International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 3, June 2022, pp. 2509~2516
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i3.pp2509-2516

Corn leaf image classification based on machine learning techniques for accurate leaf disease detection

Daneshwari Ashok Noola[1], Dayananda Rangapura Basavaraju[2]
[1]Department of Information Science and Engineering, Bijapur Liberal District Education Association Vachana Pitamaha Dr. P. G. Halakatti College of Engineering and Technology, Vijayapura, India
[2]Department of Computer Science and Engineering, Kammavari Sangha Institute of Technology, Bengaluru, India

Article Info

Article history:

Received May 20, 2021
Revised Sep 22, 2021
Accepted Oct 20, 2021

Keywords:

Corn leaf classification
Enhanced k-nearest neighbour
Fine and coarse features
Leaf disease detection
Machine learning

ABSTRACT

Corn leaf disease possesses a huge impact on the food industry and corn crop yield as corn is one of the essential and basic nutrition of human life especially to vegetarians and vegans. Hence it is obvious that the quality of corn has to be ideal, however, to achieve that it has to be protected from the several diseases. Thus, there is a high demand for an automated method, which can detect the disease in early-stage and take necessary steps. However, early disease detection possesses a huge challenge, and it is highly critical. Thus, in this research work, we focus on designing and developing enhanced k-nearest neighbor (EKNN) model by adopting the basic k-nearest neighbour (KNN) model. EKNN helps in distinguishing the different class disease. Further fine and coarse features with high quality are generated to obtain the discriminative, boundary, pattern and structural related information and this information are used for classification procedure. Classification process provides the gradient-based features of high quality. Moreover, the proposed model is evaluated considering the Plant-Village dataset; also, a comparative analysis is carried out with different traditional classification model with different metrics.

Corresponding Author:

Daneshwari Ashok Noola
Department of Information Science and Engineering, Bijapur Liberal District Education Association
Vachana Pitamaha Dr. P. G. Halakatti College of Engineering and Technology
Vijayapura, India
Email: ise.daneshwari@bldeacet.ac.in

1. INTRODUCTION

Several smart technological developments and utilization of machine learning techniques in recent years have revolutionized various fields like electronic media, medical, engineering, defense, and agriculture [1]. Among these fields, agriculture is drastically developed in recent years and one of the most benefitted fields from these smart technological developments [2]. As agriculture is the major economic source of the country, the utilization of smart agriculture techniques can provide a major boost to the economy of a country. Besides, there are several crops which can heavily impact the economy of the country. These crops are majorly produced in the country and even exported outside the country as well. One of those crops is corn which plays a significant role in agriculture field [3]. Corn is an essential traditional crop [4] which is utilized as human food, animal feed and also used as a raw material in several industries. The quality of corn kernels majorly depends on the crop yield and corn production [5]. A fast and effective technique is required to assess the quality of corn kernels [6]. However, several leaf diseases occur in the corn crop, can heavily impact the quality of crop [7] as well as yield. Moreover, the production rate of the corn crop is also affected heavily due to plant leaf diseases [8]. Therefore, the precise identification of leaf diseases is very essential to maintain the production rateand crop yield.

In [7], a corn disease identification method is adopted using the classification of leaf diseases based on support vector machine (SVM) classifier. A detailed study on segmentation of leaf diseases is also presented. In [8], a CNN architecture is adopted for corn leaf disease identification which enhances automation and digitalization in agriculture. Here, convolutional neural network (CNN) architecture enhances the classification accuracy of leaf diseases. In [9], an improvised machine learning technique is presented for the detection of crop leaf diseases. Here, handcrafted features are adopted which gives information about histogram-oriented gradient (HOG), segmented fractal texture analysis (SFTA) and local ternary patterns (LTP). In [10], a feature enhancement and a robust AlexNet technique arepresented for maize leaf disease identification. Modified neural network architecture is designed based on robust AlexNet technique. However, potential plant leaf disease detection techniques are not mature enough to use in practical applications. Several challenges need to be addressed like precise identification of plant leaf diseases, identification of various factors which can affect crop production and quality of corn and efficient feature extraction for the identification of disease type [11].

In this study, an enhanced k-nearest neighbour (EKNN) classifier is utilized for the precise detection ofcrop leaf diseases. Here, the detection of the leaf disease process is segregated into four phases using the proposed EKNN model. Moreover, the first phase describes the pre-processing process which filters the availability of noise from leaf images. Further, phase 2 discusses the segmentation process of leaf lesions to identify lesion boundaries and pattern related information. Then, phase 3 discusses the proposed feature extraction process to extract structure-related information corn leaf images. Finally, leaf classification is conducted on obtained features for the detection of leaf diseases using proposed EKNN model. This article is presented in the following manner. Section 2 describes the mathematical modelling of proposed corn leaf classification modified KNN technique. Section 3 describes the experimental results and their comparison with traditional leaf classification techniques and section 4 concludes the article.

2. MODELLING OF PROPOSED ENHANCED KNN CLASSIFIER

The health of the maize plant itself can be represented by the leaves [12]. Plants suffering from disease usually have yellowish or brownish leaves, as well as patches or decaying areas. This data is something we'd like to extract from image data [13]. During feature extraction process, features are considered to be discriminative that have low dimensions [14], and robust against change in data. Red, green, and blue (RGB) characteristics are extensively employed in image processing and pattern recognition to extract color information. RGB is highly recommended for object detection in images with substantial color variations [15]. Think of the RGB color as all the different colors that can be created using three different colored lights for red, green, and blue. The RGB values ranged from 1 to 255. We are going to normalize them from 0 to 1 in this task. The Figure 1 indicates the different diseases on corn leaf.

Figure 1. Corn leaf disease types

Digital image processing mechanism has been adopted recently for leaf disease identification; also, the structure is an eminent feature for pattern recognition [16]. Nevertheless, structure related feature provides precise information regarding crucial feature extraction such as shape, object pattern and boundaries. Moreover, structure-related information is classified into two distinctive categories for absolute leaf disease detection; the first category gives information regarding fine feature and second category provides an information about coarse features [17]. The architecture of the proposed model is given in Figure 2.

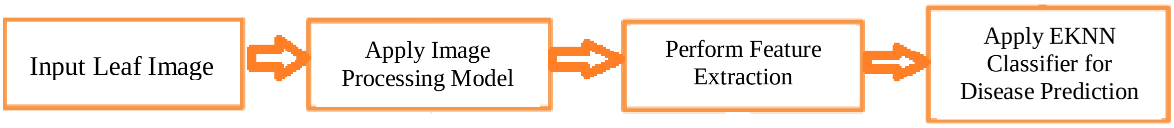

Figure 2. Proposed model architecture

The feature extraction process divides the image into a huge number of feature vectors, each of which is unaffected by image translation, scale, or rotation, but is affected by lighting changes and local geometric distortions [18]. These characteristics are similar to neurons in the primary visual cortex [19], which are responsible for primate vision's basic forms, colors, and motions for object detection. Based on the Euclidian distance of the feature vector, this technique looks for candidates from related features. The fine features are considered using (1).

Let us consider an area of pixels (central) in one image *M*, hence Fine-the feature can be given (1):

$$fine\_feat_{T,U} = \sum_{u=0}^{U-1} 2^u \, j(S_{t,u} - S_o) \tag{1}$$

In the (1) S_O and $S_{t,u}$ indicates the greyscale value of area of pixel (AoP) of central and u^{th} adjacent pixels on a circle that has radius as t. Moreover, the whole adjacent pixels are given as U that is performed using (2).

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

Further rotational constant fine feature is given in (3).

$$fine\_feat_{t,U}^{V} = \begin{cases} \sum_{u=0}^{U-1} j(S_u - S_o), & E(fine\_feat_{T,U}) \leq 2 \\ U + 1, & otherwise \end{cases} \tag{3}$$

In the (3), V represents a rotational constant form of $fine\_feat$ and obtained fine features are uniform with $E \leq 2$. Here, D is uniformity evaluator, which is defined by (4),

$$E(fine\_feat_{T,U}) = \sum_{t=0}^{T-1} |j(S_u - S_o) - j(S_{U-1} - S_o)| + |j(S_{U-1} - S_o) - j(S_0 - S_o)| \tag{4}$$

Furthermore, sign and magnitude component of given filter response are designed through a given equation; also multi-scale histogram of same is represented through (5).

$$\begin{aligned} C_u &= |S_{t,u} - S_o| \\ j_{t,u} &= j(S_{t,u} - S_o), \end{aligned} \tag{5}$$

Nevertheless, the components in (5) are encoded to achieve a histogram representation of multi-scale histogram representation and can be represented using (6).

$$coarse\_feat_{T,U} = \sum_{u=0}^{U-1} 2^u . j(C_u - O) \tag{6}$$

In the (6), in case of an input image b_t the average value is expressed through n; further central AoP is achieved are encoded with the (7), where n_J indicates central pixels value.

$$coarse\_feat_{T,U} = j(S_o - O_k) \tag{7}$$

Further, confined intensity directional order relation (DOR) is adopted for defining the intensity relationship among nearest pixels through gradient features; further confined intensity-DOR uses the confined ordinal data for determining the intensity relation among neighbouring pixel for each central AoP. Moreover, direction set (DS) based encoding strategy is used for segregating the neighbouring pixel to multisets pixel through pointing towards one direction and maintaining the rotational invariance.

Later, encoding is carried out through set wise intensity; furthermore, it is very important to identify the dominant direction through changing the greyscale value [20] for each pixel to a given average greyscale of provided shaped regions of the given image. Forgiven image N and central AoP y, average greyscale is given through (8).

$$\bar{S}_o = \phi(S_{o,w}) \tag{8}$$

$$\bar{S}_{t,u} = \phi(S_{t,u,w}) \tag{9}$$

In the (9), arbitrarily shaped of the size $w \times w$ around the AoP with notation $z$can be presented as $S_{O,W}$ where arbitrarily shaped regions with neighbouring pixel su^{th}. Further, $\varphi(\cdot)$ indicates the average greyscale of the arbitrary. Hence, this model improvises robustness in terms of noise and furtherit gives the extended structure for the leaf disease. Later neighbouring pixels are flipped through pointing to the specific dominant direction for developing the rotational constant. Moreover, the dominant direction is considered as neighbouring pixel index where the difference from AoP is on a higher value, and it is shown in the (10).

$$C = arg \max_{v \in \{0,1.....,V-1\}} \left| \bar{t}_{u,v} - \bar{t}_o \right| \tag{10}$$

In the (10) indicates class label of histogram and proposed model identifies the class among all the classthat have the biggest summation of adjacent pixels. Further, above equation employs the evaluation coefficient for identifying the neighbour pixels and further feature set through class labels. Moreover, high classification is obtained through the various class label for detecting the leaf disease in corn leaves.

3. RESULTS AND DISCUSSION

In this section of the research, we evaluate proposed methodology considering the various parameter; furtherproposed model is compared with various existing methodology. Moreover, different parameter like classification accuracy, precision and area under curve (AUC) are considered for the comparative analysis [21]. In general corn leaf identification follows the four distinctive steps, the first method is pre-processing which includes the noise reduction and faster processing [22]. Second step includes the segmentation process where disease area is identified through the prominent leaf lesion and accurate boundaries; this process has been carried out in the previous research work. In third step high quality, fine and coarse features are extracted toachieve the high quality.

Moreover, an adequate number of image sets are used for the training the proposed model for leaf disease evaluation; plant village dataset is used, and leaf disease are taken from the dataset [23]. Plant village dataset is parted into four distinctive diseases i.e., healthy classes, cercospora-leafspot gray, northern leaf blight and common rust [24]. Each class comprises several numbers of leaf disease that are depicted in Table 1, also each images have the pixel resolutions of 256×256, and entire project is simulated through MATLAB. Moreover, high-quality features are achieved through proposed model and classification process is carried out on the extracted features. Proposed models identify he disease presence in the leaves and determines the particular class of disease where it belongs.

Table 1. Dataset details

Class	Total images
HL	1162
LSG	513
NLB	985
RS	1160
Total images	3820
Class	Total images

3.1. Perfomance metrics

In this section, we evaluate the model considering the different deep learning metrics like accuracy, sensitivity, specificity, and AUC. Table 2 shows the different metrics; in here EKNN observes massive value of accuracy, sensitivity, specificity, and AUC of 99.86, 99.60, 99.88, and 99.75 respectively. The process of calculating the performance metrics is calculated as:

- The number of cases correctly diagnosed as disease is called a true positive (TP),
- The number of false positives (FP) is the number of cases that are wrongly diagnosed as disease,
- The number of instances correctly classified as healthy is called a true negative (TN),

– The number of instances wrongly diagnosed as healthy is called a false negative (FN).

Table 2. Observed performance metric of EKNN

Performance metrics	Value
Accuracy	99.86
Sensitivity	99.6
Specificity	99.88
AUC	99.75

3.1.1. Accuracy

A test's accuracy is determined by its capacity to appropriately distinguish between disease and healthy cases. We should determine the fraction of true positive and true negative in all analyzed cases to estimate a test's accuracy [25]. This can be expressed mathematically as:

$$Accuracy = TP + TNTP + TN + FP + FN$$

3.1.2. Sensitivity

A test's sensitivity refers to its capacity to appropriately identify disease cases. Calculate the fraction of true positives in leaf disease instances to estimate it. This can be expressed mathematically as:

$$Sensitivity = TPTP + FN$$

3.1.3. Specificity

A test's specificity refers to its capacity to appropriately identify healthy instances. Calculate the fraction of true negatives in healthy instances to estimate it. This can be expressed mathematically as:

$$Specificity = TNTN + FP$$

To compare two state of arttechnique is considered i.e., AlexNet and DMS-Robust AlexNet. In Table 3, first column presents the model's name, second column present the disease category, third, fourth and fifth column presents precision, recall, and F1. To extract the pixels and features from the image, the corn leaf image undergo segmentation for accurate pixel extraction. The process of image segmentation levels is indicated in Figure 3.

In Table 3, HL represents the healthy leaf, LSG represents the gray leaf spot, NLB represents North leaf blight and RS indicates common rust observed in the leaf. Moreover, for healthy leaf disease proposed model observes precision, recall and F1-score of 99.7, 99.94 and 99.82 respectively. In case of gray leaf spot in corn leaf, EKNN observes 99.87, 99.1 and 99.48 percentage of precision, recall and F1-score respectively. Further, for north leaf blight disease EKNN achieves 99.48 percentage of precision, recall and F1-core; similarly forcommon rust disease, EKNN model achieves precision of 99.77, recall of 99.88 and F1-score of 99.83.

Figure 4 shows ROC receiver operating characteristics of EKNN model for the four distinctive types ofleaf i.e., healthy leaf, gray leaf spot, North leaf blight and common rust. Further, ROC is plotted between true positive rate and true negative rate; also, different color is used to distinguish the other leaf disease. The disease detection accuracy of the proposed and existing models is represented in Figure 5. The results indicate that the proposed model accuracy is more than the traditional methods.

Table 3. Comparison of EKNN with various existing methodology

Model	Category	Precision	Recall	F1
AlexNet		HL	95.31	95.31
		LSG	94.39	94.28
		NLB	93.51	94.46
		RS	92.72	94.87
DMS-Robust AlexNet		HL	99.2	98.44
		LSG	99.16	98.21
		NLB	98.03	98
		RS	98.95	98.47
EKNN		HL	99.7	99.94
		LSG	99.87	99.1
		NLB	99.48	99.48
		RS	99.77	99.88

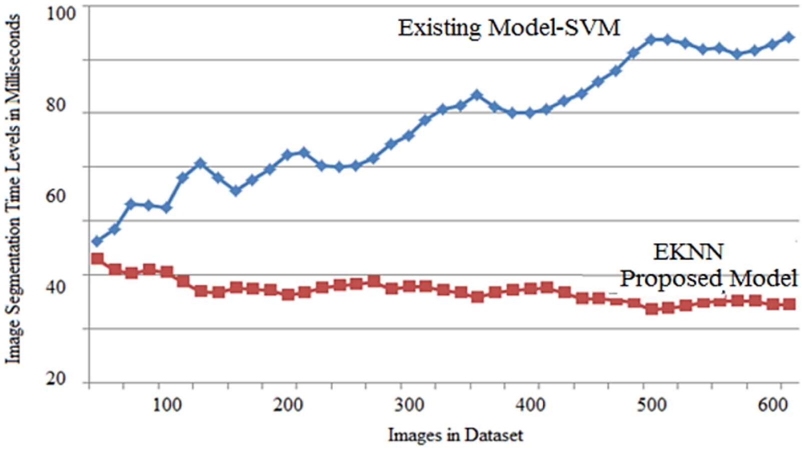

Figure 3. Image segmentation levels

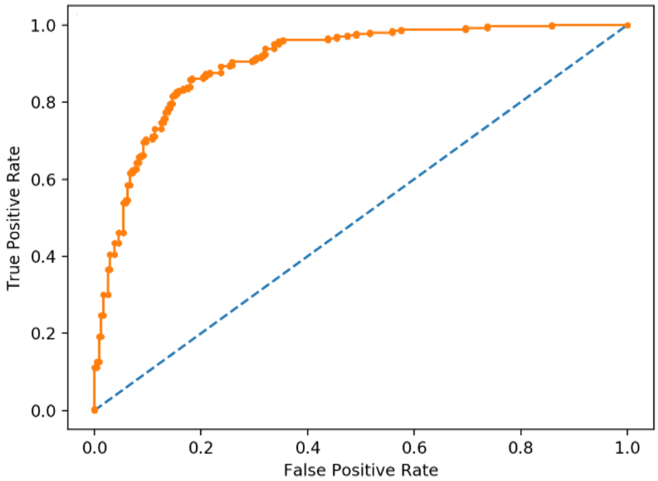

Figure 4. ROC curve using proposed EKNN model for multi-class classification

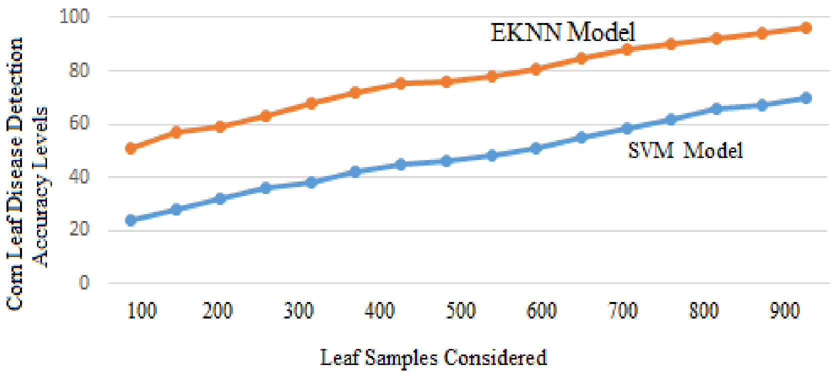

Figure 5. Disease detection accuracy levels

4. CONCLUSION

Early detection of corn leaf is considered eminent due to the major utilization of corn crop. Hence, enhanced-KNN is designed for absolute identification of corn disease and further distinguish the classes of disease. Moreover, EKNN follows the enhanced mathematical modelling for achieving the high-quality features. In here, fine, and coarse feature are achieved through the proposed model to improvise the classification accuracy. Further, confined intensity-DOR is adopted to achieve the low–dimensional through intensity relationship optimization among the neighbouring pixels. Furthermore, an optimized mechanism strategy named Directional set is developed for segregating the neighbouring pixels into the various sets through pointing into the specific direction. Proposed model EKNN is evaluated considering the various traditional mechanism and existing mechanism. Moreover, proposed model achieves observes massive value of accuracy, sensitivity, specificity, and AUC of 99.86, 99.60, 99.88, and 99.75 respectively. Further comparative analysis is carried out in terms of precision, recall and F1 score.

REFERENCES

[1] S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, "Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology," *IEEE Access*, vol. 6, pp. 8852–8863, 2018, doi: 10.1109/ACCESS.2018.2800685.

[2] A. Wu *et al.*, "Classification of corn kernels grades using image analysis and support vector machine," *Advances in Mechanical Engineering*, vol. 10, no. 12, Dec. 2018, doi: 10.1177/1687814018817642.

[3] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, vol. 266, pp. 79–88, 2017.

[4] D. Ward, P. Moghadam, and N. Hudson, "Deep leaf segmentation using synthetic data," In *Proceedings of the British Machine Vision Conference (BMVC) Workshop on Computer Vision Problems in Plant Pheonotyping (CVPPP)*, 2019.

[5] S. V. A. B. Cantero, D. N. Goncalves, L. F. dos Santos Scabini, and W. N. Goncalves, "Importance of vertices in complex networks applied to texture analysis," *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 777–786, Feb. 2020, doi: 10.1109/TCYB.2018.2873135.

[6] A. Chaudhury and J. L. Barron, "Plant species identification from occluded leaf images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 3, pp. 1042–1055, May 2020, doi: 10.1109/TCBB.2018.2873611.

[7] Z. Liu, Z. Du, Y. Peng, M. Tong, X. Liu, and W. Chen, "Study on corn disease identification based on PCA and SVM," in *Proceedings of 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2020*, Jun. 2020, pp. 661–664, doi: 10.1109/ITNEC48623.2020.9084689.

[8] K. P. Panigrahi, A. K. Sahoo, and H. Das, "A CNN approach for corn leaves disease detection to support digital agricultural system," in *Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020*, Jun. 2020, pp. 678–683, doi: 10.1109/ICOEI48184.2020.9142871.

[9] K. Aurangzeb, F. Akmal, M. Attique Khan, M. Sharif, and M. Y. Javed, "Advanced machine learning algorithm based system for crops leaf diseases recognition," in *Proceedings - 2020 6th Conference on Data Science and Machine Learning Applications, CDMA 2020*, Mar. 2020, pp. 146–151, doi: 10.1109/CDMA47397.2020.00031.

[10] M. Lv, G. Zhou, M. He, A. Chen, W. Zhang, and Y. Hu, "Maize leaf disease identification based on feature enhancement and DMS-Robust Alexnet," *IEEE Access*, vol. 8, pp. 57952–57966, 2020, doi: 10.1109/ACCESS.2020.2982443.

[11] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, Jun. 2020, doi: 10.1016/j.compag.2020.105393.

[12] X. Bai *et al.*, "Rice heading stage automatic observation by multi-classifier cascade based rice spike detection method," *Agricultural and Forest Meteorology*, vol. 259, pp. 260–270, Sep. 2018, doi: 10.1016/j.agrformet.2018.05.001.

[13] A. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, "Deep learning for image-based cassava disease detection," *Frontiers in Plant Science*, vol. 8, Oct. 2017, doi: 10.3389/fpls.2017.01852.

[14] M. Islam, Anh Dinh, K. Wahid, and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Apr. 2017, pp. 1–4, doi: 10.1109/CCECE.2017.7946594.

[15] J. W. Tan, S. W. Chang, S. Abdul-Kareem, H. J. Yap, and K. T. Yong, "Deep learning for plant species classification using leaf vein morphometric," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 1, pp. 82–90, 2020, doi: 10.1109/TCBB.2018.2848653.

[16] A. McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in *AAAI/ICML-98 Workshop on Learning for Text Categorization*, 1998, pp. 41–48, doi: 10.1.1.46.1529.

[17] A. E. Ozturk and M. Ceylan, "Fusion and ANN based classification of liver focal lesions using phases in magnetic resonance imaging," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2015, pp. 415–419, doi: 10.1109/ISBI.2015.7163900.

[18] J. D.Pujari, R. Yakkundimath, and A. S. Byadgi, "SVM and ANN based classification of plant diseases using feature reduction technique," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 3, no. 7, 2016, doi: 10.9781/ijimai.2016.371.

[19] S. H. Lee, H. Goëau, P. Bonnet, and A. Joly, "New perspectives on plant disease characterization based on deep learning," *Computers and Electronics in Agriculture*, vol. 170, Mar. 2020, doi: 10.1016/j.compag.2020.105220.

[20] A. Fuentes, S. Yoon, S. Kim, and D. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, Sep. 2017, doi: 10.3390/s17092022.

[21] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[22] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.

[23] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/3289801.

[24] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, Jun. 2019, doi: 10.1016/j.compag.2018.03.032.

[25] M. Hussain, J. J. Bird, and D. R. Faria, "A study on CNN transfer learning for image classification," in *Advances in Intelligent Systems and Computing*, vol. 840, pp. 191–202, 2019.

BIOGRAPHIES OF AUTHORS

Daneshwari Ashok Noola received the M.Tech. degree in computer science and engineering from VTU Belgaum, Karnataka, in 2011 and the B.Tech degree in computer science and engineering from VTU Belgaum, Karnataka, in 2004. Currently, she is working as Assistant Professor in the Department of Information Science and Engineering, V. P. Dr P. G. Halakatti College of Engineering and Technology, Vijayapura. Her research interests include image processing, machine learning and deep learning. She can be contacted at email: danu.noola@gmail.com.

Dayananda Rangapura Basavaraju holds a PhD in Faculty of Engineering, Computer Science, Pacific Academy of Higher education and Research University. (PAHER), Udaipur. He is currently working as professor in KSIT Bangalore and Head of VTU Research center at KSIT and even IQAC Head at the college. His areas of research include machine learning, cloud computing, artificial intelligence, image processing and data mining. He can be contacted at email: rbdayanandakit@gmail.com.