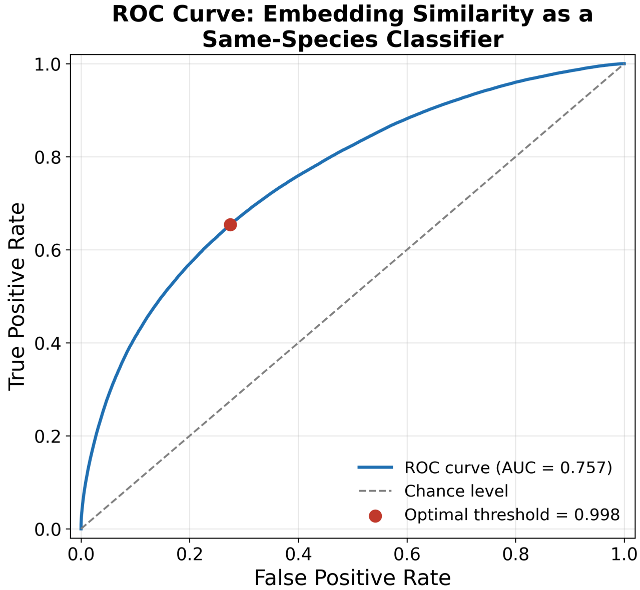

Figure A1. Receiver operating characteristic (ROC) curve for using Siamese embedding similarity to discriminate same-species from different-species image pairs. The red point marks the optimal threshold identified by Youden's J statistic.

Appendix C. Siamese Network (triplet loss) training history

Figure 1 in the main text shows the training history of the Convolutional Neural Network classifier only. Figure A2 below shows the corresponding training history of the Siamese Network, trained separately with semi-hard negative mining triplet loss (margin α = 0.2) for 10 epochs, converging to a final triplet loss of 0.198 as reported in Materials and Methods.

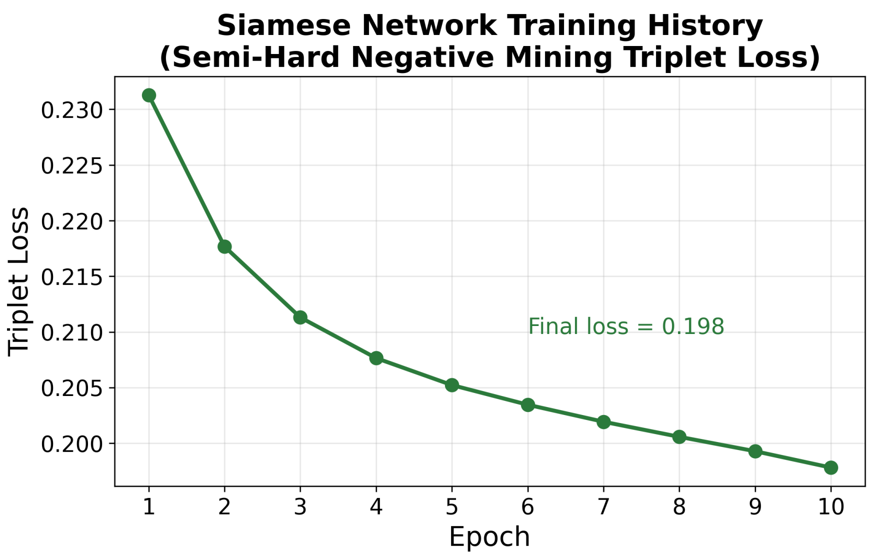

Figure A2. Siamese Network training history: triplet loss (semi-hard negative mining) across 10 training epochs.