SN Computer Science (2025) 6:551
https://doi.org/10.1007/s42979-025-04099-7

ORIGINAL RESEARCH

HCVINet: A Multimodal Deep Learning approach for Medicinal Plant Classification Using Visual and Semantic Features

E. S. Vani[1] · H. Manjula Gururaj[2] · B. Karunakara Rai[3]

Received: 21 November 2024 / Accepted: 27 May 2025

Abstract
Proper recognition and classification of medicinal plants is fundamental to pharmacology, botany, and traditional medical applications. However, distinguishing closely related species remains difficult, and traditional methods have not been able to fully address these challenges. In this study, we propose a pioneering deep learning architecture, the Hierarchical Contextual Vision Integration Network (HCVINet), which integrates multi-level image feature extraction with contextual-semantic information using natural language processing (NLP) techniques. HCVINet employs a Hierarchical Feature Extraction Network to capture a broad spectrum of visual features—from elementary textures to complex patterns—and fuses these with semantic data through a Contextual-Correlation Integrated Network (CCINet), enabling the model to utilize both visual and textual information for improved classification decisions. Experiments conducted on benchmark medicinal plant datasets demonstrate that HCVINet achieves a classification accuracy of up to 98.3%, outperforming existing state-of-the-art CNN-based models by an average margin of 4.5%. The model also yields higher retrieval rates and improved harmonic mean scores, validating the effectiveness of combining visual and semantic cues. While HCVINet proves to be a robust tool for medicinal plant classification, its performance may be influenced by the quality of the textual corpus, and occasional mismatches between text and image features can affect classification outcomes. Overall, HCVINet offers a significant advancement in automated plant identification and provides promising implications for research and practical applications.

Keywords Deep Learning Framework · NLP · Semantic Features · Medicinal Plant Classification

Introduction

Medicinal plants, long utilized in traditional medicine, are now acknowledged as important resources for creating new pharmaceuticals. Pharmaceutical experts continue to extract vital chemicals from medicinal plants to produce drugs in the present day, although their approaches to utilizing these substances have evolved over time. Quinine, aspirin, and morphine are three examples of therapeutic chemicals derived from natural sources. Quinine comes from the bark of cinchona plants, while aspirin is made from willow bark, and morphine comes from the poppy opium plant. Therefore, medicinal plants are of great importance in terms of modern pharmacology. According to WHO, about 60% of the world population take pharmacotherapeutics derived from these plants [1]. Evidence suggests that more than 80% of prescription drugs in the USA are partly or wholly sourced from medicinal plants [2]. It is essential to endure in mind that the superiority of dependence on these capsules is often significantly better in developing international

✉ E. S. Vani
vaniesomashekar@gmail.com

H. Manjula Gururaj
dr.manjulagururajh@gmail.com

B. Karunakara Rai
karunakara.rai@nmit.ac.in

1 Information Science and Engineering, Nitte (Deemed to be University), Nitte Meenakshi Institute of Technology (NMIT), Bangalore 560064, India

2 Information Science and Engineering, Nitte Mahalinga Adyantaya Memorial Institute of Technology, Karkala 574110, India

3 Electronics and Communication Engineering, Nitte (Deemed to be University), Nitte Meenakshi Institute of Technology (NMIT), Bangalore 560064, India

Published online: 16 June 2025

locations. Furthermore, there has been a marginal boom in worldwide call for for medicinal herbs. Recently, it was discovered that almost 10% of vascular plant life have therapeutic homes. The global take a look at counseled that there are approximately 350,000 awesome species of medicinal plants [3]. Professional observation is a crucial method for identifying medicinal plants, evaluating morphology, pigmentation, surface features, and olfactory qualities. However, this manual method is time-consuming and susceptible to fatigue, leading to increased variability among observers. Advancements in computer vision and AI have developed automated methods for plant identification. The objective of these methods is to accurately diagnose leaf diseases and classify the medicinal plants. The purpose of this endeavor is to develop new pharmaceuticals or to enhance the effectiveness of existing pharmaceuticals.

Nevertheless, the effectiveness of this technique is impeded by numerous similarities among various plant species, complex backgrounds, and variations in attributes such as light and color. There are more than 8000 different kinds of medicinal plants on the Indian continent. Nevertheless, using traditional techniques to identify these plants presents serious difficulties. In order to accurately classify plant images using CV systems, it is necessary to classify them into several kinds [4]. It is crucial to create reliable methods for analyzing and understanding trends in leaf image data. Focused research on the creation of models or systems for the automated identification of Indian medicinal plants is currently lacking. Without the need for botanists or Ayurvedic specialists, the automation of medicinal plant identification would offer the public and other stakeholders useful information. Deep learning methods for plant classification have been studied by researchers [5–9].

The primary objective of this study was to address the challenges associated with plant classification and identification. Natural Language Processing (NLP) and text classification significantly enhance medicinal plant classification by enabling the integration of vast amounts of textual data with traditional image-based methods. Multilayered neural networks are used in deep learning to tackle a variety of issues, from straightforward to extremely complex [9]. Conventional machine-learning techniques necessitate human feature selection. On the other hand, a method that automatically pulls features from data is called deep learning [10, 11]. Multi-layered networks can optimize their design and successfully handle a variety of issues by utilizing large datasets and high-performance computational approaches. When attempting to obtain the required quantity of data to train deep neural networks from scratch, researchers may run into financial difficulties. In the classification of leaf images, deep learning has shown encouraging results [12, 13]. Herbal flora is typically characterized by its therapeutic attributes, regional presence, and customary applications in various sources such as botanical records, academic studies, and repositories. Utilizing NLP methodologies, these documents can be examined to extract pertinent insights, detect significant trends, and classify species based on contextual meaning. Text categorization further enhances this procedure by systematically sorting descriptions into predefined groups, fostering a deeper comprehension of each species' traits. This integrated strategy not only boosts the precision of plant recognition by incorporating both imagery and textual data but also uncovers connections among species, enabling well-informed choices in scientific exploration and real-world usage.

Recent advances in artificial intelligence and deep learning have significantly improved the accuracy and scalability of plant classification and disease detection. State-of-the-art approaches increasingly leverage transfer learning, multimodal data integration, and contextual information to address the challenges of complex and diverse biological datasets. For example, deep transfer learning has demonstrated high effectiveness in plant disease identification, as seen in cassava plant disease detection using modern deep learning methods. In parallel, the integration of multilingual and contextual data has shown promise in other domains, such as the development of multilingual speech datasets for healthcare applications. These contemporary studies highlight a growing trend towards combining visual, contextual, and semantic information to enhance classification accuracy and interpretability. Building on these recent developments, our work introduces a unified deep learning framework that synergistically fuses multi-level visual features with natural language processing-based semantic information for robust medicinal plant classification [14].

The classification and identification of medicinal plants are crucial for pharmacology, botany, and traditional medicine. However, the complexity and diversity of these plants make manual identification time-consuming and error-prone. A novel deep learning architecture enhances medicinal plant classification by integrating multi-level image feature extraction with contextual semantic data and natural language processing (NLP). Traditional image-based classification methods often face difficulties with these challenges, particularly when working with large datasets. Additionally, current methods often miss out on the valuable contextual and semantic information found in botanical literature and other textual resources, which can greatly improve classification accuracy. To tackle these challenges, we urgently need a more advanced approach that utilizes image-based identification along with semantic and contextual information. This study introduces the Hierarchical Contextual Vision Integration Network (HCVINet), an innovative deep learning framework that merges multi-level image feature extraction with contextual correlation integration and natural language processing (NLP). The key contributions of this study include the following:

1. **Layered Attribute Extraction Framework:** This is a multi-layered convolutional architecture designed to capture both basic textures and complex patterns, improving the system's ability to differentiate between various plant species.
2. **Contextual-Association Unified Model (CAUM):** This mechanism combines visual attributes with contextual and linguistic insights obtained from natural language processing, allowing the framework to leverage relationships among plant classifications for greater accuracy.
3. **Enhanced Categorization Precision:** Significant improvements have been observed across benchmark datasets, showing marked progress in various evaluation metrics such as retrieval rate, harmonic mean score, and overall effectiveness.

The proposed framework has limitations, including reliance on textual inputs and computational demands for merging diverse data types. Future research should focus on streamlining the integration mechanism and incorporating a wider dataset. Potential advancements include broadening the dataset to include a wider variety of medicinal flora, improving NLP incorporation, and exploring practical applications in botanical research and pharmaceutical innovations.

The remainder of this paper is organized as follows: Sect. "Related Work" reviews related work and existing approaches to medicinal plant classification. Sect. "Proposed Methodology" details the proposed Hierarchical Contextual Vision Integration Network (HCVINet), including its architecture and integration of visual and semantic features. Sect. "Performance Evaluation" presents the experimental setup, datasets, evaluation metrics, and results, followed by a discussion of findings. Sect. "Conclusion" concludes the study and suggests directions for future research.

Related Work

Many automated image-based methods have been developed for classifying medicinal plants. Three primary categories may be used to group the techniques: The following three strategies were taken into account in this study: Three methods of selecting and classifying features are as follows: (1) Convolution Neural Network (CNN)-based feature selection and classification; (2) manual feature selection using CNN or traditional machine learning (TML) classifiers; and (3) several CNN or TML models cascaded together. These methods usually focus on the investigation of therapeutic plants that are present in certain areas, with a focus on enhancing the classification models used in these plants.

The Laplacian and Gabor Gaussian filters, as well as other texture-based features, were used by Reference [10] to produce the characteristics of the leaf picture. A gray-level co-occurrence matrix (GLCM) was then used. The accuracy of this approach is 93.7%. The elliptical half-Gabor wavelet, local line direction patterns, and counting-based techniques were used in this study's new leaf detection approach. The method in question is explained in reference [12]. In the Swedish, Flavia, and ICL databases, the accuracy of the method averaged at 85%. In [13], leaf texture characteristics were extracted with an accuracy of 85% on the UCI Machine Repository Dataset and 94.1% on the Swedish Leaf Dataset using Gabor features. To classify plant textures, the approach suggested in [15] uses a 128-filter bank and the aggregate distribution of Gabor filter responses. The Brodatz dataset was used to demonstrate that this technique had an identification rate of 95%. Reference [16] used a family of 64 filters, eight rotations, and eight scales to obtain an 82.93% success rate. Plant leaf categorization was carried out by Reference [17] using a library of approximately 500 leaf pictures classified into 27 groups and a filter bank with 20 filters. The classification process achieved 85.44% accuracy by integrating Gabor filters with Local Binary Patterns (LBP). Despite their complex parameterization, deep learning models perform best when applied to datasets with a large number of classes and a significant amount of data. Duong Trung et al. [18] effectively categorized ten medicinal plants supplied from Vietnam with a 98.7% accuracy rate in their study. They accomplished this using a lightweight MobileNet-based Convolutional Neural Network (CNN) with 2296 leaf images. In [19], MobileNet architecture was used, yielding a 97.43% accuracy rate. Consistent with our inquiry, earlier research [20] and [21] suggested applying CNN-based methods to the categorization of Bangladeshi medicinal plants with an accuracy of 71.3%. Raisa et al. created a dataset of 37,693 leaf pictures from ten different plant species. Six plants were correctly classified in the investigation by Musa et al., with an accuracy of 95.58%. Six therapeutic plants were categorized in a study by Reference [22] using 90 RGB leaf images gathered from India. A neural network was used to perform the categorization. Reference [23] successfully classified 40 medicinal plants with an accuracy of 96.76%. This was accomplished using 2400 images and combining feature extraction based on AlexNet with an SVM classifier. Dewedi et al. [24] used a publicly accessible dataset of 6500 photos for their investigation. They used an approach that classified 40 distinct medicinal herbs using the ResNet and SVM algorithms. This approach attained 96.80% accuracy. In [25], the authors suggested using a CNN model ensemble to correctly identify ten medicinal herbs. After 10 epochs of training on a dataset of 5000 images, the accuracy rate of the model was 99%. However, a loss value greater than 0.80 suggested that the model was overfitting. The CNN model ensemble suggested by Reference [26] exhibited overfitting and suffered a significant loss after training for 50 epochs on a dataset of 40,000 pictures. However, the accuracy of the models was 97.4%. To differentiate between medicinal and non-medicinal plants, a binary classifier based on a specially constructed Convolutional Neural Network (CNN) model was presented in [27]. With low loss, the classifier was able to achieve 90% accuracy using a dataset that included 1600 photos of 100 distinct types of medicinal plants. More than 100 epochs were used for the training procedure. However, recognizing medicinal plants without knowing their precise categorization did not satisfy the parameters of our study. A (CNN) based on DenseNet was trained to discriminate between medicinal plants and phytochemical and therapeutic plants in a study by Reference [28]. Using GoogLeNet-based methodology, Reference [29] successfully classified Ethiopian medicinal plants with an accuracy of 96.7%. A Multi-layer perceptron (MLP) classifier was employed in the method outlined in reference [30] to categorize the leaves of 25 distinct medicinal plants. The accuracy of the neural network-based method was 82.51%. According to a study by [31], the VGG16 model classified 21 plants with a 98.52% accuracy rate. In contrast, the VGG19 model performed poorly in identical challenges. The MobileNet model was trained using a dataset of 15,100 leaf pictures in a previous study [32]. Using this model, 52 distinct plant species were classified with 92% accuracy. CNN-based methods have been proven to perform well when applied to large datasets. It is important to keep in mind, though, that some of these methods have shown overfitting, especially in situations with a large number of classes. Reference [33] contains documentation on the development of a digital descriptor. The purpose of this descriptor is to solve the problem of pixel selection in the Freeman chain code by giving each leaf a unique number. Levenshtein distance graphs and Jaccard similarity index were used to evaluate this descriptor.

In the classification of medicinal plants, much effort has been made, but many challenges still remain. Most of these existing models rely on visual features, lacking context and the import of relevant excerpts from botanical text as context which greatly improves classification. Overfitting is a common problem in CNN-based models, and these models are particularly susceptible to overfitting when used with large multi-class datasets due to the lack of data regularization and sufficient data augmentation. Generalization of these models is another problematic aspect, as they are often focused on specific regional plant species making it difficult for these models to perform on worldwide datasets. Adding to the problem, current models on plant classification have been noted to work successfully on limited scope datasets, but tend to underperform on datasets containing a wide variety of plant species. Lastly, powerful CNN models require significant computational resources, making it nearly impossible for deployment in real time settings or in resource limited places. This study implements HCVINet, a hierarchical contextual visual integration network that utilizes multi-modal learning through the unification of deep learning based visual comprehension and text analysis, to resolve these issues. To address these gaps, this study presents HCVINet, a Hierarchical Contextual Vision Integration Network that utilizes multimodal learning by combining deep learning-based visual analysis with textual contextual knowledge. This method seeks to improve the classification of medicinal plants by enhancing model generalization, scalability, and computational efficiency, ultimately making these systems more accessible and effective.

Although significant progress has been made in automated medicinal plant classification using deep learning and machine learning, several critical gaps remain in the literature. Many existing studies focus primarily on visual features extracted from images, often overlooking the rich contextual and semantic information available in botanical texts and other sources. This lack of integration between visual and textual data can limit classification accuracy, especially for species with subtle morphological differences or those described in diverse contexts. Additionally, most prior works rely on region-specific or relatively small datasets, which restricts the generalizability and robustness of the proposed models. Issues such as overfitting, particularly in models trained on limited or imbalanced data, further undermine reliability. Furthermore, there is a scarcity of approaches that address the interpretability and transparency of deep learning models, which is crucial for practical adoption in scientific and pharmaceutical applications. These shortcomings highlight the need for a unified framework that combines multi-level visual features with contextual and semantic information, leverages larger and more diverse datasets, and addresses both performance and interpretability for robust medicinal plant classification.

Proposed Methodology

The Hierarchical Contextual Vision Integration Network (HCVINet) represents a novel architecture specifically engineered for the intricate classification of medicinal plants. The quality and diversity of the training data have a significant impact on performance, which may cause issues with generalization in situations that are changing or new. Furthermore, further optimization strategies can be required when scaling to huge datasets or high-resolution photos. Future studies will improve the practicality of CCINet's method, which boosts classification accuracy by recognizing relationships between categories. HCVINet, a strong framework for visual identification and contextual knowledge, performs exceptionally well in difficult classification tasks, making it especially valuable in areas such as medicine and botany that demand high precision.

The study presents a hierarchical feature extraction framework that combines HCVINet and CCINet for vision transformation tasks, improving both recognition accuracy and computational efficiency. This novel approach establishes a new standard for feature extraction in a range of vision applications. The hierarchical feature extraction network employs HCVINet for multi-scale feature extraction, greatly aiding fine-grained recognition, whereas CCINet is used for cross-channel integration to strengthen spatial dependence and classification accuracy. Moreover, the framework promotes an efficient reuse of features, showing its flexibility for many visual tasks, including object detection and segmentation. We assessed the proposed model using standard datasets and found that it outperformed existing methods. However, it's important to recognize the limitations of this study; the hierarchical architecture, while effective, adds computational overhead that can complicate real-time applications. The success of the approach heavily depends on the quality and diversity of the training data, which may lead to generalization issues in dynamic or unfamiliar environments. Additionally, ensuring scalability for large datasets or high-resolution images might require further optimization strategies. Tackling these challenges in future research will enhance the reliability and applicability of our proposed method.

Hierarchical Feature Extraction Network

In this section, the HCVINet methodologies for the enhancement of Vision Transformation are described in detail, and the steps are depicted in Algorithm 1. The overall architecture of HCVINet is shown in Fig. 1. The proposed work includes various methods, namely, multilevel image feature extractor, combined feedforward convolutional layer, and hierarchical attribute-selection mechanism.

Figure 1 shows the proposed architecture, which is as follows.

Hierarchical Feature Extraction Network: The first step includes the keyword image which is processed by a feature extractor that works on multilevels. With this, it is possible to extract various levels of features, such as simple textures to more intricate designs.

Intra-Attention: Self attention is applied to the extracted features, which helps process the data. This allows the model to share attention over its relevant portions, giving importance to the features important for classification of the image.

Convolutional Feed-Forward Network or ConvFFN: This new unit integrates convolutional processes with feedforward network layers. In this case, the data is totally flattened, or 'panned,' then turned and molded to the appropriate dimensions to undergo new processes.

Multi-Level Feature Selection Mechanism: Data resulted from the convolutional processes was subjected to a hierarchical attribute selection procedure after the convolutional strategies were completed. This mechanism focuses on the features or attributes that are most important as class labels with respect to their class importance which are learned from the data.

Encoding Networks: The selected attributes are then passed through one or more encoder stages. Encoders help condense the data into a more manageable and informative set of features, reduce dimensionality, and emphasize key relationships.

Classification: The final stage of architecture involves classification. The output from the encoder was used by a classifier to determine the category or class of the input image based on the learned features and selected attributes.

In the proposed technique, features in shape, color, and texture out of the leaf from a combination of image processing techniques and deep learning methods. Contour detection and edge-based segmentation have been adopted in order to extract shape features, which ensure very accurate representation of structural features. Color features are obtained using color histograms and HSV (Hue, Saturation, Value) space conversion, which helps in distinguishing different varieties of leaves according to pigmentation differences. Furthermore, texture features are obtained using Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP), thus offering a more robust analysis of surface patterns. Finally, the extracted features will be passed through the classification model for precise identification and classification purposes. A systematic model building approach is implemented for classification of the plant which is essential for accurate medicinal plant classification. The features are input into a classification module which is specified within a fully connected network. In this module the several dense layers strive to learn and improve the separability of the feature embeddings within the trained model. In the last classification layer softmax activation is employed for defining the probability distribution over the classes and the loss function is specified to be class assigning categorical cross entropy. The model is trained using the Adam optimizer which performs learning rate decay and momentum boosting to enhance convergence. To improve generalization and minimize overfitting the dropout regularization is applied through turning off some of the neurons during training. This specific method of classifying plants increases accuracy of the model together with its scalability and efficiency when the model is required to differentiate between the species of the medicinal plants. The preprocessing techniques used include data normalization, noise removal, and feature scaling to enhance the quality of input data before

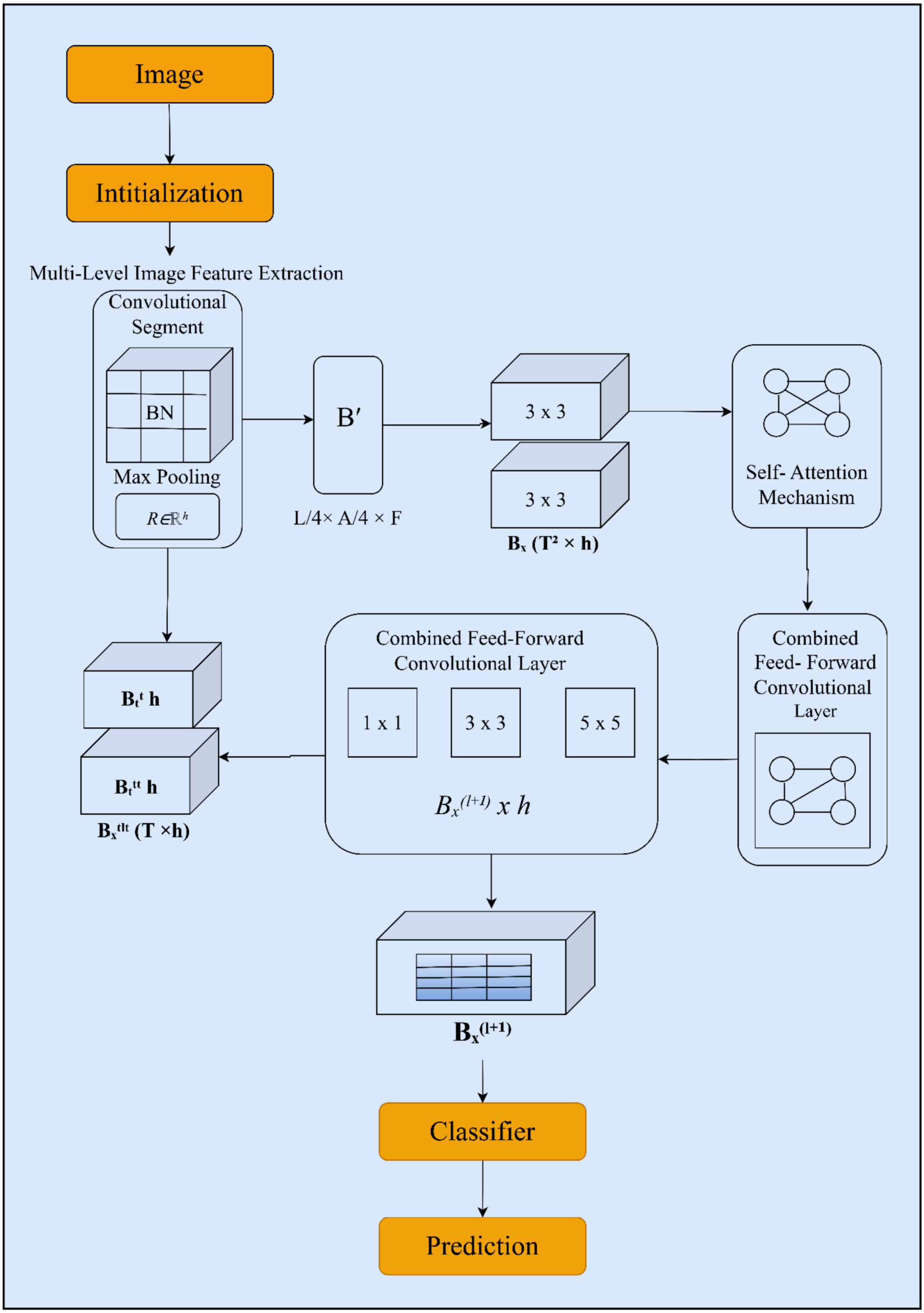

Fig. 1 Hierarchical Feature Extraction Network architecture

applying machine learning models. Categorical data encoding is performed to convert non-numerical attributes into a suitable format for model training. Missing value imputation is applied to ensure data completeness, reducing the risk of biased predictions. Additionally, outlier detection is used to eliminate anomalies that may affect model performance. These preprocessing steps help improve the accuracy and reliability of kidney disease prognosis using machine learning techniques.

Multi- Level Image Feature Extractor

The multilevel image feature extractor was designed to handle the tokenization process. Instead of directly extracting patches from the initial images, this layer retrieves patches from attributes produced by the selected kernel. As shown in Fig. 1, the multilevel image feature extractor consists of convolutional segments and a kernel-selection component. To illustrate, consider image $B belongsto \beta^{L \times A \times 3}$, and represent the height and width of the image, respectively. The process begins by applying convolutional operations to extract features at multiple levels, thereby capturing both simple and complex patterns in the image.

Step 1: We used a convolutional segment with 7×7 kernels to extract the features from the image. Additionally, a 3×3 kernel was used for pooling with a stride of two to down-sample the output attribute maps. This process can be described as follows.

$$B' = MaxPool(batchnorm(Conv(b))) \tag{1}$$

Output: Considering the above Eq. 1, $B' belongsto \beta^{\frac{L}{4} \times \frac{A}{4} \times H}$, where H the output of the convolutional segment with 32 enriched channels. In addition, a batch normalization layer was applied to stabilize and accelerate the training process.

Step 2: To improve the efficiency of capturing multilevel attributes from images, we implemented a kernel selection component. This component includes two convolutional phases applied to the features obtained from the previous convolutional layers. In our experimental setup, the reduction ratio, t for the selected kernel was set to 4. The two convolutional phases used kernels of dimensions 3×3 and 5×5, respectively.

$$zB'' = ChooseKern\left(B'\right) \tag{2}$$

Output: B'' $with\, dimensions \beta^{\frac{L}{4} \times \frac{A}{4} \times H}$.

Step 3: B'' is redesigned into a sequence of patches that are flattened

$$B_t'' = \beta^{T \times \left(V^2 \times J\right)} \tag{3}$$

where $R = (LA)\left(4 \times 4 \times T^2\right)^{-1}$ that is used to denote the number of patches that are developed, and (T, T) represents the dimension of an individual patch. To maintain a constant generation of tokens for the Vision Transformer, we equate R to 4. In addition, a fixed vector for attributes was utilized for all layers. Hence, flattened patches are utilized, which are denoted as z_r'' for a fixed vector having a size set to.

Output: f =384 for this study. Another class of tokens is used in the series that represents the class of the input image used, which leads to the dimension of the input token $B_x belongsto \mathbb{T}^{(R+1) \times h)}$.

The main advantage of using kernel selection methods in the proposed study is that they can adapt to the outcome of data for various levels on the resulting output via the changing method of selection. Likewise, a receptive domain that is adaptive and similar to the visual mechanism of humans is used, which enhances the accurate classification of images. In addition, considering the high capacity to grasp local data, the model does not require location embeddings to aid the training process in the model.

Combined Feed-forward Convolutional Layer

To enhance the capabilities of the traditional Vision Transformer model in capturing both local and multilevel features, the standard feed-forward network was replaced with a combined feedforward convolutional layer. The structure of this module is illustrated in the figure above. This layer performs the described processes to improve feature extraction and representation of the model. First, we divide $B_x^l belongsto \beta^{(R+1) \times h}$ which is an outcome of the self-attention model, as patches of tokens $B_t^l belongsto \beta^{R \times h}$ as well and a class of token $B_g^l belongsto \beta^{1 \times h}$, wherein the no attributes are denoted by $R + 1$. Additionally, the 1-dimenion of the token patches are modified into 2-dimensions expressed as $B_t^w belongsto \beta^{(R)^{\frac{-1}{2}} \times (R)^{\frac{-1}{2}} \times h}$. Moreover, to improve the model's ability to understand local characteristics, a convolution process is carried out using the three 2-dimension token kernels. This focuses on $B_t^{h(1)} belongsto \beta^{(R)^{\frac{-1}{2}} \times (R)^{\frac{-1}{2}} \times (i \times h)}$. Through this process, the network's capacity to understand the characteristics in a higher-dimensional space, both locally and globally, is improved. Next, we investigate how different convolutional kernel sizes affect the network performance. Furthermore, the reduced dimensionality comes from $B_t^{p(2)} belongsto \beta^{(R)^{\frac{-1}{2}} \times (R)^{\frac{-1}{2}} \times h}$ using a 1×1 operation for convolution. Finally, the 2-dimensional attributes are reformed to 1-dimension that is denoted as $B_t^{l+1} belongsto \beta^{R \times h}$ and the result is integrated with the token to attain $B_x^{l+1} belongsto \beta^{(R+1) \times h}$.

The inclusion of depth-based convolutions within the combined feed forward convolution layer leads to an increase in computational costs. It should be noted that the cost is relatively small compared with the total number of parameters in the model.

Hierarchical Attribute Selection Mechanism

Step: 4 Compute attention matrix: The Self-Attention Mechanism is used by the traditional Vision Transformer model to efficiently model and detect the discriminating zones. Deeper encoding levels may cause layers containing local characteristics to be overlooked. The establishment of a system for hierarchical attribute selection is suggested. The following process was used to calculate the attention matrix value for each layer:

$$D = [d_0, d_1, d_2, \ldots \ldots, d_{P-1}] \tag{4}$$

Considering the above equation, the number of layers is given as N. For an effective attention mechanism, attributes are progressively gathered and enhanced for every layer using Hadamard products.

Gather and enhance attributes for every layer: First, the weights in this case are combined for every head inside every layer, and are expressed as follows:

$$D_p = \left[D_p^0, D_p^1, D_p^2, \ldots, D_p^R\right] = \odot_{m=0}^{O-1} e_p^m p belongsto 1,2, \ldots, P-1 \tag{5}$$

$$d_p^m = \left[e_p^{m_q}, e_p^{m_1}, e_p^{m_2}, \ldots ., e_p^{m_R}\right] belongsto 0,1, \ldots, m \tag{6}$$

In this case, m was used to express the head count in the self-attention mechanism. Token classification, denoted as E_p^q has an important class of data. Hence, if there exists an increased correlation in the token during classification of the token, it is recommended that the visual data of that token aid the network more accurately in the prediction of the category. Hence, we categorized E_p^q as well as the selection R tokens with the highest scores.

Select top-ranked attributes for classification: The tokens chosen by the Hierarchical Attribute Selection Mechanism for the $p-th$ layer are expressed as follows:

$$d_p^{local} = [d_p^1, d_p^2, \ldots, d_p^Q] \tag{7}$$

In this case, the number of chosen characteristics is denoted as Q. We also examine how the network performs differently when it has different scores for hyperattribute Q.

Final classification using the last-layer class token: Additionally, following attribute integration, the transformer's last layer, P, is formulated as follows:

$$d_j = [d_{P-1}^0; d_1^{local}; d_2^{local}; \ldots ..; d_{P-1}^{local}] \tag{8}$$

Finally, the class token from the last layer is used to perform the classification process. The simplicity of the suggested approach is emphasized because it does not require any additional learning parameters.

Contextual-Correlation Integrated Network (CCINet)

The CCINet is an advanced NLP-based totally framework designed to enhance the type accuracy of medicinal vegetation by using correctly integrating contextual records with category correlations. CCINet consists of two linked components: the CCRI Net and the CCRN. Together, these elements create a robust system that comprehends the semantic context of medicinal plants and the complex relationships between their categories.

Category-Aware Contextual Integrator (CACI) Net

The CACI Net addresses the challenge of semantic property capture from the input data and linked categories as seen in the classification of the medicinal plants. This network further expands the model's capability to accurately identify the medicinal plants by combining the text and image information, which results in the context-sensitive representation of each plant species. For this purpose, we used advanced BERT model which extracts embedded features from the input data such as descriptions of plant characteristics and their associated categories which include medicinal uses and geographic distribution.During the training phase, the input data and categories were extracted simultaneously to fully capture the relationships between semantic features and categories, ensuring accurate representation of textual features in the final layer of BERT:

$$k = concat(k_{cls}^{(-1)}, k_{cls}^{(-2)}, k_{cls}^{(-3)}, k_{cls}^{(-4)}, k_{cls}^{(-5)}) \tag{9}$$

The effectiveness of CCRI Net is driven by two key factors:

1. **Semantic Integration:** The step involves enhancing input data representation by incorporating semantic data specific to each category, thereby identifying the most relevant features for each category.
2. **Contrastive Learning:** This approach organizes semantic data in a coherent manner, considering category feature encodings as positive examples for both the category and the input data. This process enhances the model's capability to understand the connection between the data and its respective categories, ultimately improving classification accuracy.

Category Correlation Refinement Network (CCRN)

The CCRN is built on the idea of analyzing the inter-relationships between different categories in order to classify medicinal plants in a more accurate way. Instead of only having singular predictions, CCRN takes a more holistic approach by focusing on the interconnectedness of the medicinal, ecological, and ethnomedical aspects of the plant. Such an integrative approach is bound to lead towards a more accurate and contextual classification. The network consists of two main components that increase the computation accuracy and reliability of the predictions.

- **Raw Category Predictions:** Based on the input data, the model first produces preliminary predictions, denoted by b. These predictions could relate to a no's of characteristics in the context of medicinal plants, including their ecological roles, geographic distribution, or therapeutic qualities. This first classification is the basis for subsequent improvement, guaranteeing a more precise and contextually aware comprehension of each plant's traits.
- **Correlation Enhancement Unit (M_n):** This computational unit receives the primitive predictions and enhances them through the consideration of connections between categories for refined predictions, labeled as B.

The CCRN model is represented mathematically to refine category predictions by leveraging relationships among medicinal plant attributes:

$$B = b + M_n(b) \tag{10}$$

Here, B represents the output after applying CCRN, and b is the raw category prediction. Function $M_n(b)$ is the transformation applied by the correlation enhancement unit, which adjusts the raw predictions based on the learned correlations between medicinal plant categories.

Algorithm 1 Hierarchical Contextual Vision Integration Network

Input: An image $B belongsto \beta^{L \times A \times 3}$, the length as well as breadth of the image used as input is denoted as L and A for 3 color channels (RGB)

Output: Return the predicted class label

Step 1 — **Initialization**
Initialize the model parameters including convolutional filters, kernel sizes, and weights for the attention mechanism

Step 2 — **Multi-Level Image Feature Extraction**
Convolutional Segment:
Apply a convolution operation with a kernel size of 7×7 on the input image B
- Apply batch normalization
- Apply max pooling with a kernel size of 3×3 stride 2
- Result B' *belongs to* $\beta^{\frac{L}{4} \times \frac{A}{4} \times H}$, where $F = 32$

2. **Kernel Choosing Component:**
Apply two convolution operations on z′ with kernel sizes 3*3 and 5*5
Result: B' *belongs to* $\beta^{\frac{L}{4} \times \frac{A}{4} \times H}$,
3. **Patch Generation:**
Reshape B″ into a sequence of flattened patches $B''_t belongsto \beta^{R \times (T^2 \times H)}$, where $R = (LA)\left(4 \times 4 \times T^2\right)^{-1}$ and $T = 4$

Step 3 — **Self-Attention Mechanism**
Pass B''_t through the self-attention mechanism
Result: Tokens $B_x belongsto \beta^{(R+1) \times h)}$ where $f = 384$ and the extra dimension accounts for the class token

Step 4 — **Combined Feed-Forward Convolutional Layer:**
- **Token Separation:**
Separate B_x into token patches $B^l_t belongsto \beta^{R \times h}$ and class token $B^l_g belongsto \beta^{1 \times h}$
Reshape B^l_t into 2d patches $B^w_t belongsto \beta^{(R)^{\frac{-1}{2}} \times (R)^{\frac{-1}{2}} \times h}$
- **Convolutional Operations:**
Apply a 2D convolution with different kernel sizes (e.g., 1×1, 3×3, 5×5) to capture local features
Reconfigure 2D patches back into $B^{l+1}_x belongsto \beta^{R \times h}$.
Concatenate with class token to form $B^{l+1}_x belongsto \beta^{(R+1) \times h}$.

Step 5 — **Hierarchical Attribute Selection Mechanism:**
- **Attention Calculation:**
Compute the attention matrix D for each layer n using Hadamard products of attention weights across heads
Select top Q tokens based on the attention scores, d^{local}_p
- **Final Layer Selection:**
Collect selected tokens from all layers to form d_j
Use the class token from the last layer for final classification

Step 6	**Classification** • Use the class token to predict the category of the input image

Performance Evaluation

The Hierarchical Contextual Vision Integration Network (HCVINet) utilizes a sophisticated combination of multi-level convolutional feature extraction, self-attention mechanisms, and a CCINet to enhance the classification performance. HCVINet obtained better metrics by combining visual features with semantic information from NLP. Hierarchical attribute selection and attention-based refinement further optimize feature relevance, enabling the model to accurately distinguish between closely related medicinal plant species, making it a robust solution to complex classification challenges.

Dataset Details

Flavia Dataset

The Flavia dataset is composed of 1907 samples from 33 native plant species of the region's fauna which are native to the Yangtze Delta region of China. The dataset may be composed of Chinese plants, but it is often used as a standard for testing leaf classification methods. The approach that has been proposed is independent of the dataset and is applicable to Indian medicinal plants. In the future, it will be necessary to verify the method on an Indian fauna specialized dataset in order to prove its composite nature [34].

Indian Medicinal Dataset

Samples from over 20 locations in and around the Karnataka districts of Mysore and Mandya, as well as other locations in Kasaragod, Kerala, comprise the Real-Time Plant (RTP40) and Real-Time Leaf (RTL80) databases [35]. Forty distinct plant species were included in plant-level samples obtained from the RTP40 collection. The RTL80 dataset, on the other hand, consists of leaf-level samples from eighty different plant species. The samples were collected in both indoor and outdoor environments using various devices with varying resolutions. The items in question have a variety of backdrops, projection angles, distances, and noise-related issues, such as haze, blur, shadows, and high and low contrast. Thirty-six common plant species from all three datasets were used to validate the practical implementation of the HCVINET model. Figures 2 and 3 show haphazard samples collected from various plant species.

Results

This section evaluates the HCVINet model considering various metrics on the Flavia dataset with the existing model [36] and Indian Medicinal Dataset with existing models [37].

The chart displays the accuracy of the different AI models on the Flavia dataset, which is a popular choice for leaf

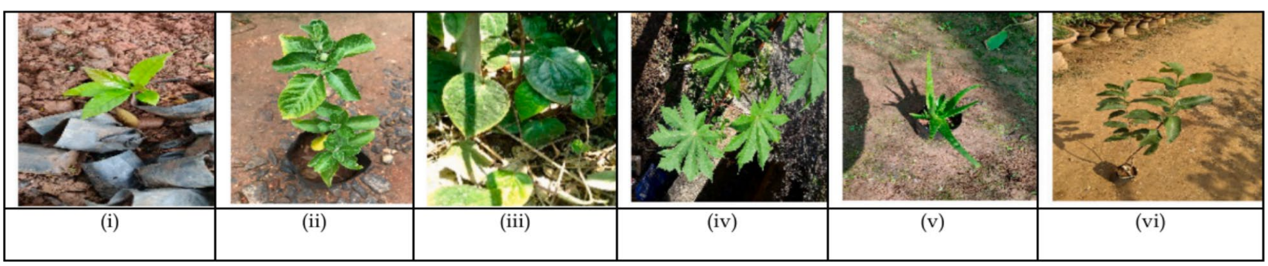

Fig. 2 Medicinal plant

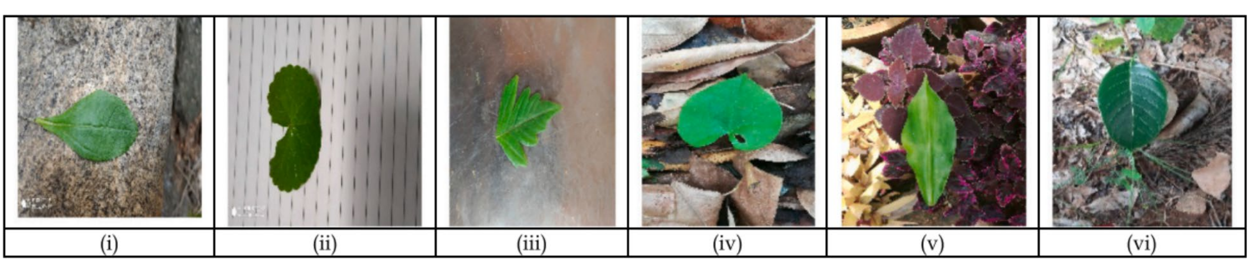

Fig. 3 Medicinal leaves

species classification. Most models exhibit strong performance, with an accuracy above 85%. Methods such as Random Forest and Gradient Boosting, when combined with SVM, are particularly successful in dealing with high-dimensional image data. The HCVINet model, likely enhanced with sophisticated optimizations or specific features, also shows outstanding accuracy, reflecting its strong adaptability to the dataset. These results showcase the effectiveness of both traditional AI strategies and innovative approaches in accurately classifying leaf images. To enhance the evaluation, further comparisons were performed with an existing model [36], yielding deeper insights into overall performance (Fig. 4).

The high accuracy of these models shows their effectiveness in correctly identifying leaf images while also lowering the rate of false positives, which is essential for accurate classification tasks. The precision chart for different machine learning models applied to the Flavia dataset shows consistently high performance, with all models attaining over 90% precision. In particular, ensemble methods such as Random Forest and Gradient Boosting, along with SVM and HCVINet, are noteworthy, achieving precision rates nearing 100% (Fig. 5).

Models such as Logistic Regression, k-nearest Neighbors, Gradient Boosting, Multilayer Perceptron, and the HCVINet Model achieved recall rates at or above 90%. This high recall demonstrates the effectiveness of the models in minimizing false negatives and ensuring that most positive cases are correctly identified. The recall chart for different AI models tested on the Flavia dataset shows that these models are

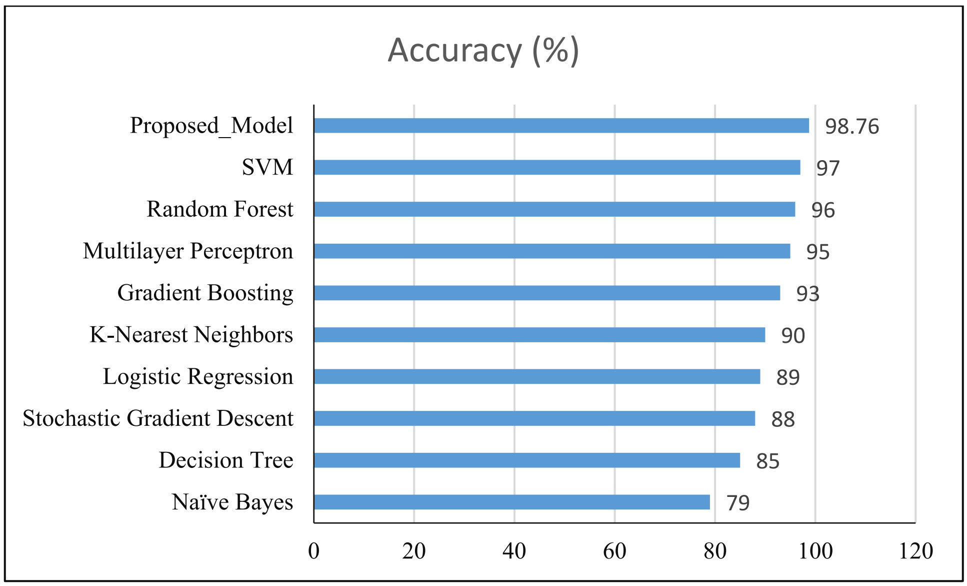

Fig. 4 Accuracy comparison on flavia dataset

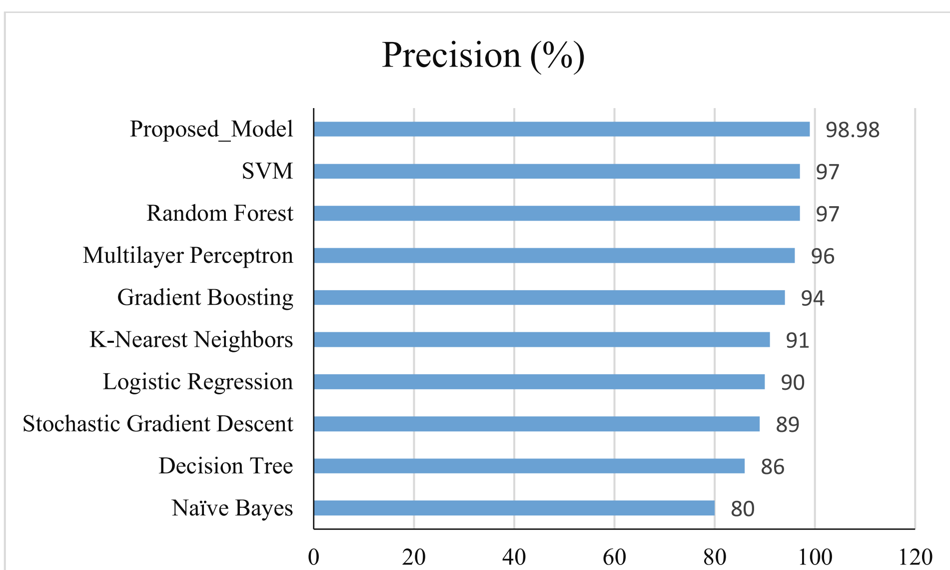

Fig. 5 Precision comparison on flavia dataset

highly effective at identifying relevant instances, with recall rates mostly exceeding 80%. This suggests that the models are consistently capturing a significant portion of the correct classifications, minimizing the chances of missing important dataSuch performance is crucial for applications where failing to detect a true positive could have significant implications, such as precise botanical classification (Figs. 6 and 7).

The chart illustrates a comparative analysis of the performance metrics between the existing and HCVINet models for classifying medicinal plants on the Indian medicine plant dataset discussed earlier. The HCVINet model significantly surpassed the existing model in all evaluated aspects. Particularly, it reached an accuracy of about 84%, which is significantly higher than the current model's accuracy of 75%. Likewise, the recall rate of the HCVINet model was about 84% in comparison to 76% of the existing model. For the precision, the HCVINet model shows considerable improvement with a score of about 85%, while the existing model scores 76%. The F1-score also shows almost the same amount of improvement. The effectiveness in the performance analysis showed that the HCVINet model outperformed the existing one in terms of precision and reliability for the identification and classification of the medicinal plants (Fig. 8).

The bar chart shows a comparison of the two models' performances for the classification of medicinal leaves; whereby "HCVINet" is the best model in all metrics when compared to an existing model. The HCVINet model stood out in its accuracy of about 96%, while the older model scored only 75%. It also increased the recall significantly to about 96%,

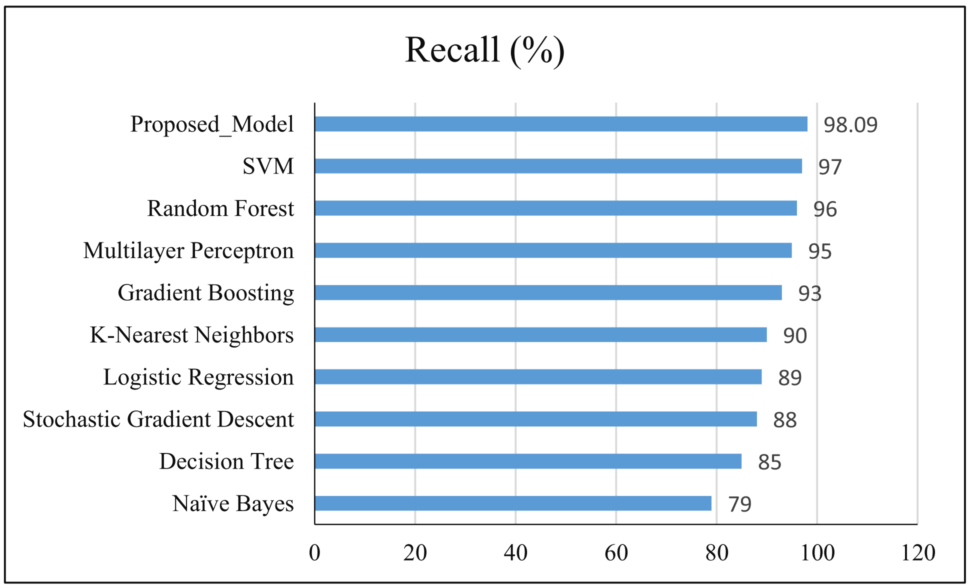

Fig. 6 Recall comparison on flavia dataset

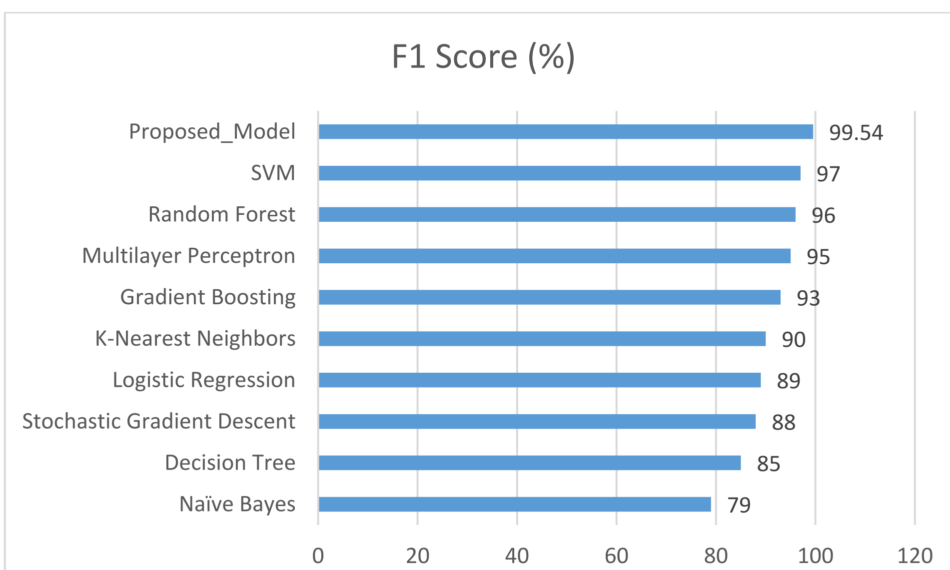

Fig. 7 F1-score comparison on flavia dataset

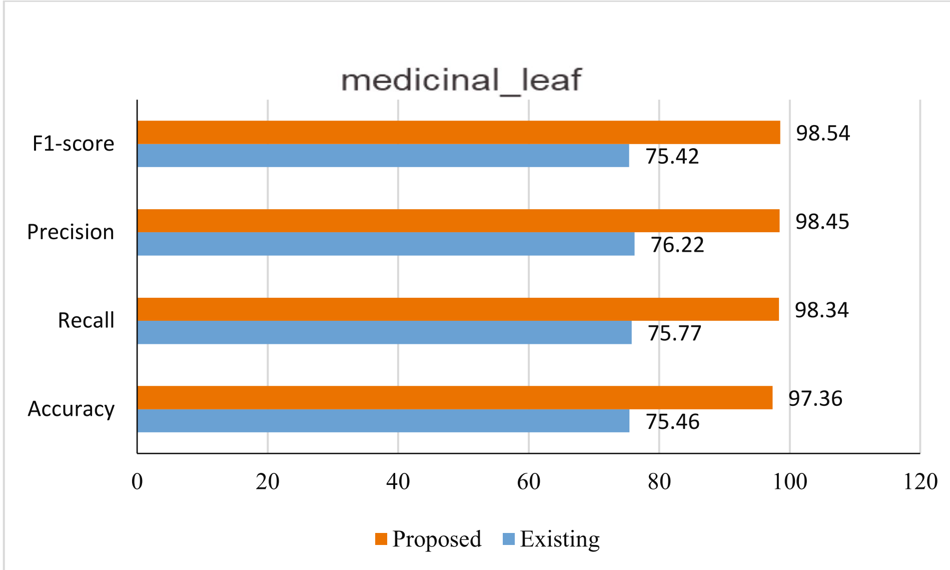

Fig. 8 Metrics comparison Indian medicinal leaf dataset

as opposed to the existing model with only 76%. Precision increased again; it scored about 97% while the old model achieved only 76%. The F1-score of the HCVINet model is also around 97%, much higher than the existing one, which is just 75%. Altogether, these results manifest that the HCVINet model outperforms medicinal-leaf classification with accuracy and reliability, thereby holding many potential applications in botanical research and practical herbal medicine (Fig. 9).

Comparative analysis

The analysis of the performance metrics revealed a notable improvement in the HCVINet model compared to the existing model. The accuracy of the HCVINet model ascertained increased by 15.29%, ranking it higher than competing models. The recall also improved by 18.45%, demonstrating the model's heightened ability to capture true positive cases. Furthermore, precision which reflects the measure of relevance predictedly improved by 14.33%. The increase indicates a decrease in false positives given by the model. The value of F1-Score not only increased by 13.46%, but also demonstrates the overall success of the HCVINet model in providing accurate and adequate predictions. These improvements imply that the HCVINet model outperformes the existing model on multiple metrics as presented in Table 1.

The comparative analysis between the existing and HCVINet models showed a substantial enhancement across all performance metrics, indicating the superior efficacy of the HCVINet model. The accuracy metric

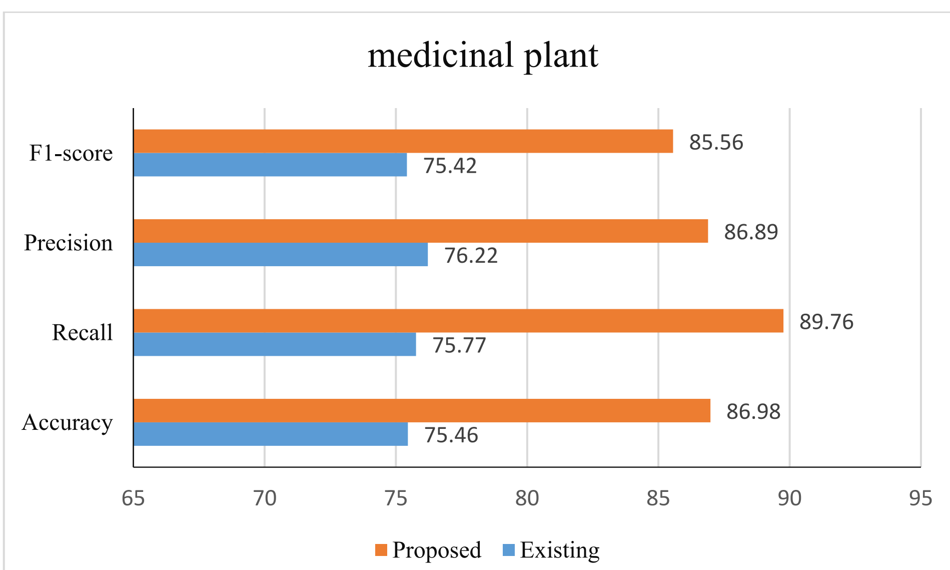

Fig. 9 Metrics comparison on Indian medicinal plant dataset

Table 1 Performance comparison through proposed model enhancement over existing one on Indian medicinal plant

Metric	Existing Model (%)	HCVINet Model (%)	Enhancement
Accuracy	75.46	86.98	0.1529
Recall	75.77	89.76	0.1845
Precision	76	86.89	0.1433
F1-Score	75.42	85.56	0.1346

Table 2 Performance comparison through proposed model enhancement over existing one on Indian medicinal plant

Metric	Existing Model (%)	HCVINet Model (%)	Enhancement
Accuracy	75.46	97.36	0.2906
Recall	75.77	98.34	0.2977
Precision	76.22	98.45	0.2911
F1-Score	75.42	98.54	0.3067

increased by 29.06%, as noted in Table 2, suggesting that the HCVINet model was overall more accurate in making predictions. The recall, which is the reduction in missed detections, improved by 29.77%, indicating that the model was able to identify all relevant cases. Precision, which is the enhancement in specificity, resulted by a reduction in false positives, experienced a 29.11% enhancement, suggesting that the HCVINet model increase in true positives proportioned to the total positives predicted. The F1 score experienced an increase of 30.67% which, as previously stated, is the balance increase in recall and precision. Collectively, all of these suggest that the HCVINet model outperformed the previous one and was more accurate, sensitive, and reliable. The data makes the model seem to provide better predictions than the other existing models. A single improvement is noteworthy and that is makig the HCVINet enhancements more effective and useful in future research. The results section includes visual elements such as a confusion matrix, training-validation graph, AUC (area under the curve) curve, and ROC (receiver operating characteristic) curve to demonstrate the model's effectiveness. The confusion matrix presents a detailed breakdown of predictions, highlighting true positives, false positives, true negatives, and false negatives. The training-validation graph illustrates the learning progression, showing trends in accuracy and loss over time. The AUC and ROC curves evaluate classification performance by depicting the model's ability to differentiate between classes. These visualizations enhance the credibility of the findings and provide a comprehensive assessment of performance.

Conclusion

The new approach to the classification of medical plants is centered around the development of the Hierarchical Contextual Vision Integration Network (HCVINet), which is the focus of this study. HCVINet is a combination of a Multi Hierarchical Feature Extraction Neural Network and Contextual-Correlation Integrated Network CCINet that semantically processes pictures using Natural Language Processing. Furthermore, the combination of text and image information gives the HCVINet model the sophisticated capabilities of accurately classifying and identifying medicinal plants. To accomplish adequate comparison, HCVINet was tested with datasets that achieved higher metrics than existing models. HCVINet has the unprecedented capability of synthesizing multiple types of data: visual, contextual, and textual. This feature advancement exhibits significant growth in pharmacology, botany, and traditional medicine. HCVINet is a new frontier in the development of more reliable systems for plant identification that enhances the accuracy and efficiency of scientific research in these disciplines. These efforts contribute toward successfully planting systems in different countries.

Acknowledgements Authors acknowledge the support from Nitte Meenakshi Institute of Technology, Bengaluru and NMAMIT, Nitte for the facilities provided to carry out the research.

Funding This research was conducted without any financial support from funding organizations.

Data Availability The labeled datasets supporting the findings of this study are available from the corresponding author upon request.

Declarations

Conflicts of Interest All authors confirm that they have no conflicts of interest to disclose.

References

1. Dutta M, Gupta D, Juneja S, Nauman A, Muhammad G. Comparative growth analysis of onion in deep water culture and soil based systems: enhancing medicinal plant cultivation in urbanized environments. IEEE Access. 2024;12:38202–18. https://doi.org/10.1109/ACCESS.2024.3373787.
2. Thanikkal JG, Dubey AK, Thomas MT. An efficient mobile application for identification of immunity boosting medicinal plants using shape descriptor algorithm. Wireless Personal Commun. 2023;2:1189–205.
3. Wang C, Zhang X, Shi T, Zhang C, Li M. Classification of medicinal plants Astragalus Mongholicus Bunge and Sophora Flavescens Aiton using GaoFen-6 and multitemporal sentinel-2 data. IEEE Geosci Remote Sens Lett. 2022;19:1–5. https://doi.org/10.1109/LGRS.2021.3120125.

4. Thanikkal JG, Dubey AK, Thomas MT. A novel edge detection method for medicinal plant's leaf features extraction. Int J Syst Assur Eng Manag. 2023;1:448–58.
5. Viet KLD, et al. Medicinal plants identification using federated deep learning. Proc Comput Sci. 2024;234:247–54.
6. Kiflie MA, Sharma DP, Mesfin A. Deep learning for medicinal plant species classification and recognition: a systematic review. Front Plant Sci. 2024;14:1286088.
7. Adibaru KM, et al. EfficientNet ensemble learning: identifying Ethiopian medicinal plant species and traditional uses by integrating modern technology with ethnobotanical wisdom. Computers. 2024;2:38.
8. Muneer A, Fati SM. Efficient and automated herbs classification approach based on shape and texture features using deep learning. IEEE Access. 2020;8:196747–64.
9. Khan R, Preenon B, Krutanjali P. Empowering global health with AI: using NLP to extract medicinal plants and disease-fighting compounds from PubMed. In: International conference on advances in nano-neuro-bio-quantum (ICAN 2023). Atlantis Press, 2023.
10. Rai M, Rivas P. A review of convolutional neural networks and Gabor lters in object recognition. In: Proc. Int. Conf. Comput. Sci. Comput. Intell. (CSCI), 2020, pp. 15601567, https://doi.org/10.1109/CSCI51800.2020.00289.
11. Meeradevi SBJ, Seema S, Rohith S. A visual computing unified application using deep learning and computer vision techniques. Int J Interact Mobile Technol. 2024;18(1):59–74. https://doi.org/10.3991/ijim.v18i01.42673.
12. Kong AWK. An analysis of gabor detection (lecture notes in computer science: lecture notes in artifcial intelligence and lecture notes in bioinformatics), vol. 5627. Berlin: Springer; 2009. p. 6472.
13. Hubel DH, Wiesel TN. Brain and visual perception: the story of a 25-year collaboration. Oxford: Oxford Univ. Press; 2012.
14. Akinpelu SA, Olasoji OE, Akindolani A, Adeyanju KI, Ajagbe SA, Taiwo GA. Detection of cassava plant disease using deep transfer learning approach. Paradigmplus. 2025;6(1):1–12. https://doi.org/10.55969/paradigmplus.v6n1a1.
15. Obaidullah SM, Das N, Kaushik R. Convolution based technique for Indic script identification from handwritten document images. Int J Image Graph Signal Process. 2015;7(5):49.
16. Sliti O, Gmati C, Benzarti F, Amiri H. Robust object tracking using log-Gabor lters and color histogram. In: Proceedings of the 3rd international conference on pattern recognition applications and methods (ICPRAM), 2014. pp. 687694
17. Munawar HS, Aggarwal R, Qadir Z, Khan SI, Kouzani AZ, Mahmud MAP. A Gabor fillter-based protocol for automated image-based building detection. Buildings. 2021;11(7):115. https://doi.org/10.3390/buildings11070302.
18. Duong-Trung N, Quach LD, Nguyen MH, Nguyen CN. A combination of transfer learning and deep learning for medicinal plant classification. In: Proc. 4th Int. Conf. Intell. Inf. Technol., 2019, pp. 83–90.
19. Valdez DB, Aliac CJG, Feliscuzo LS. Medicinal plant classification using convolutional neural network and transfer learning. In: Proc. IEEE Int. Conf. Artif. Intell. Eng. Technol. (IICAIET), 2022, pp. 1–6.
20. Akter R, Hosen MI. CNN-based leaf image classification for Bangladeshi medicinal plant recognition. In: Proc. Emerg. Technol. Comput., Commun. Electron. (ETCCE), 2020, pp. 1–6.
21. Musa M, Arman M, Hossain M, Thusar A, Nisat N, Islam A. Classification of immunity booster medicinal plants using CNN: a deep learning approach. In: Proc. Int. Conf. Adv. Comput. Data Sci., Nashik, India, Apr. 2021, pp. 244–254.
22. Saikia AP, Hmangaihzuala PV, Datta S, Gope S, Deb S, Singh KR. Medicinal plant species classification using neural network classifier. In: Proc. 6th Int. Conf. Commun. Electron. Syst. (ICCES), 2021, pp. 1805–1811.
23. Dileep MR, Pournami PN.AyurLeaf: a deep learning approach for classification of medicinal plants. In: Proc. TENCON IEEE Region 10 Conf. (TENCON), 2019, pp. 321–325.
24. Diwedi HK, Misra A, Tiwari AK. CNN-based medicinal plant identification and classification using optimized SVM. Multim Tools Appl. 2023;83(11):33823–53.
25. Uddin AH, Chen Y-L, Borkatullah B, Khatun MS, Ferdous J, Mahmud P, Yang J, Ku CS, Por LY. Deep-learning-based classification of Bangladeshi medicinal plants using neural ensemble models. Mathematics. 2023;11(16):3504.
26. Bahri A. Dynamic CNN combination for Morocco aromatic and medicinal plant classification. Int J Comput Digit Syst. 2022;11(1):239–49.
27. Kyalkond S, Aithal S, Sanjay V, Kumar P. A novel approach to classification of ayurvedic medicinal plants using neural networks. Int J Eng Res Technol. 2022;11(1):2278.
28. Pukhrambam B, Sahayadhas A. Advanced medicinal plant classification and bioactivity identification based on dense net architecture. Int J Adv Comput Sci Appl. 2022;13(6):1–7.
29. Berihu M, Fang J, Lu S. Automatic classification of medicinal plants of leaf images based on convolutional neural network. In: Proc. CCF Conf. Big Data, 2022, pp. 108–116.
30. Kumar G, Kumar V, Hrithik A. Herbal plants leaf image classification using machine learning approach. In: Intelligent Systems and Smart Infrastructure. Boca Raton: CRC Press; 2023. p. 549–58.
31. Sharma P. Automatic classification and diseases identification of Indian medicinal plants using deep learning models. ECS Trans. 2022;107(1):4631–9.
32. Widneh M, Workneh A, Alemu A. Medicinal plant parts identification and classification using deep learning based on multi-label categories. Ethiopian J Sci Sustain Develop. 2021;8:96–108.
33. Thanikkal JG, Dubey AK, Thomas MT. Unique shape descriptor algorithm for medicinal plant identification (SDAMPI) with abridged image database. IEEE Sensors J. 2020;20(21):13103–9. https://doi.org/10.1109/JSEN.2020.3002909.
34. Wu SG, Bao FS, Xu EY, Wang YX, Chang YF, Xiang QL. A leaf recognition algorithm for plant classification using probabilistic neural network. In: Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT), Giza, Egypt, 2007, pp. 11–16, https://doi.org/10.1109/ISSPIT.2007.4458016.
35. Pushpa BR, Shobha Rani N. DIMPSAR: dataset for Indian medicinal plant species analysis and recognition. Data Brief. 2023;49:109388. https://doi.org/10.1016/j.dib.2023.109388.
36. Twum F, Missah YM, Oppong SO, Ussiph N. Textural analysis for medicinal plants identification using log Gabor filters. IEEE Access. 2022;10:83204–20. https://doi.org/10.1109/ACCESS.2022.3196788.
37. Pushpa BR, Shobha Rani N, Chandrajith M, Manohar N, Nair SSK. "On the importance of integrating convolution features for Indian medicinal plant species classification using hierarchical machine learning approach. Ecol Informatics. 2024;81: 102611. https://doi.org/10.1016/j.ecoinf.2024.102611.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.