Multimedia Tools and Applications (2024) 83:13775–13789
https://doi.org/10.1007/s11042-023-15931-8

A weighted KNN model for identification of medicinal plant species

Savitha Patil[1] · **M. Sasikala**[1]

Received: 1 February 2022 / Revised: 2 May 2023 / Accepted: 29 May 2023 /
Published online: 10 July 2023

Abstract

Medicinal plants can provide immense contribution towards the growth of modern medicine and pharmaceutical industries to protect people from current deadly diseases like cancer and cardiovascular diseases. However, presence of thousands of plant species globally and similarity in their features like color, texture and shape makes their identification critical and immensely challenging. Moreover, utilization of traditional methods to classify plant leaf under expert's guidance is costly, challenging and time taking process. Therefore, in this article, a Weighted KNN Classification (WKNNC) Model is adopted for the accurate identification of plant leaf images based on machine learning techniques. High quality morphological and discriminative features are obtained by using Region of Interest (ROI) of the images, which is extracted from segmentation process. The proposed WKNNC model works upon Local Intensity Relation (LIR) and directional group encoding method to obtain high quality features. Further, the obtained feature weights provide high classification accuracy. Folio Leaf dataset is utilized to evaluate performance of proposed WKNNC model. The obtained classification accuracy is compared against several state-of-art-techniques and proposed WKNNC model outperforms all of them.

Keywords Folio leaf dataset · Medicinal plant leaf images · Weighted KNN model · Classification · Feature extraction

1 Introduction

Plants have a significant part in the growth of several fields across the world such as in agriculture, medicine, foodstuff preparation, cosmetic industry, environmental growth and ecological security etc. Furthermore, plants plays vital role in human growth as well and subsequently utilized in daily life activities [5]. Plants not just only provide foodstuff but also supports in several productive activities for each human beings like they maintain

✉ Savitha Patil
savitha019@gmail.com

M. Sasikala
sasi_mum@rediffmail.com

[1] CSE, Sharanbasava University, Kalaburagi, Karnataka 585105, India

equilibrium between oxygen $\left(O_2\right)$ and carbon-di-oxide $\left(CO_2\right)$. Another vital characteristic of plants is that the medicinal plant species are one of the major contributor towards medicine production due to their ability to synthesize several chemical compounds and presence of antioxidant capabilities makes them fight with cardiovascular diseases, cancer diseases and defend humans against insects, fungi, herbivores mammals etc. Moreover, there are thousands of medicinal plant species are available across the world which can provide a new dimension to medicine industry. However, accurate identification of medicinal plant and their type is the most vital and necessary.

Modern medicines, which are utilized in medical treatment, are majorly made up of synthetic drugs, which are costly and have few side effects as well. Furthermore, according to World Health Organization (WHO), synthetic drugs are limited in terms of chronical disease treatment [33]. On the other side, medicinal plants are one of the biggest contributor towards traditional medicine which has almost 25% share of global prescribed medicine and widely utilized in pharmaceutical industry [16]. There are numerous benefits of traditional medicine utilization which is made up of medicinal plants like consists no side effects, keeps people healthy due to their comparatively minor adverse reactions and substantially cheaper than synthetic drugs [19]. Furthermore, medicinal plants consists of several bioactive chemical compounds which have characteristics like remain antioxidants, anti-allergenic, antibacterial, anti-hepatotoxic and anti-inflammatory. However, most difficult and challenging task to identify medicinal plant species and their exact type and their capabilities due to presence of abundant amount of herbs globally. Moreover, due to their high similarities in color, texture and shape, it becomes highly critical and complex to identify accurately the type of medicinal plant. Besides, accurate classification between normal and medicinal herbs is a laborious task [7, 18].

Furthermore, the classification of medicinal plants can be done using traditional methods or with the help of professionals. Generally, countries like India, Malaysia and Thailand still utilizes traditional methods to identify medicinal plants with the help of professionals [3]. However, traditional methods always require expert knowledge and classification from traditional ways is an extremely challenging and laborious task. Besides, automated classification of herbs based on machine learning techniques can be an effective solution to classify various medicinal plants, their color and shapes. Among various classification areas for herbs, leaf classification is widely utilized. Although, leaf classification is immensely challenging process for researchers, especially based on machine learning techniques. However, presence of several leaf varieties and their unique characteristics makes leaf classification an interesting topic for researchers [22]. There are several features like shapes, color and texture are available in various plant leaves but their similarities between those features make this task challenging and critical.

Thus, in recent years, several research community experts and researchers have shown immense interest in extracting high quality features from leaf images and obtaining plant leaf classification accurately. Some of the literatures are discussed in this article as follows. In [20], a deep learning technique is introduced for automated plant leaf classification based on the shape and texture feature. Here, Support Vector Machine (SVM) classifier is employed for leaf image classification and Malaysian herb dataset is utilized. The classification accuracy achieved using SVM classifier is 93%. In [23], a machine learning technique is adopted for accurate identification of herbs. Here, typical features such as texture, shape, vein and their combinations are obtained with the machine learning classifiers. Here, the model is tested on different datasets. In [34], a Dual Channels Convolutional Neural Network technique is introduced for effective segmentation and classification of plat leaf images. A Non-Maximum Suppression Algorithm is also adopted for the enhancement of performance efficiency. Here, Support Vector Machine (SVM) classifier is utilized to perform classification process. In [2, 4, 6, 17, 28, 29, 36], a neural network architecture is employed based on radial functions for plant leaf identification. Here, distance between feature points and centroid are calculated to obtain high quality features. Here, SVM classifier is employed to measure classification accuracy.

1.1 Motivation and contribution

However, it is evident from the above literatures that there are very few techniques, which have employed weighted KNN classifier for identification of medicinal, plant species. Although other classifiers provide decent accuracy results but weighted KNN classifier provide comparatively superior performance in terms of plant leaf identification. Therefore, a Weighted KNN Classification (WKNNC) Model is adopted for the accurate identification of plant leaf images based on machine learning techniques. Here, the proposed WKNNC model consists of four significant stages to achieve high quality results. In first stage, dataset images are pre-processed with the help of filtering techniques to reduce interference. Moreover, in second stage, high quality segmentation is achieved using an Improvised Snake Model (ISM) by obtained Region of Interest (ROI) which is demonstrated in detail in our previous research article. In next stage, typical features like shape, color and texture are obtained through WKNNC model for all dataset images. Although it was quite challenging due to presence of different medicinal plant species with same color and shape images. However, the proposed WKNNC model has performed well and obtained significant feature weights. Then, finally, those obtained feature weights are further utilized for classification process to identify type of medicinal plant species precisely. The classification accuracy and other metrics results obtained through proposed WKNNC model are very impressive and compared with several state-of-art-machine learning techniques.

This paper is presented in the following manner. Section 2, describes about the literature survey presented regarding identification of medicinal plant species and identification problems and how those problems can be sorted with the help of the proposed WKNNC Model. Section 3, describes about the methodology utilized for the efficient implementation of WKNNC model to enhance performance accuracy. Section 4 discusses about the experimental results and their comparison with state-of-art- classification techniques and Section 5 concludes the paper.

2 Related work

Medicinal plants contain anthocyanins, tocopherols, carotenoids and phenolics like bioactive chemical compounds which have the characteristics of provide protection against several diseases. Therefore, medicinal plant is extensively useful for human beings. However, there are numerous variety of herbs are available across the world. So among those herbs, accurate identification of non-medicinal and medicinal plant species are difficult and challenging process. Moreover, it becomes even more challenging when traditional classification methods are utilized for the identification of medicinal plant species with the help of experts due to high similarities between plant features. Therefore, identification of plant species with an automated methods based on machine learning techniques is highly essential. Several researchers have provided their efforts in bringing various plant leaf identification techniques and some of the latest research work are presented below.

In [3, 11, 13, 30, 32, 37, 38], a detailed review on plant leaf image classification is presented. This article describes about the various characteristics of plant leaf images, their type, public datasets, different feature extraction methods, various classification methods and ways to improve classification accuracy. Here, methods such as deep learning model and sparse reconstruction methods are discussed in brief for the identification of leaf images. In [1, 12, 14, 21, 26, 27, 31], a comprehensive review article is presented on identification of plant species and classification of plant leaf images. Here, several machine learning techniques are studied to enhance classification performance. Moreover, efficient feature extraction and pre-processing techniques are utilized to enhance classification accuracy. In [8], a plant leaf classification technique is adopted with the help of Discrete Contour Evolution (DCE) algorithm. Here, NP-hard problem is optimized and contour points are obtained using similarity transformation metrics and leaf energy function is minimized with the help of convex-concave relaxation framework. In [1, 9, 39], a deep learning model is employed with the help of Leaf Vein Morphometric features for plant leaf identification. Furthermore, different models of CNN architecture are employed for feature extraction process. Similarly, several classifiers are used to measure and compare classification accuracy. In [10], a Unique Shape Descriptor Algorithm is utilized for the accurate detection of plant species. Here, model is tested on Abridged Image Database. Here, a descriptors is employed to eliminate pixel selection issue. The performance results are compared with several existing classification techniques. In [39], a plant species identification technique is adopted with the help of Morphological Features based on the adaptive boosting methods. Here, adaboost classifier is utilized with Multilayer Perceptron. Here, pre-processing techniques are employed for high quality feature extraction like centroid, minor axis length, solidity, major axis length and perimeter from leaf images. In [24], a deep learning model along with Artificial Neural Network (ANN) arc hitecture is presented for precise plant leaf image classification. Here, ANN is utilized to automate feature extraction process. Here, KNN classifier is utilized to measure performance. In [25], a Leaf Recognition method is presented based on Local Line Direction Pattern. Along with that elliptical half Gabor is employed to extract local patterns. SVM classifier is utilized to measure classification accuracy.

However, very few techniques have utilized weighted KNN classifier for plant leaf image identification based on machine learning techniques which is a quite are challenging and complex. And plant leaf detection have many challenges such as precise leaf detection, typical feature extraction like color, texture and boundary. Therefore, Weighted K-Nearest Neighbour Classification (WKNNC) model is presented for the precise detection of plant leaf images and high quality feature extraction such as color, texture and shape features.

3 Modelling for weighted K- nearest neighbor classification (WKNNC) model

This section discusses about the mathematical modelling of proposed Weighted K- Nearest Neighbor Classification (WKNNC) Model for the accurate identification of plant leaf images from variety of medicinal plant species. In proposed WKNNC model, plant leaf images are trained first and then exact Region of Interest (ROI) for all the medicinal leaf images are obtained. Those obtained ROI is further utilized for typical feature extraction process to get features like color, texture and shape. Finally, those feature weights are utilized for classification process through proposed weighted KNN model. Here, Fig. 1 demonstrates the flow diagram of proposed WKNNC model.

3.1 Feature extraction using proposed WKNNC model

Here, the obtained typical gives rich information about medicinal plant species such as their exact boundaries, texture, size, color and precise color information. Moreover, obtained feature are segmented into two categories where first category is morphological features and another category is discriminative features. Here, morphological features consists of color, shape, structure and texture related information. However, discriminative features gives information about the pitch, spectral centroid, magnitude and sign related information of filter responses and mainly concentrates on the central pixels of an image. A detailed mathematical modelling is presented in the following manner for feature extraction and classification process.

Consider that for a central region of image K, the Morphological Features (MF) are expressed as follows,

$$\mathbb{M}_{p,R} = \sum_{r=0}^{R-1} q\left(s_{p,r} - s_t\right).2^r \tag{1}$$

$$q(s) = \begin{cases} 1, s \geq 0 \\ 0, s < 0 \end{cases} \tag{2}$$

here, the central region of image K in gray scale domain is expressed as s_t. Moreover, the r^{th} adjacent pixels of a circle inside an image are also in gray scale domain and expressed as $s_{p,r}$ whose radius is denoted as p. The sum of adjacent pixels are expressed as R and Morphological Features (MF) are represented by $\mathbb{M}$. Moreover, rotational constant morphological features $\mathbb{M}$ are represented by following equation,

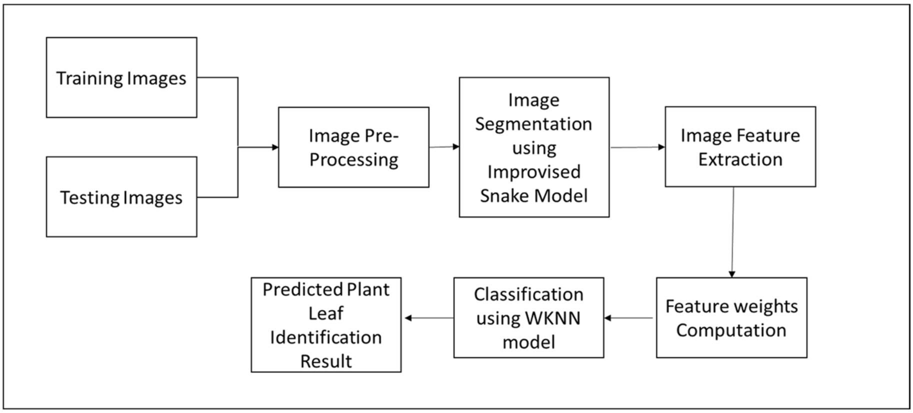

Fig. 1 Block diagram of proposed WKNNC model

$$\mathbb{M}_{p,R}^{G} = \begin{cases} \sum_{r=0}^{R-1} y\left(s_r - s_t\right), & D\left(\mathbb{M}_{p,R}\right) \le 2 \\ R+1 & , \quad otherwise \end{cases} \tag{3}$$

where, rotational invariant nature of $\mathbb{M}$ is expressed by G and D is denoted as uniformity evaluator for obtained morphological features $\mathbb{M}$ is $D \le 2$. Here, uniformity evaluator D is defined by following equation,

$$D\left(\mathbb{M}_{p,R}\right) = \left|q\left(s_{R-1} - s_t\right) - q\left(s_0 - s_t\right)\right| + \sum_{r=0}^{R-1} \left|q\left(s_r - s_t\right) - q\left(s_{R-1} - s_t\right)\right| \tag{4}$$

here, morphological features are evaluated for each dataset image by creating histogram gradient of every pixel in an image. Whereas discriminative features are evaluated by calculating magnitude and sign of multiple filter responses together to form multi-scale filter response which is demonstrated in Eq. (5),

$$q_{p,r} = q\left(s_{p,r} - s_t\right), f_r = \left|s_{p,r} - s_t\right| \tag{5}$$

Then, the multi-scale filter response is achieved by encoding magnitude and sign of each filter response, which is shown in following equation,

$$\mathbb{D}_{p,R} = \sum_{r=0}^{R=1} q\left(f_r - t\right).2^r \tag{6}$$

where, t is denoted as average of f_r for the given input image. Then, encoding of central region of given input image is represented as follows,

$$\mathbb{D}_{p,R} = q\left(s_t - t_K\right) \tag{7}$$

where, the average value for pixels of central region for a given input image K in gray scale domain is represented as t_K. Moreover, combining sign and magnitude of filter response and the pixels which focuses on central region, multi-scale filter response can be obtained. Then, the obtained response is transformed into a vector which can build image and provide fine discriminative features.

Then, local intensity relationship between neighboring pixels of the central region of each image are evaluated with the help of Local Intensity Relation (LIR) method while keeping morphological features as rotational constant. Further, a directional group encoding method is employed to isolate neighboring pixels into various groups which points towards a specific prevailing direction and group wise encoding is performed on local intensity relationships between neighboring pixels. However, precise encoding requires knowledge of prevailing direction which can be done by substituting average pixel value of a local region inside an image considering gray scale domain in place of each pixel value considering gray scale domain. Further, average pixel value of a local region inside an image considering gray scale domain near to central pixel region y inside a given image K is evaluated as follows,

$$\overline{s}_t = \Theta\left(S_{t,v}\right) \tag{8}$$

$$\overline{s}_{p,r} = \Theta\left(S_{p,r,v}\right) \tag{9}$$

where, dimensions of local pixel region are $v \times v$ near to central pixel region y and local pixel region is expressed as $S_{t,v}$ and similarly, local pixel region close to r^{th} adjacent pixel is expressed by $S_{p,r,v}$. Here, average pixel value of a local region inside an image considering gray scale domain is expressed by $\Theta(\bullet)$. The proposed WKNNC model eliminates interference in given images and gives typical information related to medicinal plant leaf images. Further, adjacent pixels of central region are varied in such a way that pixels points towards a specific prevailing direction to determine a rotational invariant structure. Then, the prevailing direction is determined as the gray scale difference of adjacent pixel index from central pixel region remains maximum which is demonstrated in Eq. (10),

$$W = \arg \max_{r \in \{0,1 \ldots\ldots, R-1\}} \left| \overline{s}_{p,r} - \overline{s}_t \right| \tag{10}$$

Further, discriminative features are obtained using maximum filter response. Therefore, once prevailing direction is determined, revolve the sequence till W comes to initial position,

$$\begin{array}{c} s'_{p,0}, \ldots \ldots \ldots s'_{p,R-1} \\ := \left(\overline{s}_{p,W}, \ldots .., \overline{s}_{p,R-1}, \overline{s}_{p,0}, \ldots .., \overline{s}_{p,W-1}\right) \end{array} \tag{11}$$

where, $:=$ is utilized for element wise operations. Then, varied groups are formed with evenly placed pixels. Low dimensionality of encoded features is achieved by placing 4 number of pixels in every group. Thus, for each sequence, number of groups are defined as $a = R/4$ which is demonstrated in following equation,

$$\overline{s}'_n = \begin{cases} \left(\overline{s}'_{p,0}, \overline{s}'_{p,a}, \overline{s}'_{p,2a}, \overline{s}'_{p,3a}\right), & n = 1 \\ \left(\overline{s}'_{p,1}, \overline{s}'_{p,a+1}, \overline{s}'_{p,2a+1}, \overline{s}'_{p,3a+1}\right), & n = 2 \\ . & \\ . & \\ \left(\overline{s}'_{p,a-1}, \overline{s}'_{p,2a-1}, \overline{s}'_{p,3a-1}, \overline{s}'_{p,R-1}\right), & n = a \end{cases} \tag{12}$$

where, a vector whose vector elements represents a^{th} group values of adjacent pixels in gray scale domain are expressed as $\overline{s}'_n \in n = 1, 2, \ldots ..a$. Finally, for each group encoding among adjacent pixels of intensity relations are determined by following equation,

$$H_{p,R,n} = \mathbb{f}\left(\lambda\left(\overline{s}'_n\right)\right) \tag{13}$$

where, function $\lambda(\bullet)$ is utilized for sorting operation which changes vector element order in non-descending manner. This function gives index vector of sorted vector elements. Then, function $\mathbb{f}(\bullet)$ is used for the transformation of attained index vector into a distinctive integer.

Furthermore, Local Intensity Relation (LIR) method has multiple benefits like LIR method is able to obtain multi-scale filter response with low dimensionality. Another importance is neighboring pixels are isolated into various groups which points towards a specific prevailing direction and this is achieved by using directional group encoding method and LIR methods makes the model robust enough to perform well even in presence of noise. Finally, LIR method is utilized for encoding of morphological and discriminative features. Here, discriminative features contain magnitude, sign and intensity related information of pixels in an image.

3.2 Leaf Classification using proposed WKNNC model

Moreover, morphological and discriminative features obtained using proposed EKNNC model from medicinal leaf input images are expressed as $\mathbb{M}_{p,R}$ and $\mathbb{D}_{p,R}$ respectively. Then, for images of each plant species, the obtained feature set $\mathcal{F} = \mathbb{M}_{p,R}, \mathbb{D}_{p,R}$ is expressed by below equation,

$$\mathcal{F} = I\Omega \tag{14}$$

where, Ω is expressed as coefficient valuation vector which is defined by $\Omega = \left\{\Omega_1, \Omega_2, \Omega_3, \dots \dots \dots ..\Omega_c\right\}^J$. Here, the coefficient valuation vector Ω is utilized for the evaluation of coefficients for all the input medicinal leaf images to generate weights from feature set $\mathcal{F}$. Then, coefficient valuation vector Ω becomes Ω_u for training sample z_u which is utilized to determine feature weights for obtained feature set $\mathcal{F}$. Here, Ω_u value becomes larger when similarities between $\mathcal{F}$ and z_u remains at their peak. Then, obtained feature set is classified using proposed WKNNC model and $L1$-norm regularization is employed to make process faster and efficient,

$$\overline{\Omega} = \underset{\Omega}{\operatorname{argmin}}\left[\|\mathcal{F} - I\Omega\|_2^2 + \xi\|\Omega\|_1\right] \tag{15}$$

where, $\overline{\Omega}$ is defined by coefficient set $\overline{\Omega} = \left\{\overline{\Omega}_1, \overline{\Omega}_2, \overline{\Omega}_3, \dots \dots ..\overline{\Omega}_c\right\}^J$. Here, coefficient valuation vector set is utilized for selecting $\mathcal{K}$- closest adjacent pixels for obtained feature set $\mathcal{F}$ considering every image of plant leaf dataset. Then, the coefficient with largest $\mathcal{K}$ value is selected from the coefficient set $\overline{\Omega} = \left\{\overline{\Omega}_1, \overline{\Omega}_2, \overline{\Omega}_3, \dots \dots ..\overline{\Omega}_c\right\}^J$ and expressed as $\widetilde{\Omega} = \left\{\widetilde{\Omega}_1, \widetilde{\Omega}_2, \widetilde{\Omega}_3, \dots \dots \dots ..\widetilde{\Omega}_{\mathcal{K}}\right\}$. Then, $\mathcal{K}$- closest adjacent pixels for obtained feature set $\mathcal{F}$ for respective training image is computed as,

$$J_{\mathcal{K}} = \left\{\left(z_u^{wknn}, \zeta_u^{wknn}\right)\right\}_{u=1}^{\mathcal{K}} \tag{16}$$

where, chosen assessment coefficient $\widetilde{\Omega}_u$ is always $\widetilde{\Omega}_u \geq 0$. Then, chosen coefficient $\widetilde{\Omega}_u$ of u^{th} closest adjacent pixels z_u^{wknn} is utilized as feature weights u^{th} closest adjacent pixels z_u^{wknn} considering obtained feature set $\mathcal{F}$. Finally, classification estimation is shown by below equation which is achieved using proposed WKNNC model,

$$\zeta_{\mathcal{F}} = \underset{\varphi}{\operatorname{argmax}} \sum_{z_u^{wknn}, \zeta_u^{wknn}} \widetilde{\Omega}_u \times \Psi\left(\varphi = \zeta_u^{wknn}\right) \tag{17}$$

where, required class label of feature set $\mathcal{F}$ is defined as $\zeta_{\mathcal{F}}$ for classification process. Then, a class with largest sum of feature weights is selected between all the present classes with the help of proposed WKNNC model and the chosen class belongs to $\mathcal{K} - th$ closest adjacent pixels. As discussed, coefficient valuation vectors are employed for choosing closest adjacent pixels with the help of proposed WKNNC model and class labels is utilized for

Table 1 Algorithm for medicinal plant leaf classification using proposed weighted KNN classifier

Input:

Training data $J = (z_u, \zeta_u)_{u=1}^{c}$ where $\zeta_u \in \widetilde{\Omega}$, the group of $\mathcal{F}$ class labels considering coefficient valuator $\widetilde{\Omega} = \{\widetilde{\Omega}_1, \widetilde{\Omega}_2, \widetilde{\Omega}_3, \ldots\ldots\ldots..\widetilde{\Omega}_{\mathcal{K}}\}$

The obtained feature set $\mathcal{F} = \mathbb{M}_{p,R}$, $\mathbb{D}_{p,R}$ and neighborhood distance $\mathcal{K}$

Output: Predict the class label with the help of proposed WKNNC model

Step 1: Compute coefficient valuator for all the given leaf images to analyze $\mathcal{F}$,

$$\overline{\Omega} = \arg\min_{\Omega}[\|\mathcal{F} - I\Omega\|_2^2 + \xi\|\Omega\|_1]$$

Step 2: Find out $\mathcal{K}$ –closest adjacent pixels with the help of coefficient valuator for feature set $\mathcal{F}$ of each given leaf image,

$$\overline{\Omega} = \{\overline{\Omega}_1, \overline{\Omega}_2, \overline{\Omega}_3, \ldots\ldots..\overline{\Omega}_c\}^J$$

Step 3: Search valuation coefficient which contains highest $\mathcal{K}$ value,

$$\widetilde{\Omega} = \{\widetilde{\Omega}_1, \widetilde{\Omega}_2, \widetilde{\Omega}_3, \ldots\ldots\ldots..\widetilde{\Omega}_{\mathcal{K}}\}$$

Step 4: Find out $\mathcal{K}$ closest adjacent pixel correspond to $\mathcal{K}$ largest valuation coefficient $\widetilde{\Omega}$,

$$J_{\mathcal{K}} = \{(z_u^{wknn}, \zeta_u^{wknn})\}_{u=1}^{\mathcal{K}}$$

Step 5: chosen coefficient $\widetilde{\Omega}_u$ of u^{th} closest adjacent pixels z_u^{wknn} is utilized as feature weights u^{th} closest adjacent pixels z_u^{wknn} considering obtained feature set $\mathcal{F}$.

$$\zeta_{\mathcal{F}} = \arg\max_{\varphi} \sum_{z_u^{wknn},\ \zeta_u^{wknn}} \widetilde{\Omega}_u \times \Psi\left(\varphi = \zeta_u^{wknn}\right)$$

classification process for feature set $\mathcal{F}$ for every medicinal plant leaf image. Table 1 demonstrates algorithm for classification process using proposed WKNNC model.

4 Result and discussion

This section discusses about the performance results of proposed Weighted K- Nearest Neighbor Classification (WKNNC) model for medicinal plant leaf images and shows comparison with state-of-art-classification techniques. Initially, input images are trained based on Improvised Snake model (ISM) to achieve high quality Region of Interest (ROI) images. Then, after segmentation of trained images, high quality morphological and discriminative features are obtained with the help of proposed WKNNC model. Moreover, to improve quality of obtained features methods like Local Intensity Relation (LIR) method and directional group encoding method are introduced. This methods utilized for obtaining multi-scale filer response and encode intensity relationships and improves quality of obtained features. Then, obtained feature weights are utilized for classification process with the help of proposed WKNNC model. Significant classification results are achieved and compared with several existing classification techniques in terms of classification accuracy.

4.1 Dataset details

This section discusses about the dataset utilized for identification of medical plant leaf species. Here, proposed WKNNC model is tested on Folio Leaf dataset. Moreover, this dataset is huge which consists of total 640 leaf images and total number of 32 different medicinal plant species. Every plant species in Folio Leaf dataset contains 20 different images. Farm located in 'University of Mauritius' is selected for all the medicinal leaf images present in this Folio dataset. All the images are taken from high resolution camera of Smartphone and each leaf image is having a pixel resolution of 1152×1726. Moreover, Folio leaf dataset is a public dataset which is available on website of UCI machine learning Repository. Here, proposed WKNNC model utilizes 70% images for training and remaining 30% image data for testing purpose. Furthermore, proposed WKNNC model effectively classify all the medicinal plant leaf images from obtained feature weights and estimate classes for respective leaf species accurately.

4.2 Performance metrics

This section gives information related to performance results of proposed weighted KNN classification (WKNNC) model for the accurate identification of medicinal plant leaf images. Table 2 demonstrates all the performance metric results obtained using proposed WKNNC model. The proposed WKNNC model works upon machine learning techniques. Numerous performance results are obtained and compared with various state-of-art-classification techniques considering metrics like accuracy, specificity, sensitivity, precision, recall, Fscore.

4.3 Comparative analysis

This section discusses about the comparison of proposed WKNNC model against various state-of-art-classification techniques in terms of classification accuracy, which is shown in Table 3. Here all the values of classification accuracy are mentioned in table.

Table 2 Observed performance metric of proposed WKNNC model

Performance Metrics	Value
Accuracy	0.9862
Sensitivity	0.7764
Specificity	0.9929
Precision	0.8381
Recall	0.7764
Fscore	0.7695

Table 3 Comparison of WKNNC model with various existing methodology

Model Name	Clas-sifcation accuracy
Multi-feature+SVM	0.9200
Multi-feature+PCA+SVM	0.9525
Alexnet+transfer learning	0.9531
GoogLe V3 +transfer learning	0.9540
k-nearest neighbor	0.8730
CNN	0.9460
PCNN	0.9667
AlexNet	0.9460
Vgg16	0.9488
DCCNN	0.9637
WKNNC	0.9862

The performance of proposed WKNNC model is analyzed against traditional techniques such as Multi-feature + SVM [30], Multi-feature + PCA+ SVM [37],Alexnet+transfer learning [38], GoogLe V3 +transfer learning [38], k-nearest neighbor [11], CNN [35], PCNN [31], AlexNet [34], Vgg16 [34], DCCNN [34]. In Table 3, first column describes name of the model, second column describes the classification accuracy. Moreover, average of classification accuracy is mentioned in Table 2 for comparison against various existing classification techniques, which is obtained from the classification results of each medicinal plant species from the dataset. The classification accuracy obtained with the help of proposed WKNNC model is 0.9862 which is quite high in contrast to other available classification techniques. The performance results shows the proposed WKNNC model outperforms all the mentioned classification techniques in terms of classification accuracy. Similarly, Table 4 represents classification accuracy results against different CNN models.

4.4 Graphical representation

This section demonstrates qualitative performance results which is shown in terms of graphs for Receiver Operating Characteristics (ROC) curve using proposed WKNNC

Table 4 Comparison of WKNNC model with varied CNN models

Model Name	Clas-sifcation accuracy
AlexNet	0.9308
Vgg16	0.9313
DCCNN	0.9494
AlexNet with ISC-MRCNN	0.9467
Vgg16 with ISC-MRCNN	0.9488
DCCNN with ISC-MRCNN	0.9637
WKNNC	0.9862

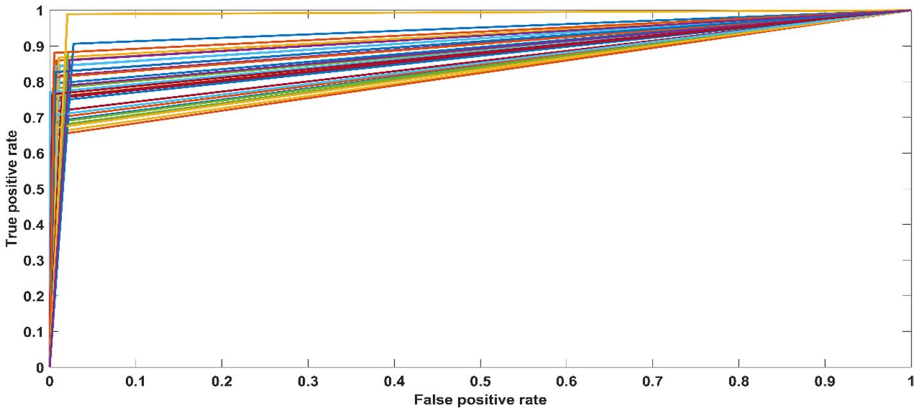

Fig. 2 Receiver operating characteristics using proposed WKNNC model

model. Along with that intermediate results like mask, boundary box, overlay and RoI results are also demonstrated considering varied medicinal plant species. Here, Fig. 2 shows ROC plot which defines obtained area under curve results all 32 plant species. Figure 3 demonstrates confusion matrix results considering all 33 medicinal plant species. Moreover, Fig. 4 shows qualitative performance of classification results considering plant species such as chrysanthemum, coeur demoiselle, jackfruit, mulberry leaf and thevetia.

Here, column 1 demonstrates type of leaf image, column 2 demonstrates original leaf image, column 3 demonstrates their masked image, column 4 demonstrates boundary box image, column 5 demonstrates overlay image and finally, column 6 demonstrates segmented ROI image for mentioned medicinal leaf image type. It is evident from the Fig. 4 results that obtained ROI outperforms several exiting techniques in terms of segmentation quality.

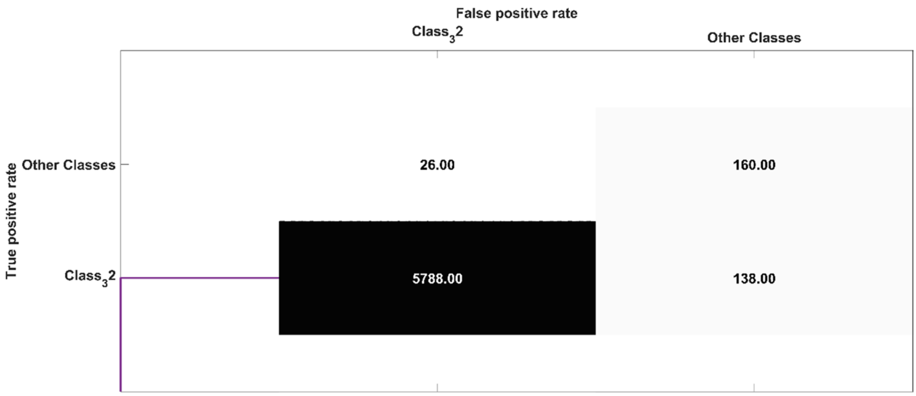

Fig. 3 Confusion matrix results using proposed WKNNC model

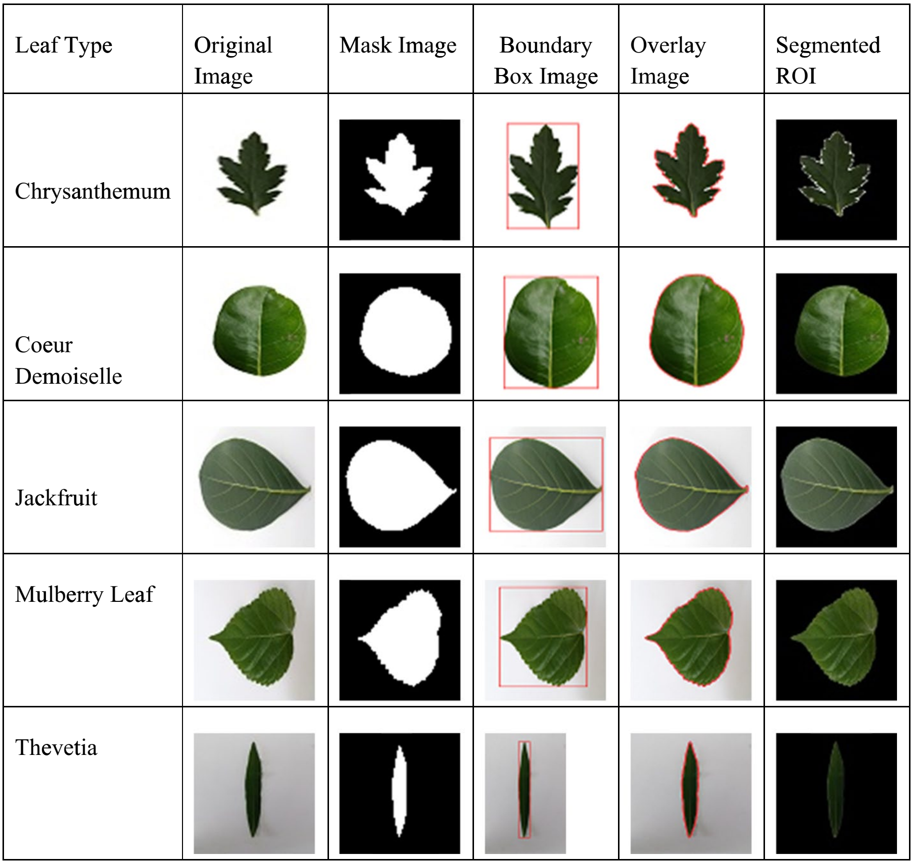

Fig. 4 Qualitative performance of classification results considering plant species

5 Conclusion

The identification of medicinal plants from thousands of available plant species is a vital and challenging issue due to their abundance utilization in production of traditional medicine, which can defend human beings from several diseases. Therefore, a weighted K- Nearest Neighbor Classification (WKNNC) Model for the precise detection of plant leaf images is presented to classify medicinal plant species based on machine learning techniques. A comprehensive mathematical modelling is presented to achieve high quality features and classification accuracy. The proposed WKNNC model utilizes Local Intensity Relation (LIR) method to obtain morphological and discriminative features which gives detailed information related to the features like color, shape, boundary and texture. Feature also provide information related to magnitude, sign and intensity relationships between pixels. Along with that a directional group encoding method is also presented for encoding of local intensity relationships and ultimately multi scale filer response is achieved to obtain discriminative features. Here, Folio medicinal leaf dataset is used for testing of proposed WKNNC model. The classification accuracy obtained with the help of proposed WKNNC model is 0.9862 which is quite high in contrast to other available classification techniques. Along with that, different performance metrics results and obtained ROC curve is demonstrated.

Data availability The datasets generated during and/or analyzed during the current study are available in the reference [38].

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Agrawal P, Girshick R, Malik J (2014) Analyzing the performance of multilayer neural networks for object recognition. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8695. Springer, Cham. https://doi.org/10.1007/978-3-319-10584-0_22
2. Ahmed A, Hussein SE (2020) Leaf identification using radial basis function neural networks and SSA based support vector machine. PLoS One 15(8):0237645. https://doi.org/10.1371/journal.pone.0237645
3. Anami BS, Nandyal SS, Govardhan A (2010) A combined color, texture and edge features based approach for identication and classification of Indian medicinal plants'. Int J Comput Appl 6(12):45–51
4. Azlah MA, Chua LS, Rahmad FR, Abdullah FI, Wan Alwi SR (2019) Review on techniques for plant leaf classification and recognition. Computers 8(4):77. https://doi.org/10.3390/computers8040077
5. Journal of Botany (2014) Retracted: phytoliths as emerging taxonomic tools for identification of plants: an overview. J Bot PB - Hindawi Publishing Corporation KW - ER 2014, p 231914. https://doi.org/10.1155/2014/231914
6. Chaudhury A, Barron JL (2020) Plant species identification from occluded leaf images. IEEE/ACM Trans Comput Biol Bioinform 17(3):1042–1055. https://doi.org/10.1109/TCBB.2018.2873611
7. Durairajah V, Gobee S, Muneer A (2018) Automatic vision based classification system using DNN and SVM classifiers. In 2018 3rd International Conference on Control, Robotics and Cybernetics (CRC), Penang, Malaysia (pp. 6–14). https://doi.org/10.1109/CRC.2018.00011
8. Farabet C, Couprie C, Najman L, LeCun Y (2013) Learning hierarchical features for scene labeling. IEEE PAMI 35(8):1915–1929
9. Guyer DE, Miles GE, Gaultney LD et al (1993) Application of machine vision to shape analysis in leaf and plant identification. Trans ASABE 36(1):163–171
10. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (ICCV), Santiago, Chile (pp. 1026–1034). https://doi.org/10.1109/ICCV.2015.123
11. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 770-778. https://doi.org/10.1109/cvpr.2016.90
12. Hinton G, Osindero S, Teh Y-W (2006) "A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
13. Huixian J (2020) The Analysis of Plants Image Recognition Based on Deep Learning and Artificial Neural Network. IEEE Access 8:68828–68841. https://doi.org/10.1109/ACCESS.2020.2986946
14. Im C, Nishida H, Kunii TL (1998) Recognizing plant species by leaf shapes-a case study of the acer family. In Proceedings fourteenth international conference on pattern recognition (Cat. No. 98EX170), Brisbane, QLD, Australia (Vol. 2, pp. 1171–1173). https://doi.org/10.1109/ICPR.1998.711904
15. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. MM 2014: Proceedings of the 22nd ACM international conference on Multimedia (pp. 675–678). https://doi.org/10.1145/2647868.2654889
16. Karami N, Javid A, Haghirosadat B (2017) A review of medicinal plants effective in the treatment or apoptosis of cancer cells. Cancer Press 3(1):22. https://doi.org/10.15562/tcp.41
17. Kumar M, Gupta S, Gao XZ, Singh A (2019) Plant species recognition using morphological features and adaptive boosting methodology. IEEE Access 7:163912–163918. https://doi.org/10.1109/ACCESS.2019.2952176
18. Larese MG, Namías R, Craviotto RM, Arango MR, Gallo C, Granitto PM (2014) Automatic classi_ cation of legumes using leaf vein image features. Pattern Recognit. 47(1):158–168

19. Lulekal E, Kelbessa E, Bekele T, Yineger H (2008) An ethnobotanical study of medicinal plants in Mana Angetu District, southeastern Ethiopia. J Ethnobiol Ethnomed.https://doi.org/10.1186/1746-4269-4-10
20. Muneer A, Fati SM (2020) Efficient and Automated Herbs Classification Approach Based on Shape and Texture Features using Deep Learning. IEEE Access 8:196747–196764. https://doi.org/10.1109/ACCESS.2020.3034033
21. Oide M, Ninomiya S (2000) Discrimination of soybean laeflet shape by neural networks with image input. Comput Electorn Agric 29(1–2):59–72
22. Patil AA, Bhagat KS (2016) Plant identication by leaf shape recognition:a review. Int J Eng Trends Technol 8(35):359–361
23. Pushpanathan K, Hanafi M, Mashohor S et al (2021) Machine learning in medicinal plants recognition: a review. Artif Intell Rev 54:305–327. https://doi.org/10.1007/s10462-020-09847-0
24. Qingfeng W, Kunhui L, Changle Z et al. (2007) Feature extraction and automatic recognition of plant leaf using artificial neural network. Adv Art Intell
25. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, Le Cun Y, (2014) Overfeat: Integrated recognition, localization and detection using convolutional networks, In ICLR
26. Simard PY, Steinkraus D, Platt JC (2003) Best practices for convolutional neural networks applied to visual document analysis. Seventh international conference on document analysis and recognition. Proceedings, Edinburgh, UK, 2003, pp. 958–963. https://doi.org/10.1109/ICDAR.2003.1227801
27. Soderkvist OJO (2001) Computer Vision Classification of Leaves from Swedish Trees. Linkoping University, Linkoping
28. Tan JW, Chang SW, Abdul-Kareem S, Yap HJ, Yong KT (2020) Deep learning for plant species classification using leaf vein morphometric. IEEE/ACM Trans Comput Biol Bioinform 17(1):82–90. https://doi.org/10.1109/TCBB.2018.2848653
29. Thanikkal JG, Dubey AK, Thomas MT (2020) Unique shape descriptor algorithm for medicinal plant identification (SDAMPI) with abridged image database. IEEE Sensors J 20(21):13103–13109. https://doi.org/10.1109/JSEN.2020.3002909
30. Wang L, Huai Y, Peng Y-C (2015) Identi_cation of foliage plant species based on multi-feature fusion of leaf images''. J Beijing Forestry Univ 37(01):55–61. https://doi.org/10.13332/j.cnki.jbfu.2015.01.006
31. Wang Z, Sun X, Zhang Y, Ying Z, Ma Y (2016) Leaf recognition based on PCNN. Neural Comput Appl 27(4):899–908. https://doi.org/10.1007/s00521-015-1904-1
32. Wang X, Du W, Guo F, Hu S (2020) Leaf recognition based on elliptical half Gabor and maximum gap local line direction pattern. IEEE Access 8:39175–39183. https://doi.org/10.1109/ACCESS.2020.2976117
33. WHO (1999) WHO monographs on selected medicinal plants, 1st edn. World Health Organization, Geneva
34. Yang X et al (2020) Instance Segmentation and Classification Method for Plant Leaf Images Based on ISC-MRCNN and APS-DCCNN. IEEE Access 8:151555–151573. https://doi.org/10.1109/ACCESS.2020.3017560
35. Zhang C, Zhou P, Li C, Liu L (2015) A convolutional neural network for leaves recognition using data augmentation. 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing, Liverpool, UK, (pp. 2143–2150). https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.318
36. Zhang Shanwen, Huang Wenzhun, Huang Yu-an, Zhang Chuanlei (2020) Plant species recognition methods using leaf image: Overview. Neurocomputing 408:246–272 ISSN 0925-2312
37. Zheng Y (2017) Method of leaf identication based on multi-feature dimension reduction. Trans Chin Soc Agric 48(03):30–37. https://doi.org/10.6041/j.issn.1000-1298.2017.03.004
38. Zheng Y, Zhang L (2018) Plant leaf image recognition method based on transfer learning with convolutional neural networks. Trans Chin Soc Agricult Mach 49(1):354–359. https://doi.org/10.6041/j.issn.1000-1298.2018.S0.047
39. Ni ZW (1997) Deep study on activation function in BP network. Journal of Anhui University Natural Science Edition 21