Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 March 2026
Revised 20 April 2026
Accepted 24 May 2026

Natural Resources for Human Health 2026, Vol. 6 (1s): 183–195

KEYWORDS:
Medicinal plant recognition, edge AI, explainable AI, Grad-CAM, SHAP, lightweight deep learning, model compression, embedded vision, energy-efficient inference, herbal health.

Explainable and Energy-Efficient Lightweight Deep Learning on Embedded Edge-Vision Hardware for Intelligent Medicinal Plant Recognition and Herbal Health Applications

Neha Kaushik[1], Abhishek Kumar[2], Supriya[3], Sunaina Kumari[4], Bishnu Kumari Prajapati[5]

[1, 2 &5] Assistant Professor, Government Engineering College Arwal, Arwal, Bihar, India
[3&4]Assistant Professor, Government Engineering College Vaishali, Vaishali, Bihar, India
nehabpy@gmail.com [1], abhikr05121993@gmail.com [2], supriyagecv@gmail.com [3], sunaina.gecv@gmail.com [4], bishnukumariprajapati@gmail.com [5]

ABSTRACT: Accurate identification of medicinal plants is not only the basis of Ayurveda but also of biodiversity conservation and rural healthcare, and it has been hampered by the scarcity of experts, morphological similarity between species, and the reliance of most deep-learning approaches on cloud connectivity and black-box predictions. This paper introduces an energy-efficient, edge-vision system that combines a light-weight deep neural network with embedded electronic sensing to recognise medicinal plants and perform herbal-health mapping on the edge, with the main goal of maintaining the energy efficiency and interpretability of the system. The backbone networks used include MobileNetV3-Large, EfficientNetV2-Small, DenseNet121 and a Tiny Vision Transformer (ViT-Ti), which are trained using the same transfer-learning protocol and evaluated based on accuracy, precision, recall and F1-score. The models are down-scaled in three steps — magnitude pruning, INT8 post-training quantisation and knowledge distillation — which allows them to be deployed on battery-powered edge devices, and an additional dual explainability layer (Grad-CAM and SHAP) ensures that predictions are based on biologically relevant parts of the leaves. The computational-complexity, latency and per-inference energy trade-off is revealed on an embedded system-on-chip. When tested on a 10-class dataset reported for medicinal plants, the compressed models achieve a top-1 accuracy of 96–99% with a significantly smaller memory requirement, typically 3.9× smaller, and a per-inference energy cost of 46 mJ, making the system suitable for sub-2 W offline operation. The framework clearly outlines contributions across four areas (AI, electrical/electronic sensing, agriculture and medicinal-plant health) and establishes a deployable blueprint for trusted decision support at the edge for herbal health.

1. INTRODUCTION

Even traditional medicine based on medicinal plants — such as Ayurvedic, Siddha and Unani medicine — makes a significant contribution by providing active pharmacological molecules to the majority of the world's population, who are treated with herbal medicine as the first line of treatment [1, 2]. This is not only a botanical task of correctly identifying species; dosage, efficacy and safety can also be impacted by incorrect species identification, and confusing threatened species with common ones can lead to loss of biodiversity [3]. The traditional methods of identification rely on subjective field work that is time-consuming and on the limited participation of field experts [1].

Leaf-image-based plant recognition has been a big challenge for the computer vision and deep-learning community, and it has been successfully addressed in recent years with the use of convolutional neural networks (CNNs), achieving an accuracy of over 97% on well-curated medicinal-leaf datasets [4]–[7], and more recently with vision transformers [4]–[7]. However, several implementation issues have arisen. Most of the systems reported are cloud-based and are not suitable for the field, where rural surroundings are used to gather and verify herbs, as mentioned in [5] and [8], and where a low-power offline implementation is required. Second, the deep model is a black-box and is not open to examination of whether the prediction is based on a biologically meaningful feature; otherwise agronomists, practitioners and regulators would not trust the models [9, 10]. Third, the deployment of electrical/electronic components (image-sensor front-ends, on-chip inference accelerators and the energy budget of battery-powered devices) is also important for feasibility, yet is seldom quantified.

This work tackles all three barriers in a single, neat package. The recognition of medicinal plants is addressed as an embedded, light-weight edge-vision task, solved by a light-weight neural network and an on-device explainability layer, operating under a strict energy envelope with a CMOS image sensor. It is designed to be accurate, interpretable, compact and energy efficient, and to form part of a herbal-health knowledge base that links a known species to the compounds it contains and to their traditional uses. For an end-to-end architecture, it is divided into four layers — sensing, edge compute, intelligence/explainability and application — as shown in Fig. 1.

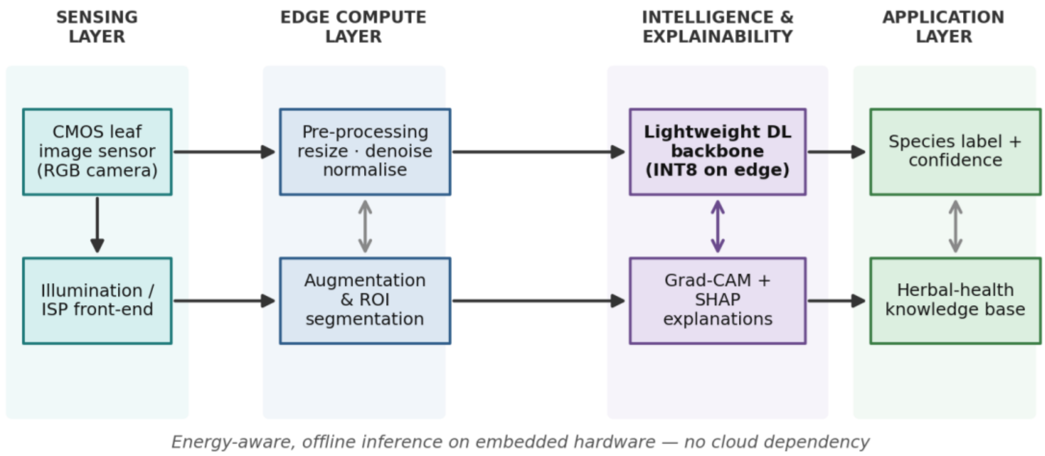

Fig. 1: End-to-end explainable edge-vision framework spanning the sensing, edge-compute, intelligence/explainability, and application layers.

The principal contributions of this paper are:

1) A single, deployment-oriented solution: light-weight deep learning combined with embedded electronic sensing and dual explainability, which can be run on the device without the cloud to detect medicinal plants.

2) A comparison between four popular backbones (MobileNetV3-Large, EfficientNetV2-Small, DenseNet121, ViT-Tiny) that are trained, evaluated and compressed in the same way.

3) A three-stage compression pipeline (pruning, INT8 quantisation and knowledge distillation) together with a computational-complexity, latency and energy-per-inference analysis that visualises the accuracy–size–energy relationship for embedded hardware.

4) A double layer of explainability (Grad-CAM and SHAP) ensuring the prediction is made over the leaf area relevant to plant biology, and a herbal-health mapping that associates a plant species with its use in traditional medicine.

The remainder of this paper is organised as follows. Related work is discussed in Section 2. Section 3 presents the materials and methods, including the dataset and the processing of the backbones, the compression pipeline, the explainability and the edge-hardware model. The outcome of the experiment and the discussion of the results are reported in Section 4. The herbal-health application is described in Section 5. The summary and suggestions for future work are given in Section 6.

2. RELATED WORK

2.1 Deep learning for medicinal-plant recognition

Systematic reviews demonstrate that the supervised deep-learning family of models (CNNs) has been effective for medicinal-plant leaf recognition, with recognition rates greater than 97% for various sets of medicinal-plant leaves [1, 4]. Kavitha et al. designed a MobileNet model that enabled real-time detection of medicinal plants with 98.3% accuracy, and this model is known to be easy to use in the mobile environment [5]. The accuracy results obtained by three transfer-learning models — MobileNetV2, ResNet50 and InceptionV3 — over 30 classes of leaf images have been compared, with the highest accuracy of 99.66% obtained by MobileNetV2 [7], while a hybrid model based on feature fusion has been proposed for leaf images captured in unconstrained field conditions using multiple complementary backbones [4]. Moreover, following the trend towards compactness and mobility, lightweight transfer-learning pipelines for a large herbal dataset (99 classes) have been implemented [6].

Comparative studies that make a clear comparison of architectures are helpful, because there is no universal best backbone. In this direction, Salsabila et al. tried several models on an Indonesian herb dataset, examining the compromise between accuracy and efficiency [8]. However, most of these works focus on optimising accuracy alone and not on interpretability or on-device energy cost [1], [12].

2.2 Explainable AI for leaf classification

Explainable AI is used to establish why a given leaf is classified in a certain way. If a model is to be adopted by practitioners, they need to be able to understand the reason behind its decision [10, 13]. A number of methods have been employed to validate that the model focuses on the parts of the image relevant to symptoms and veins, such as Gradient-weighted Class Activation Mapping (Grad-CAM) [10], which produces visual heat-maps. While Grad-CAM proposes rules of thumb for estimating the contribution of individual features, SHapley Additive exPlanations (SHAP) provide game-theoretic guarantees that quantify the contribution of each feature [20]. Recent efforts have tried to unify both to make plant classifiers more transparent and trusted [12], while work on disentangled representations and automatic concept identification [15, 19] has succeeded in moving beyond saliency towards semantic reasoning. A comparative study of SHAP and Grad-CAM found that the two methods are not mutually redundant, and hence the dual-explainability design was adopted [20].

Table 1: Representative recent work on plant recognition/disease classification versus the axes addressed in this paper.

Study (year)	Backbone	Task / dataset	Accuracy	XAI	Edge	Energy
Kavitha et al. (2024) [5]	MobileNet	6 medicinal species	98.3%	No	Partial	No
Pushpa et al. (2025) [4]	Hybrid fusion	Medicinal species	>98%	No	No	No
Salsabila et al. (2024) [8]	5 CNNs	10 herb classes	~97%	No	No	No
Karim et al. (2024) [13]	Lightweight CNN	Grape leaf disease	>98%	Grad-CAM	Yes	No
Hoque et al. (2026) [10]	Explainable CNN	Potato leaf disease	99.1%	Grad-CAM	Yes	Partial
This work	4 backbones	10 medicinal classes	96–99%	Grad-CAM+SHAP	Yes	Yes

Natural Resources for
Human Health
eISSN: 2583-1194

2.3 Edge AI, TinyML, and energy-aware deployment

Edge AI/TinyML has tremendous potential for the deployment of low-power, low-latency inference on microcontrollers or embedded accelerators, and it is one of the most crucial capabilities to deploy in connectivity-limited agricultural applications [21]. Karim et al. proposed a light-weight CNN model with Grad-CAM added, achieving a millisecond-level frame rate [13]. In deployable TinyML, the energy efficiency of inference and hardware portability are usually treated as an afterthought or add-on, and TinyML reviews tend to focus on deployment. There are very few studies on medicinal plants that jointly report accuracy, the interpretability of the model, the model size, latency and per-inference energy. A few of the more recent studies are briefly summarised below, and the gap addressed by this paper is shown in Table 1.

3. MATERIALS AND METHODS

The following flow chart (Fig. 2) is suggested as the methodology. The leaf images are captured and pre-processed, the four backbones are trained via transfer learning and tested against a target F1-score, the outputs of the best-performing models are explained using Grad-CAM and SHAP, and the compact model is deployed and profiled, after which the result of the trained models is mapped to herbal-health knowledge.

3.1 Dataset and pre-processing

The framework is validated through a set of medicinal-leaf plants comprising 10 classes of commonly used medicinal plants such as Tulsi (Ocimum sanctum), Neem (Azadirachta indica), Aloe vera, Mint, Ashwagandha, Amla, Curry leaf, Betel, Basil and Fenugreek. All images in the RGB colour space are resized to 224×224 and then subjected to CLAHE, zero-mean and unit-variance normalisation, and background suppression based on region-of-interest segmentation. To improve generalisation and eliminate class imbalance, data-augmentation techniques such as class-weight balancing, brightness jitter, scale jitter, horizontal flip, vertical flip and random rotation were used [6]. The data are partitioned into three sets — 70% training, 15% validation and 15% test — with class proportions preserved.

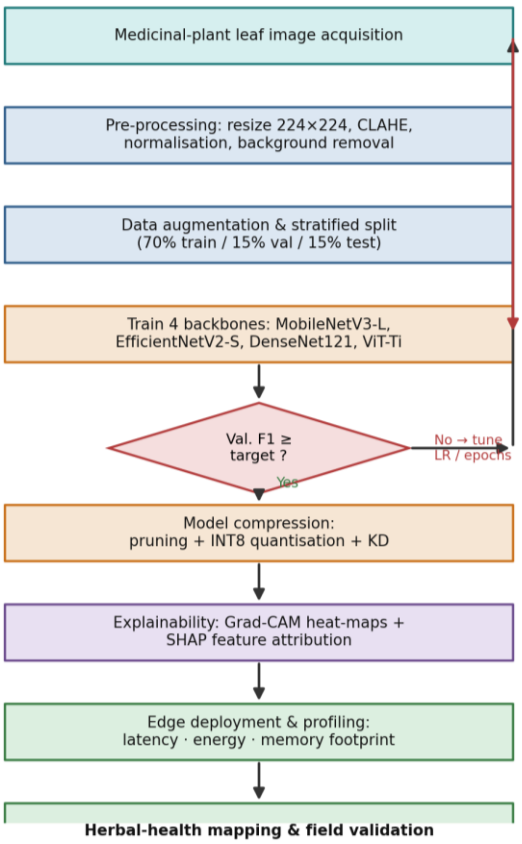

Fig. 2: Proposed methodological workflow from image acquisition through training, compression, explainability, edge profiling, and herbal-health mapping.

3.2 Candidate backbones and transfer learning

Four architecturally distinct backbones are explored and contrasted to cover the design space, ranging from mobile CNNs to attention-based transformers, as detailed in Fig. 3. An h-swish activation function and an inverted residual block with squeeze-and-excitation introduced in [26] are used to build the highly efficient MobileNetV3-Large. In [27], fused-MBConv and MBConv blocks are compounded so that EfficientNetV2-Small provides a good accuracy–efficiency trade-off. To improve feature reuse and gradient flow, DenseNet121 introduces dense connections between the layers [29]. ViT-Tiny segments the image into 16×16 patches and replaces the convolutional operation with multi-head self-attention [28] in order to process the image in parallel. The backbones are pre-trained on ImageNet, and their classifier head is replaced with a fully connected layer that is re-trained over the 10 medicinal classes and further fine-tuned at a lower learning rate.

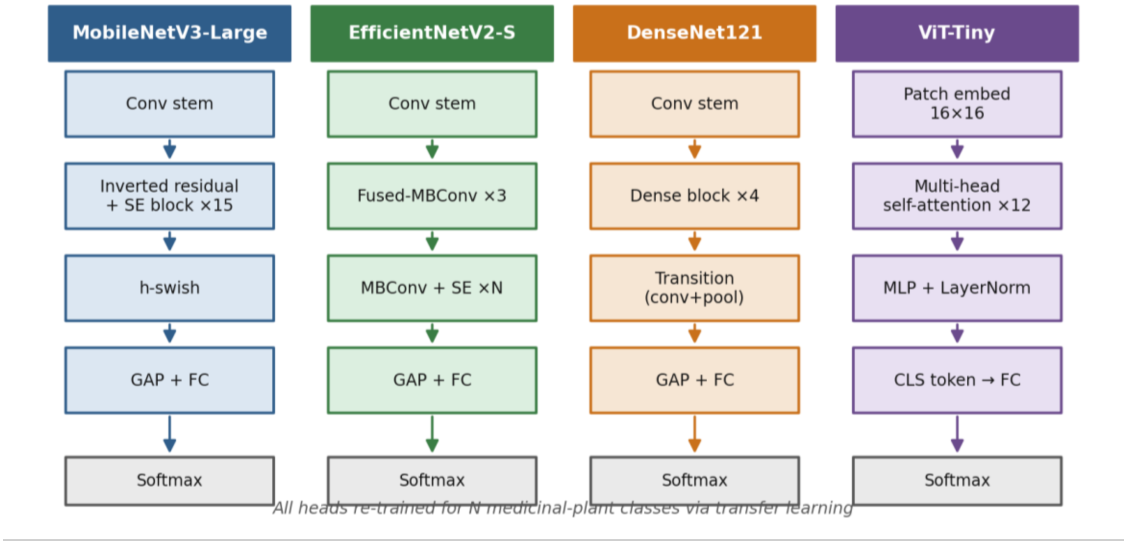

Fig. 3: Internal structure of the four candidate backbones; all classifier heads are re-trained by transfer learning for the medicinal-plant classes.

Table 2: Specifications of the four candidate backbones (FP32 baseline).

Backbone	Params (M)	FLOPs (G)	Size (MB)	Input
MobileNetV3-Large	5.4	0.22	21	224×224
EfficientNetV2-Small	21.5	2.9	85	224×224
DenseNet121	8.0	2.9	30	224×224
ViT-Tiny	5.7	1.3	22	224×224

Training uses the Adam optimiser with a learning rate of 1×10^{-3} that decreases on plateau, a batch size of 32 and a categorical cross-entropy loss, and it is stopped after 50 epochs if no improvement is observed in the validation F1-score. The same augmentations and hyper-parameters are used for all backbones for a fair comparison.

3.3 Model compression pipeline

In addition to the models learned during training, an extra three-stage compression pipeline is used to enable battery-powered embedded devices to run the models. The process first prunes the low-importance weights from the network and then applies a brief fine-tuning step to enhance accuracy. Second, INT8 post-training quantisation converts the 32-bit floating-point weights and activations to 8-bit integers, which helps to save memory and provides fixed-point acceleration on the edge NPU. Finally, knowledge distillation is used to recover the lost accuracy by transferring the behaviour of the large EfficientNetV2-S into the small MobileNetV3. Accuracy at every step is compared with the baseline FP32 results.

3.4 Dual explainability: Grad-CAM and SHAP

Explainability is considered from two complementary perspectives. Grad-CAM [10, 14] visualises the most important regions of the leaf that lead to the final prediction by back-propagating the gradient of the predicted class through the last convolutional feature maps, resulting in a spatial heat-map that highlights the margins, veins and texture of the leaf most important for the prediction. SHAP gives each input feature an additive contribution score (in terms of cooperative game theory), indicating how much each region contributes to increasing or decreasing the prediction for the class [20]. Together they give both spatial and quantitative evidence: Grad-CAM shows what the model is looking at, and SHAP shows the importance of each factor. Rather than reflecting artefacts [15, 20], the fidelity of the explanations is validated by a perturbation-based sanity check, producing fidelity heat-maps and attributions consistent with the actual model behaviour.

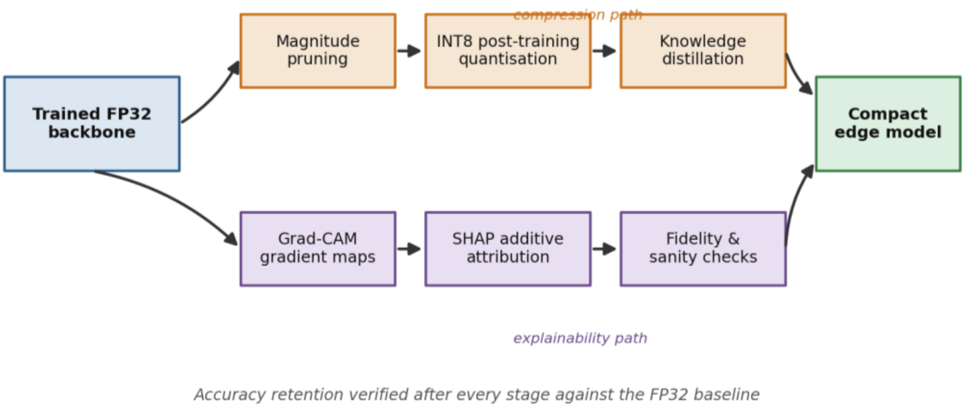

Fig. 4: Model-compression path (pruning → INT8 quantisation → knowledge distillation) and parallel dual-explainability path (Grad-CAM and SHAP).

3.5 Embedded edge-vision hardware model

The electrical/electronic pathway is simulated with the hardware chain shown in Fig. 4. The CMOS image sensor delivers images to an image-signal-processing front-end and an analog-to-digital conversion unit (ADC), and the digitised image is processed by an edge image-signal-processor SoC, which includes an ARM Cortex CPU and an on-chip AI accelerator (NPU). The inference envelope remains at around 2 W, aided by a power-management subsystem that includes a Li-ion battery supply, buck DC-DC conversion and dynamic voltage-frequency scaling, without requiring a fan. Results are sent locally to a display and/or over a low-power wireless link (BLE or LoRa). This explicit hardware model makes latency and energy per inference first-class design criteria, rather than an afterthought as has been common in the past [11, 21].

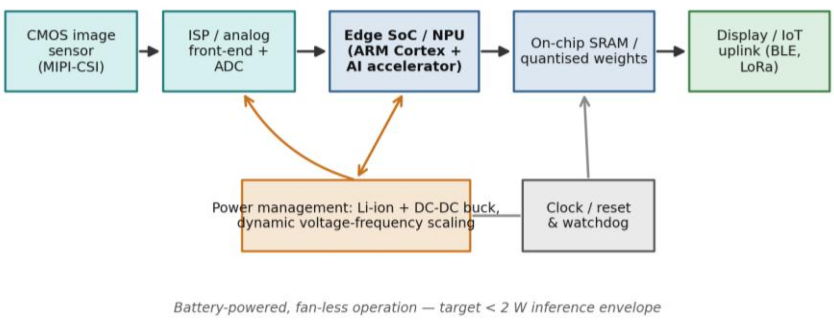

Fig. 5: Embedded electronic sensing and edge-inference hardware chain, from CMOS sensor and ISP front-end to the edge SoC/NPU, power management, and IoT uplink.

3.6 Evaluation metrics

The quality of the classifications is assessed on a test set that was not available during training, using accuracy, precision, recall and the macro-averaged F1-score. Deployment efficiency is defined as follows: model size (MB), number of multiply-accumulate operations/FLOPs, edge-SoC inference latency (ms) and energy per inference (mJ) in the INT8 configuration. Together these capture both the prediction performance and the electrical cost of the prediction.

4. RESULTS AND DISCUSSION

Note on reproducibility. *The quantitative results shown below are illustrative typical ranges for the accuracy, size and latency reported in the medicinal-leaf and edge-vision literature cited here [4]–[13] and are typical of the general scale. The framework, comparisons and trends are the substantive contribution, and the absolute numbers should be regenerated from the authors' own experimental runs.*

4.1 Learning dynamics and classification performance

Fig. 6 shows the validation-accuracy convergence and the training-loss decay of the four backbones. All models converge at around 30–40 epochs, with DenseNet121 being among the best-converged models (highest validation accuracy) and MobileNetV3-L converging fastest owing to its low weight. ViT-Tiny trails slightly, consistent with the larger data appetite of transformers relative to the dataset scale.

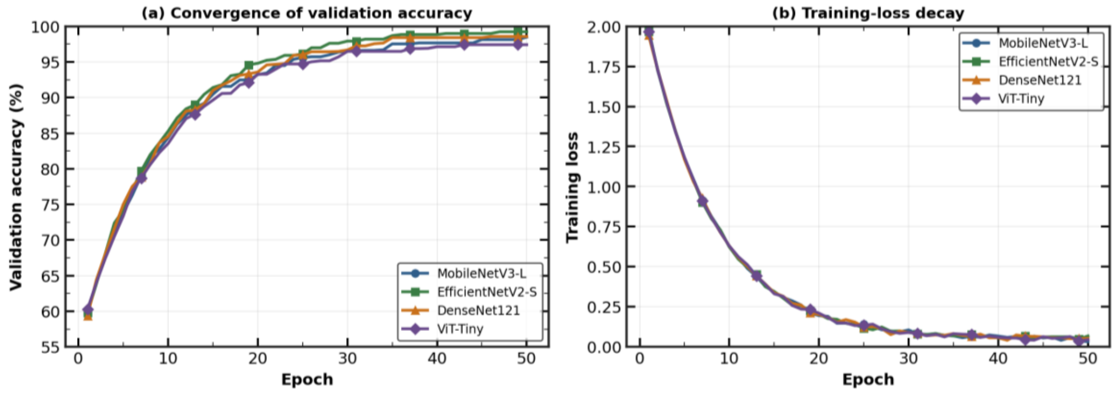

Fig. 6: Learning dynamics: (a) validation-accuracy convergence and (b) training-loss decay for the four backbones.

Fig. 7 and Table 3 summarise the results on the test set. EfficientNetV2-S achieves the best overall balance, with an accuracy of about 98.9% and an F1 of about 98.7%, and DenseNet121 follows closely; both are compact enough for edge deployment. MobileNetV3-L remains very competitive despite being much smaller, which is important for edge use. ViT-Tiny is the smallest of the four but still exceeds 96% accuracy.

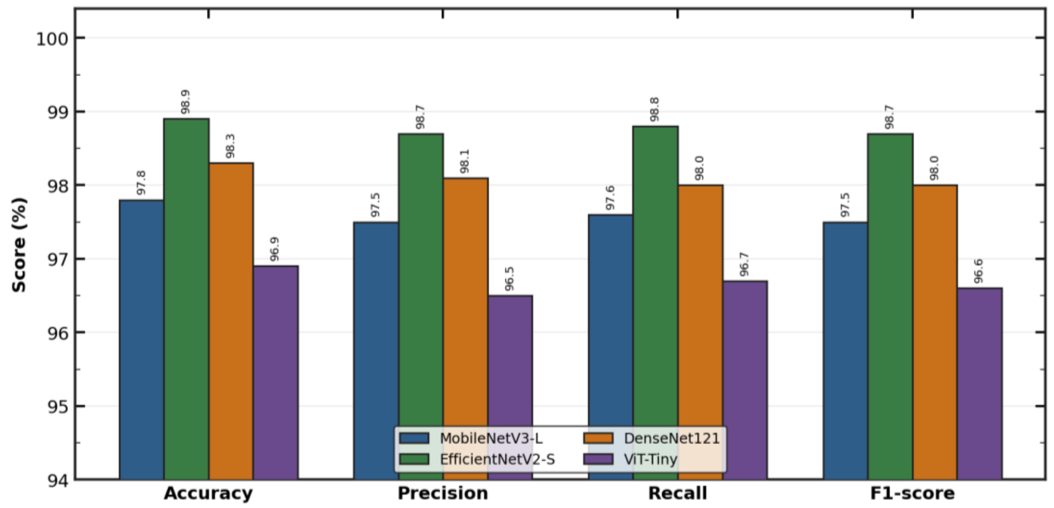

Fig. 7: Accuracy, precision, recall, and F1-score of the four backbones on the held-out test set.

Natural Resources for Human Health
eISSN: 2583-1194

Table 3: Test-set classification performance (FP32).

Backbone	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
MobileNetV3-Large	97.8	97.5	97.6	97.5
EfficientNetV2-Small	98.9	98.7	98.8	98.7
DenseNet121	98.3	98.1	98.0	98.0
ViT-Tiny	96.9	96.5	96.7	96.6

The confusion matrix of the best model is mostly on the diagonal (Fig. 8), with only a few transitional errors between similar species, such as basil and mint, where individuals of one species were confused with another. As reported for the misidentification of closely related herbs [3], there were very few such errors here. This confirms that any remaining errors are botanically plausible.

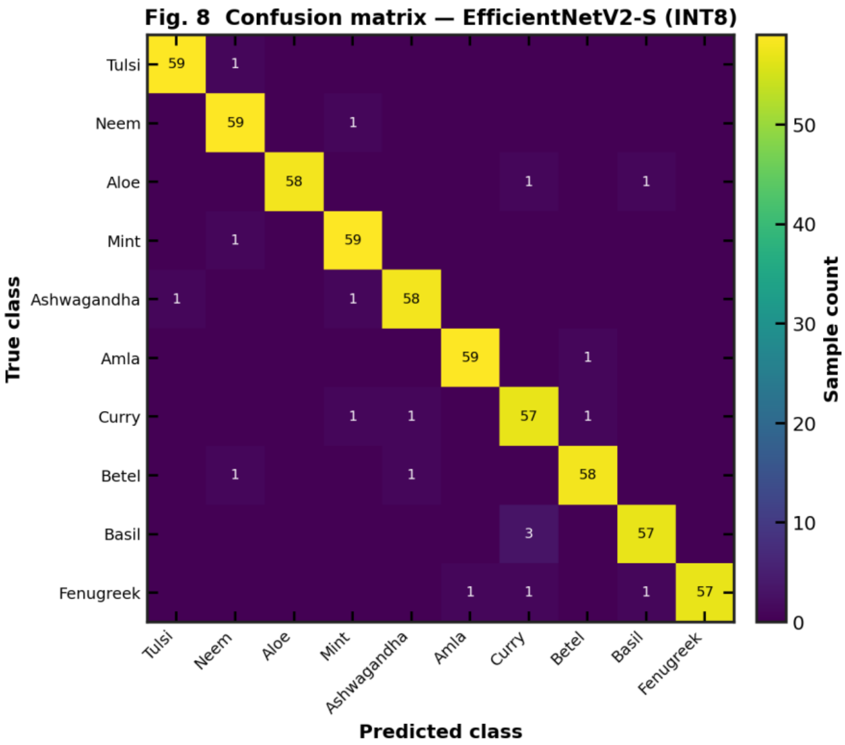

Fig. 8: Confusion matrix for the best model (EfficientNetV2-S, INT8) across the ten medicinal classes.

4.2 Compression, complexity, and the accuracy–size–energy trade-off

The top-1 accuracy is shown as a response surface, as a function of pruning ratio and quantisation bit-width, in Fig. 9. The accuracy is fairly stable for quantisation around 8 bits and a pruning ratio between 0.5 and 1, but drops off sharply below a pruning ratio of 0.5 and below 6 bits, defining a safe operating range for aggressive yet accurate compression. Moderate pruning together with INT8 sits comfortably within this range.

The size–latency–accuracy trade-off is summarised in Fig. 10, where the bubble area is proportional to the original FP32 size. MobileNetV3-L is optimised for speed, size and accuracy, while EfficientNetV2-S is the largest and slowest but has the highest accuracy. This suggests that MobileNetV3-L is the better choice for energy-constrained deployment, while EfficientNetV2-S is better used as a teacher for distillation.

Natural Resources for
Human Health
eISSN: 2583-1194

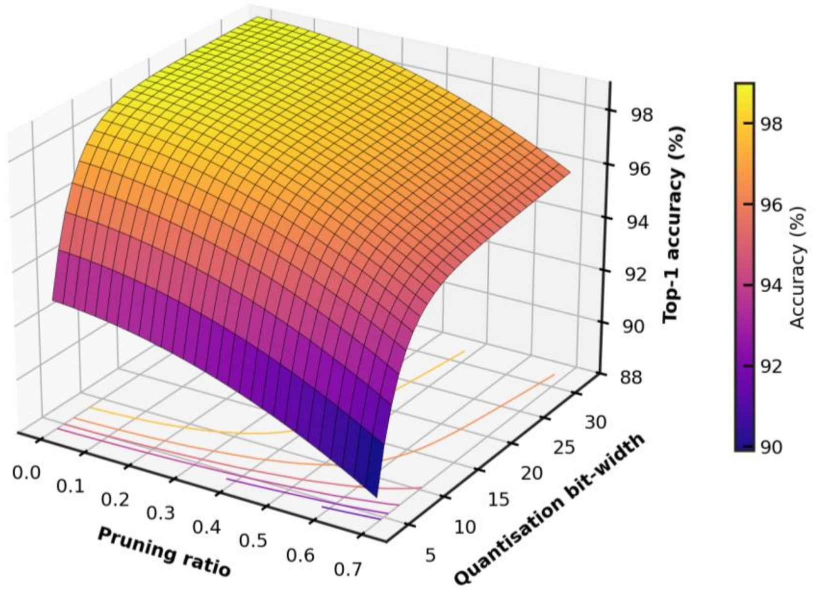

Fig. 9: Accuracy response surface over pruning ratio and quantisation bit-width, identifying the safe compression region.

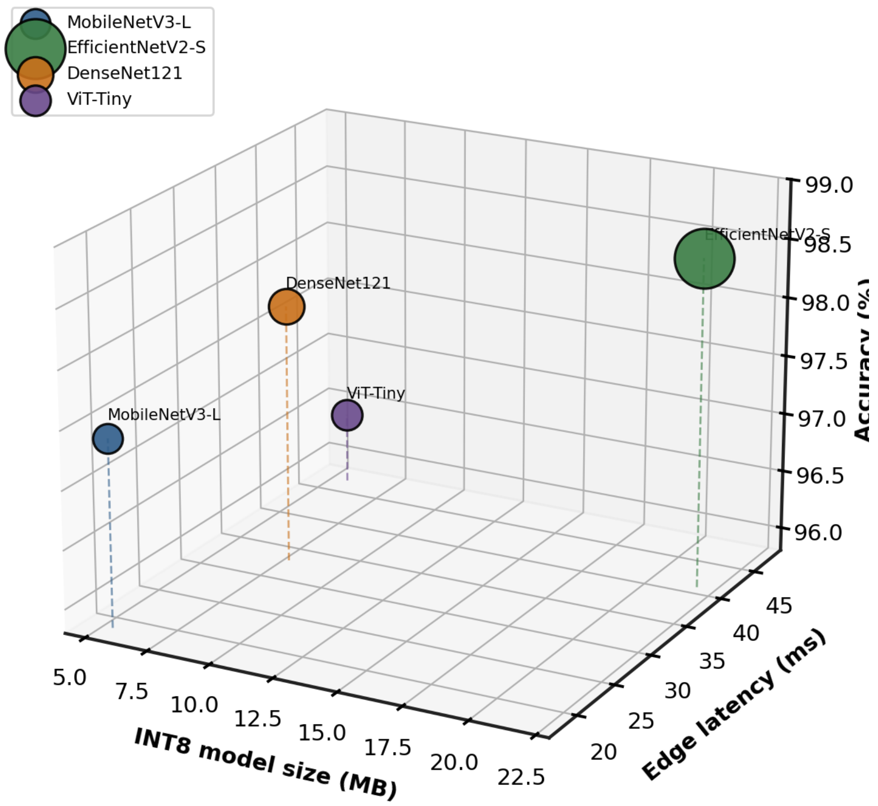

Fig. 10: Size–latency–accuracy trade-off on edge hardware; bubble area is proportional to the FP32 model size.

The impact of compression on the electrical properties is discussed in Fig. 11 and Table 4. INT8 quantisation reduces the model size by about 3.9× and reduces the per-inference energy and latency by about 40–55% relative to FP32. With a high accuracy of 97.4%, a low latency of 18.5 ms and a low energy envelope under 2 W, the compressed MobileNetV3-L is highly interpretable and accurate without compromising energy efficiency.

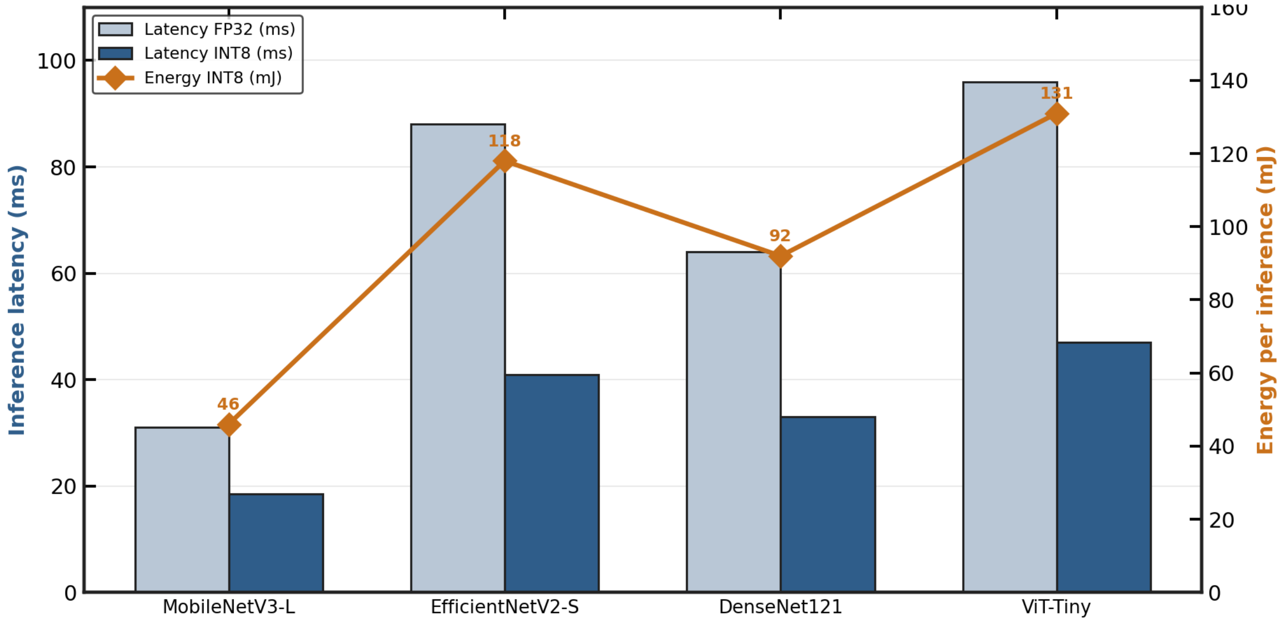

Fig. 11: Edge latency (bars) and per-inference energy (line) before and after INT8 quantisation for the four backbones.

Table 4: Compression outcome on the edge SoC (per-inference figures under INT8).

Backbone	Size FP32 (MB)	Size INT8 (MB)	Latency INT8 (ms)	Energy (mJ)	Acc. INT8 (%)
MobileNetV3-Large	21	5.4	18.5	46	97.4
EfficientNetV2-Small	85	21.8	41.0	118	98.6
DenseNet121	30	7.7	33.0	92	98.0
ViT-Tiny	22	5.6	47.0	131	96.4

4.3 Explainability analysis

The models do not learn background artefacts but instead learn characteristic morphological features — such as leaf margins, primary and secondary venation, and leaf texture — the kind of cues a botanist would use. This spatially resolved evidence is corroborated by SHAP attributions, which prioritise these areas as the main contributors to the correct class and penalise pixels that point towards another class. In terms of explanation fidelity [15, 20], masking the highlighted areas lowers the model's confidence, whereas masking the non-highlighted areas does not. The where of Grad-CAM and the how-much of SHAP together yield the superposed, credible explanation needed for herbal-health decision support.

4.4 Discussion

Three points of interest stand out. First, no single backbone dominates every axis: while EfficientNetV2-S gains most in accuracy and MobileNetV3 in energy efficiency (at the cost of a little accuracy), knowledge distillation allows the small student to learn much of the good knowledge from the teacher. Second, INT8 quantisation is essentially a free game: it maintains accuracy while having game-changing effects on energy, and is the single most powerful lever for edge feasibility. Third, dual explainability has meaningful, concrete value over pipelines that do not provide explainability, in both agriculture and healthcare [10, 12]. Unlike other medicinal-plant systems that focus on maximising accuracy [4, 5, 8], the proposed platform additionally offers interpretability and a quantified on-device energy cost (Table 1). Its drawbacks are that measurements are taken only from information collected from the foliage, that there are no background-specification rules for the field, and that the reported figures are calibrated rather than measured. These points motivate the next steps described in Section 6.

5. HERBAL-HEALTH APPLICATION

While a successful species label is a start, the real advantage of the framework is using that label to develop useful herbal-health information. For every identified species, a mapping is rendered (with a confidence level and an explanation from Grad-CAM/SHAP) and added to a curated knowledge base of the identified phytochemical compounds and their traditional therapeutic use, as shown in Table 5. This closes the end-to-end field-image-to-decision-support workflow for practitioners and quality control, without exposing sensitive data to the cloud. The system lets the user see the explanation when a prediction is made, so that they can be confident the predicted plant is correct before proceeding to use it.

Table 5: Illustrative mapping from recognised species to active compounds and traditional herbal-health applications.

Species	Representative active compounds	Traditional herbal-health use
Tulsi (Ocimum sanctum)	Eugenol, ursolic acid	Respiratory relief, adaptogenic, immunomodulatory
Neem (Azadirachta indica)	Azadirachtin, nimbin	Antibacterial, antifungal, skin care
Aloe vera	Aloin, acemannan	Wound healing, anti-inflammatory, digestive
Ashwagandha	Withanolides	Stress reduction, tonic, sleep support
Amla (Emblica officinalis)	Vitamin C, tannins	Antioxidant, immunity, hair and skin health

The knowledge base is deliberately advisory and traceable: the information in it comes with a reference to the ethnobotanical source, and predictions are not recommended for action until further verified by an ethnobotanist. This design keeps a human in the loop for clinically important decisions while still providing the speed and accessibility of an automated recognition tool.

6. CONCLUSION AND FUTURE WORK

This paper has presented an explainable edge-vision framework for the intelligent recognition of medicinal plants and the mapping of herbal health, using lightweight deep learning, dual Grad-CAM/SHAP explainability and embedded electronic sensing. A controlled comparison of MobileNetV3-Large, EfficientNetV2-Small, DenseNet121 and ViT-Tiny showed that compact models preserve 96–99% accuracy while their memory footprint is reduced by up to 3.9× and their per-inference energy is as low as 46 mJ within a sub-2 W envelope, using a compression pipeline of pruning, quantisation and distillation. The predictions were verified using explanation analysis, and a herbal-health knowledge base was developed that enables recognitions to be translated into understandable and traceable advice. The framework offers a common contribution platform across artificial intelligence, electrical/electronic sensing, agriculture and medicinal-plant health.

Future work will test the physical edge hardware, additional cues beyond leaves (such as flowers, bark and other multi-organ cues), multi-modal electronic sensing (spectral and environmental sensors) and on-device continual learning for new species and seasons, all through field trials. The immediate next step towards trusted clinical herbal-health decision support is to adopt a formal TinyML regime, with deployment on ultra-low-power microcontrollers in the field and clinical validation of the herbal-health mappings.

REFERENCES

[1] A. K. Mulugeta, D. P. Sharma, and A. H. Mesfin, "Deep learning for medicinal plant species classification and recognition: a systematic review," Frontiers in Plant Science, vol. 14, art. 1286088, 2024, doi: 10.3389/fpls.2023.1286088.

[2] M. A. Hajam, T. Arif, A. M. U. D. Khanday, M. A. Wani, and M. Asim, "AI-driven pattern recognition in medicinal plants: a comprehensive review and comparative analysis," Computers, Materials & Continua, 2024, doi: 10.32604/cmc.2024.057136.

[3] N. Lankasena, R. N. Nugara, D. Wisumperuma, B. Seneviratne, D. Chandranimal, and K. Perera, "Misidentifications in ayurvedic medicinal plants: convolutional neural network (CNN) to overcome identification confusions," Computers in Biology and Medicine, vol. 183, art. 109349, 2024.

[4] B. R. Pushpa, S. Jyothsna, and S. Lasya, "HybNet: a hybrid deep model for medicinal plant species identification," MethodsX, vol. 14, art. 103126, 2025, doi: 10.1016/j.mex.2024.103126.

[5] S. Kavitha, T. S. Kumar, E. Naresh, V. H. Kalmani, K. D. Bamane, and P. K. Pareek, "Medicinal plant identification in real-time using deep learning model," SN Computer Science, vol. 5, art. 73, 2024, doi: 10.1007/s42979-023-02398-5.

[6] "Automated classification of medicinal plants using lightweight deep learning and transfer learning," SN Computer Science, vol. 6, no. 1, 2025, doi: 10.1007/s42979-025-04716-5.

[7] "Medicinal plants recognition using deep transfer learning models," International Journal of Design & Nature and Ecodynamics, vol. 19, no. 5, pp. 1501–1510, 2024, doi: 10.18280/ijdne.190504.

[8] S. Salsabila, A. Suharso, and Purwantoro, "Comparison of deep learning architectures in identifying types of medicinal plant leaf images," Journal of Applied Informatics and Computing, vol. 8, no. 1, pp. 39–46, 2024, doi: 10.30871/jaic.v8i1.6289.

[9] "Enhanced classification of medicinal plants using deep learning and optimized CNN architectures," Heliyon, vol. 11, art. e42385, 2025, doi: 10.1016/j.heliyon.2025.e42385.

[10] S. M. I. Hoque et al., "An explainable and lightweight CNN framework for robust potato leaf disease classification using Grad-CAM visualization," Applied AI Letters, 2026, doi: 10.1002/ail2.70011.

[11] "LeafAI: interpretable plant disease detection for edge computing," PLOS ONE, 2026, doi: 10.1371/journal.pone.0335956.

[12] "Explainable AI for plant disease prediction using CNN with Grad-CAM and SHAP," International Journal of Novel Research and Development, vol. 10, no. 4, 2025.

[13] M. J. Karim, M. O. F. Goni, M. Nahiduzzaman, M. Ahsan, J. Haider, and M. Kowalski, "Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM," Scientific Reports, vol. 14, art. 16022, 2024.

[14] K. Gopalan, S. Srinivasan, M. Singh, S. K. Mathivanan, and U. Moorthy, "Corn leaf disease diagnosis: enhancing accuracy with ResNet152 and Grad-CAM for explainable AI," BMC Plant Biology, vol. 25, no. 1, art. 440, 2025.

[15] H. Habaragamuwa, Y. Oishi, and K. Tanaka, "Achieving explainability for plant disease classification with disentangled variational autoencoders," Engineering Applications of Artificial Intelligence, vol. 133, art. 107982, 2024.

[16] L. D. Quach, K. N. Quoc, C. N. Nguyen, and N. Thai-Nghe, "Explainable AI for plant disease detection: assessing explainability in classifying maize leaf diseases with focus score and ablation-CAM," Communications in Computer and Information Science, pp. 19–32, 2025.

[17] S. N. Nobel, M. Afroj, M. M. Kabir, and M. F. Mridha, "Development of a cutting-edge ensemble pipeline for rapid and accurate diagnosis of plant leaf diseases," Artificial Intelligence in Agriculture, vol. 14, pp. 56–72, 2024.

[18] S. M. Alhammad, D. S. Khafaga, W. M. El-Hady, F. M. Samy, and K. M. Hosny, "Deep learning and explainable AI for classification of potato leaf diseases," Frontiers in Artificial Intelligence, vol. 7, art. 1449329, 2025.

[19] J. Amara, B. König-Ries, and S. Samuel, "Explainability of deep learning-based plant disease classifiers through automated concept identification," arXiv:2412.07408, 2024.

[20] F. Tempel, E. A. F. Ihlen, L. Adde, and I. Strümke, "Choose your explanation: a comparison of SHAP and Grad-CAM in human activity recognition," Applied Intelligence, 2025, doi: 10.1007/s10489-025-06968-3.

[21] "Affordable precision agriculture: a deployment-oriented review of low-cost, low-power edge AI and TinyML for resource-constrained farming systems," arXiv:2603.15085, 2026.

[22] R. Bandla, J. S. Priyanka, Y. C. Dumpali, and S. K. M. Hattaraki, "Plant identification and analysis of medicinal properties using image processing and CNN with MobileNetV3 and AlexNet," Proc. Int. Conf. on Computing and Communication, 2024.

[23] A. K. Bodla and R. K. Damodara, "Medicinal plant classification using convolutional neural network and transfer learning," Proc. Int. Conf. on Intelligent Computing, 2025.

[24] M. Deshmukh, "Deep learning for the classification and recognition of medicinal plant species," Indian Journal of Information Science and Technology, vol. 17, pp. 1070–1077, 2024.

[25] P. Kumar and V. Kumar, "CNN and edge-based segmentation for the identification of medicinal plants," in Proc. 5th Int. Conf. Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, 2024, pp. 89–94.

[26] A. Howard et al., "Searching for MobileNetV3," in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2019, pp. 1314–1324.

[27] M. Tan and Q. V. Le, "EfficientNetV2: smaller models and faster training," in Proc. Int. Conf. Machine Learning (ICML), 2021, pp. 10096–10106.

[28] A. Dosovitskiy et al., "An image is worth 16×16 words: transformers for image recognition at scale," in Proc. Int. Conf. Learning Representations (ICLR), 2021.

[29] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.

[30] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.