

Crop leaf disease identification based on ensemble classification

Navneet Kaur^a, V. Devendran^b and Sahil Verma^c

^a *Lovely Professional University, Jalandhar, Punjab, India*

^b *Lovely Professional University, Jalandhar, Punjab, India*

^c *Chandigarh University, Mohali, Punjab, India*

Abstract

Livestock and horticulture are well-known contributors to the global economy, particularly in countries where farming is the sole motivation for income. Yet, it is regretful that infection degeneration has affected this. Vegetables are a significant source of power for people and animals. Leaves and stems are the most common way for plants to interact with the surroundings. As a consequence, researchers and educators are responsible for investigating the problem and developing ways for recognizing disease-infected leaves. Growers everywhere across the world will be able to take immediate action to avoid their produce from getting heavily affected, so sparing the globe and themselves from a potential global recession. Because manually diagnosing ailments might not have been the ideal solution, a mechanical methodology for recognizing leaf ailments could benefit the agricultural sector while also enhancing crop output. The goal of this research is to evaluate classification outcomes by combining composite classification with hybrid Law's mask, LBP, and GLCM. The proposed method illustrates that a group of classifiers can surpass individual classifiers. The attributes employed are also vital in attaining the best findings because ensemble classification has demonstrated to be much more reliable. The experiments used sick leaf pictures of bell pepper, potato, and tomato from the PlantVillage database.

Keywords

Leaf disease, ensemble classification, feature extraction

1. Introduction

The most fundamental and among the most significant duties in agribusiness is the appropriate identification of infection of crop leaves with diseases. It's amazing that plant diseases are still detected manually in today's technological world, and it's possible that doing so for crops in abundance or in the outdoors would be problematic. As an outcome, the tool for preventing illness became crucial, pushing investigators to create a structure that is more successful than the manual technique in diagnosing illnesses. For this aim, many databases in the form of photos are available. The disease's initial point might be conceived of as the infectious spots on the leaflets [2]. As a reason, having a thorough understanding of the disorder is essential. Crop diseases detection with the visible light is a time-consuming and error-prone operation [3]. As a response, the importance of a computerised system has to be stressed. Among the most main advancements for developing systems proficient of replicating humans is machine learning [33]. This is performed by employing a variety of strategies. Construction of an automatic system capable of classifying leaf diseases using image processing method can increase yield. Leaf photos can be taken using a camera phone or any other suitable photo-capturing instrument. This is accomplished so that a usable dataset can be compiled and disease hotspots can be identified. Several image processing techniques should be included

Algorithms, Computing and Mathematics Conference, August 19 – 20, 2021, Chennai, India.

EMAIL: sahilverma@ieee.org (Sahil Verma)

ORCID: 0000-0003-3136-4029 (Sahil Verma)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

because they can be used to locate problematic zones and collect useful features to analyse the condition. To find the appropriate ill area, a procedure called as photo segmentation is used. Thereafter, the attributes are extracted in order to estimate the disease using various categorization techniques. State-of-the-art approaches, as well as their execution on a large dataset, were studied to express these problems to the researchers. The primary priority of our study is on how preventative care of diminishing plant leaf health can be used to control production. The mission is to create the most efficient system feasible. The purpose of this research is to focus on ensemble classification [30,32] and the use of feature vectors in light of the possible benefits of diverse machine learning techniques.

The following is a breakdown of the paper's structure. The image processing methods are discussed in Section 2. Section 3 delves deeply into the linked research and literature. The recommended methodology is presented in Section 4. The suggested work's practical results are presented in Section 5. The findings and possible implications of the current study are discussed in Section 6.

2. Overview of Image Processing

Amongst the most globally acknowledged ways for analyzing and identifying plant leaf disorders is computer vision. A number of experiments were available to undertake advanced research in the field of plant disease detection.

2.1. Acquisition

It's a crucial phase in the image processing operation. In this procedure, photos from the world wide web or high-definition sensors are used to capture high-quality photographs. The PlantVillage dataset, which is a landmark dataset provided by Penn State University, was referenced in the bulk of the publications. The purpose of this programme is to harness AI advancements and current practices to provide rural communities with solutions. Multiple retail digital cameras were used to take high-quality images of the ill plant leaves [4].

2.2. Pre-processing

To boost the image's quality, image enhancement strategies such as image filtering and image contrast improvement are applied. It can sometimes be important to use this to remove unwanted parts from an image.

2.3. Segmentation

The image is divided into pieces with comparable characteristics. To focus solely on the ill region of the image, subdivision is essential. The retrieved attributes will be efficient in discriminating across infected and non-infected areas if the photograph has been adequately partitioned. Edge-based, threshold-based, and colour scheme segmentation have all worked brilliantly in detecting leaf disease. The Sobel operator and canny edge detection [1] are two edge-based segmentation techniques which have been used. For this goal, a range of techniques have been used in several study articles. Only a few of the common computer vision segmentation algorithms are K-means clustering [6], Fuzzy c-means clustering [5], and the Otsu method [8] [9]. Growth in a seeded area has also been shown to be beneficial [7].

2.4. Feature Extraction

It's the most important stage of image processing after segmentation. The Gray Level Co-occurrence Matrix (GLCM) is a typical feature extraction method for diagnosing leaf illness that assesses numerous texture parameters such as entropy, energy, contrast, homogeneity, correlation, and etc [11]. Many investigators have integrated textural, pigment, and form data to predict leaf disorders [11]. Speeded-up robust features (SURF), histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT), dense SIFT (DSIFT), and pyramid histograms of visual words (PHOW) have all been used to identify soybean diseases [10].

2.5. Classification

The classification procedure is the final step. Categorization is among the most important components of image processing. It's a way of identifying images of plant leaves as disorders that have been discovered. The researchers put a range of categorisation methods to the test in a variety of circumstances. This type of classification scheme must be able to distinguish among contaminated and non-infected leaf pictures [10]. Machine learning approaches are divided into two categories: supervised and unsupervised [12]. The inputs and also the corresponding label readings must be included in the training dataset for supervised algorithms. In contrast, the unsupervised technique, which does not require label values, will develop classification assumptions on its own.

3. Related Work

Several research have been undertaken on the taxonomy of leaf diseases. These were carried done using a range of datasets, including readily available ones. Extensive study has also been done on real - world datasets. [37] presented a survey of various plant leaf disease detection schemes. Using 12 feature vectors such as mean, standard deviation, skewness, kurtosis, shape features such as Hu moment variants, and texture features using LBP and GLCM, the XGBoost classifier had an accuracy of 86.58 percent and the SVM classifier had an accuracy of 81.67 percent for three rice diseases [13]. For extracting features, the Histogram of Oriented Gradient (HOG) was used in [14], with Random Forest achieving a maximum accuracy of 70.14 percent. [15] devised a method for distinguishing between diseased and healthy leaves based on K-means clustering and feature extraction methodologies such as GLCM, Haralick, Gabor, and 2DWT. The IPM dataset and Plant Village were used in this investigation. For varying reasons, optimization algorithms such as feature selection and optimal segmentation have also been used. [16] found that selection of features utilising the newly designed Spider Monkey optimization improved computational effectiveness and categorization efficiency when compared to traditional methods. Because the extraneous parts merely harm performance, spider monkey optimization is used to choose only the most relevant elements. To increase segmentation and classification, as well as the accuracy of the outputs, [17] uses Particle Swarm Optimization. Using the optimised extracted features, [18] proposed an effective technique for boosting classification accuracy. [19] employed a delta segmentation method, colour histograms, LBP textural properties, and trained models to differentiate the disease-affected area. [20] also included a whole new image segmentation technique. The developed method's accuracy rate was shown to be much higher than that of existing methods. Using an SVM classifier, [21] developed a method for segmentation and features extraction. It also uses Gaussian filters, long transforms, and 2D DWT with a dataset of 500 photographs. [22] proposed a feature set consisting of a two-feature set separated into ten characteristics. It used the K means clustering approach to partition the diseased area. [6] used the K means clustering technique for fragmenting the lesion from the image using the theory of super-pixel segmentation and derivation of Pyramid of Histogram of Oriented Gradients(PHOG) features on two data sets of apple and cucumber. Using One Class classifiers trained on vine leaves, [24] suggested an approach for recognising four ill diseases. [23] looked at numerous machine learning techniques to detect sicknesses on rice leaves, including logistic regression, Nave Bayes, decision trees, and KNN. Apart from image processing, machine learning and deep learning has been the most trending topic these days for other domains as well and has been utilized for various purposes. WSN algorithm [34] has been proposed using machine learning. the authors have proposed a model [35] to reconstruct medical images. To predict traffic flow [36], deep learning has been utilized.

4. Proposed Methodology

The technique for the proposed work is depicted in Figure 1.

4.1. Dataset Collection

The dataset utilised for training and validation is PlantVillage [15] [18], which comprises sick bell pepper, potato, and tomato leaves. Plant Village is essentially a Penn State University research and development branch.

4.2. Segmentation using K means clustering

K means segmentation [6] is an unsupervised method for fragmenting similar regions in digital images. It separates the image into K clusters, each with a set of centroids of its own. Unsupervised is clearly used for data that has not been tagged or labelled. The purpose of this technique is to reduce the total distance between all locations and the cluster centre.

4.3. Feature Extraction

The feature extraction procedure is used to show the distinctive features in an image. Feature extraction methods used included Law's Texture Mask, GLCM, and LBP. The Laws texture feature [31] is a strategy for identifying the image's supplementary characteristics that has been used in research such as the classification of wood faults [27], mammography classification [26], and bone texture analysis [25]. The texture energy is calculated using a set of 5*5 convolutional filters. It employs filter masks within a predetermined window size. It was chosen due of its superior ability to extract texture information from images. The four essential aspects that can be analysed are the image's level, edge, spot, and ripple. GLCM [15] [28] is one of the oldest methods for analysing textures. It's a grid that's created over a photograph to show how co-occurring pixels are distributed. LBP is also a statistically based feature. LBP defines the pattern with the tiniest primitives. LBP was designed to deal with two-dimensional texture information. LBP [29] is a visual description that was developed in 1994. For basic LBP, a 3*3 pixel proximity is acceptable. First, the photograph must be converted to monochrome. 8 pixel vicinity will be assessed around a center pixel. Using this centre pixel as a threshold, a set of 8 binary digits will be created.

4.4. Ensemble Classification

When opposed to pure or solo classifications, ensemble learning [30,32] techniques have significantly beaten them. In the proposed approach, models such as RF, ANN, SVM, KNN, Logistic regression, and Nave Bayes have been used. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were employed to reduce the dimensionality of the data. The proposed methodology is depicted in Fig 1.

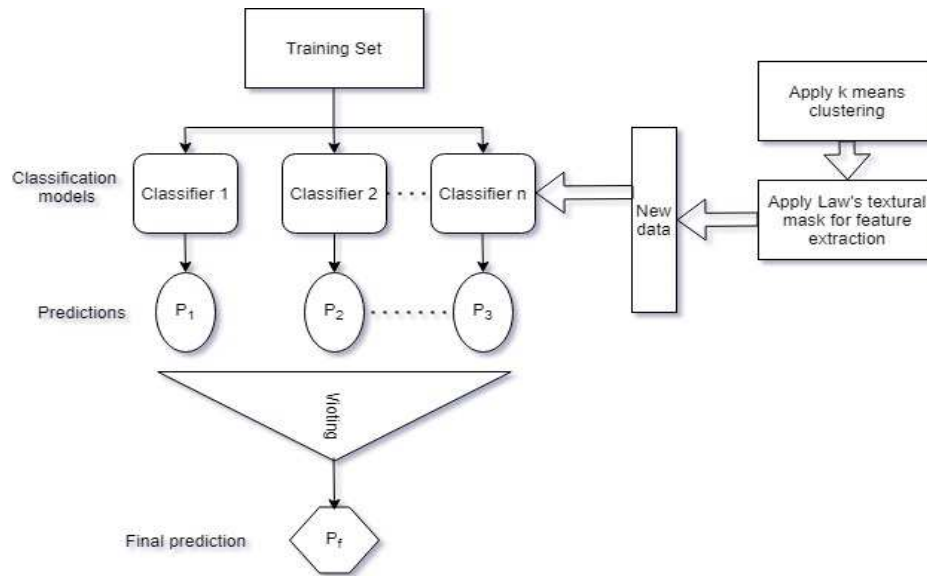


Figure 1: Proposed Methodology

5. Experimental Results

5.1. Leaf Images Dataset

There are a total of 20,639 photos in the set, divided into two categories for bell peppers, three categories for potatoes, and ten categories for tomatoes. 70 percent of the photos were utilised for training, while the remaining 30% were used for testing. In order to evaluate the results, the methodologies are combined.

5.2 Evaluation Metrics

We employed a variety of evaluation indicators to assess the classification model's performance:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN},$$

$$\text{Precision} = \frac{TP}{TP+FP},$$

$$\text{Recall} = \frac{TP}{TP+FN},$$

where TP = True positive, TN = True Negative, FP = False Positive, FN = False Negative

5.3. Abbreviations used in results

Table 1: Abbreviations

Approach used	Abbreviation used in results
Law's mask + GLCM + LBP + PCA + RF	Pca-Rf3
Law's mask + GLCM + LBP + PCA + (ANN, SVM, Logistic Regression, KNN, Naïve Bayes)	pca-ensemble-3
Proposed features (3*3 Law's mask) + LDA + RF	Lda-rf-3

Proposed features (3*3 Law's mask) + LDA + (ANN, SVM, Logistic Regression, KNN, Naïve Bayes)

Lda-ensemble-3

Proposed features (3*3 Law's mask) + RF

Rf-3

Proposed features (3*3 Law's mask) - (ANN, SVM, Logistic Regression, KNN, Naïve Bayes)

Ensemble-3

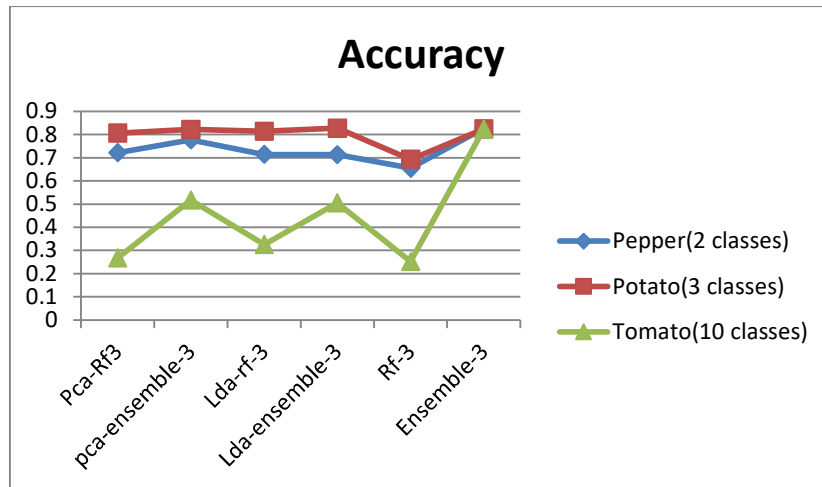


Figure 2: Comparison chart for accuracy

Ensemble 3 has the highest accuracy of 82.66 for pepper, as shown in Fig 2. For potato, Lda-ensemble-3 achieves the maximum accuracy of 82.81. For tomato, Ensemble 3 has the highest accuracy of 82.50.

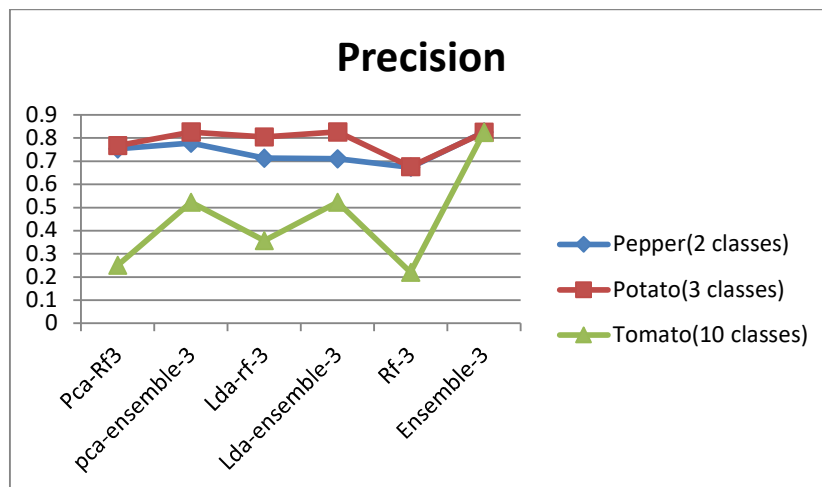


Figure 3: Comparison chart for precision

Ensemble 3 has the highest precision of 82.53 for pepper, as seen in Fig 2. For potato, Lda-ensemble-3 achieves the maximum precision of 82.62. For tomato, Ensemble 3 obtains the maximum precision of 82.41.

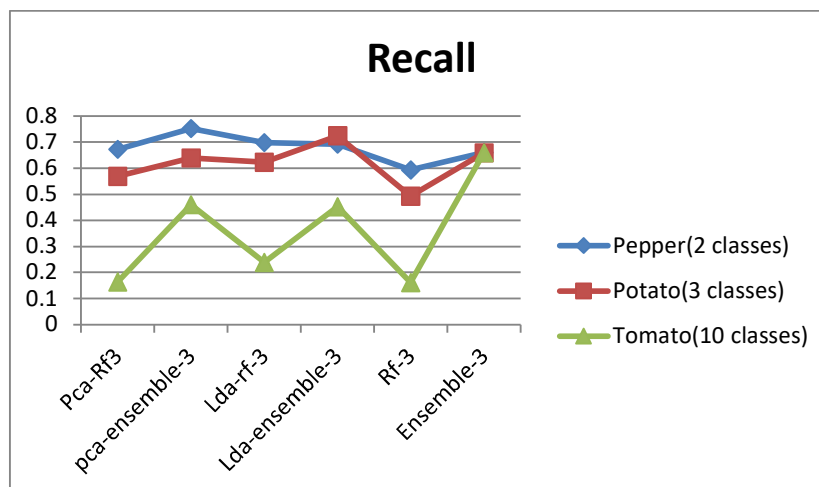


Figure 4: Comparison chart for recall

According to Figure 4, pca-ensemble-3 has the best recall of 75.27 for pepper. For potato, Lda-ensemble-3 achieves the maximum recall of 72.41. For tomato, Ensemble 3 had the highest recall of 65.84.

6. Conclusion

The paper's biggest contribution is the effective construction of an ensemble - based strategy that incorporates various feature extraction strategies. All of the trials were carried out using the PlantVillage dataset, which included two disease categories from bell peppers, three from potatoes, and ten from tomatoes. Image capture, segmentation, feature extraction, and categorization are all involved, but the feature extraction and classification phases receive the most attention. The use of feature extraction algorithms like GLCM and LBP has been considered. Classifiers such as RF, SVM, ANN, KNN, logistic regression, and Nave Bayes have been employed to create an efficient classifier model. The ensemble classification using several characteristics has been applied, and the results have been evaluated. When combined with the proposed work, our ensemble classifier produced the best results in terms of accuracy, precision, and recall. Ensemble 3 has the maximum accuracy of 82.66 for pepper, as previously indicated. For potato, Lda-ensemble-3 achieves the maximum accuracy of 82.81. For tomato, Ensemble 3 has the highest accuracy of 82.50.

7. References

1. R. C. Shinde , J. Mathew C and C. Y. Patil, Segmentation technique for soybean leaves disease detection, *International Journal of Advanced Research*. 3 (5) (2015) 522-528.
2. J. G. A. Barbedo, A review on the main challenges in automatic plant disease identification based on visible range images, *Biosystems Engineering*. 144 (2016) 52-60.
3. R. Kaur, M. Kaur, A brief review on plant disease detection using in image processing, *International Journal of Computer Science and Mobile Computing*. 6 (2) (2017) 101–106.
4. J. G. A. Barbedo, A novel algorithm for semi-automatic segmentation of plant leaf disease symptoms using digital image processing, *Tropical Plant Pathology*. 41 (2016) 210-224.
5. X. Bai, X. Lia, Z. Fu, X. Lv and L. Zhang, A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images, *Computers and Electronics in Agriculture*. 136 (2017) 157-165.
6. S. Zhang, H. Wang, W. Huang and Z. You, Plant diseased leaf segmentation and recognition by fusion of superpixel, K-means and PHOG, *Optik*. 157 (2018) 866-872.

7. J. Pang, Z. Bai, J. Lai and S. Li, Automatic segmentation of crop leaf spot disease images by integrating local threshold and seeded region growing, *International Conference on Image Analysis and Signal Processing*. (2011) 590-594.
8. L. Wang, F. Dong, Q. Guo, C. Nie and S. Sun, Improved rotation kernel transformation directional feature for recognition of wheat stripe rust and powdery mildew, *7th International Congress on Image and Signal Processing*. (2014).
9. R. Masood, S. A. Khan and M. Khan, Plants disease segmentation using image processing, *International Journal of Modern Education and Computer Science*. 8 (1) (2016) 24-32.
10. R. D. L. Pires, D. N. Gonçalves, J. P. M. Oruê, W. E. S. Kanashiro, J. F. Rodrigues Jr., B. B. Machado and W. N. Gonçalves, Local descriptors for soybean disease recognition, *Computers and Electronics in Agriculture*. 125 (2016) 48-55.
11. K. Huang, Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features, *Computers and Electronics in Agriculture*. 57 (1) (2007) 3-11.
12. https://en.wikipedia.org/wiki/Machine_learning
13. M. A. Azim, M. K. Islam, Md. M. Rahman and F. Jahan, An effective feature extraction method for rice leaf disease classification, *Telecommunication, Computing, Electronics and Control*. 19 (2) (2021) 463-470.
14. S. Ramesh, R. Hebbar; Niveditha M., Pooja R., Prasad Bhat N., Shashank N. and Vinod P.V., Plant Disease Detection Using Machine Learning, *International Conference on Design Innovations for 3Cs Compute Communicate Control*. (2018).
15. S. Kaur, S. Pandey and S. Goel, Semi-automatic leaf disease detection and classification system for soybean culture, *IET*. 12 (6) (2018) 1038-1048.
16. S. Kumar, B. Sharma, V. K. Sharma, H. Sharma and J. C. Bansa, Plant leaf disease identification using exponential spider monkey optimization, *Sustainable Computing: Informatics and Systems*. 28 (2020).
17. V. P. Kour and S. Arora, Particle Swarm Optimization Based Support Vector Machine (P-SVM) for the Segmentation and Classification of Plants, *IEEE Access*. 7 (2019) 29374 – 29385.
18. M. A. Khan, M. I. U. Lali, M. Sharif, K. Javed, K. Aurangzeb, S. I. Haider, A. S. Altamrah and T. Akram, An Optimized Method for Segmentation and Classification of Apple Diseases Based on Strong Correlation and Genetic Algorithm Based Feature Selection, *IEEE Access*. 7 (2019) 2169-3536.
19. H. Ali, M. I. Lali, M. Z. Nawaz, M. Sharif and B. A. Saleem, Symptom based automated detection of citrus diseases using color histogram and textural descriptors, *Computers and Electronics in Agriculture*. 138 (2017) 92-104.
20. V. Singh and A. K. Misra, Detection of Plant Leaf Diseases Using Image Segmentation and Soft Computing Techniques, *Information Processing in Agriculture*. 4 (1) (2017) 41-49.
21. K. Singh, S. Kumar and P. Kaur, Support vector machine classifier based detection of fungal rust disease in Pea Plant (*Pisum sativum*), *International Journal of Information Technology*. 11 (2019) 485-492.
22. Md. T. Habib, A. Majumder, A. Z. M. Jakaria, M. Aktera, M. S. Uddin and F. Ahmed, Machine vision based papaya disease recognition, *Journal of King Saud University - Computer and Information Sciences*. 32 (3) (2020) 300-309.
23. K. Ahmed, T. R. Shahidi, S. M. I. Alam and S. Momen, Rice Leaf Disease Detection Using Machine Learning Techniques, *International Conference on Sustainable Technologies for Industry*. (2019).
24. X. E. Pantazi, D. Moshou and A. A. Tamouridou, Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers, *Computers and Electronics in Agriculture*. 156 (2019) 96-104.
25. M. Rachidi, A. Marchadier, C. Gadois, E. Lespessailles, C. Chappard and C. L. Benhamou, Laws' masks descriptors applied to bone texture analysis: an innovative and discriminant tool in osteoporosis, *Skeletal Radiology*. 37 (2008) 541-548.
26. A.S. Setiawan, Elysia, J. Wesley and Y. Purnama, Mammogram Classification using Law's Texture Energy Measure and Neural Networks, *Procedia Computer Science*. 59 (2015) 92-97.

27. K. Kamal, R. Qayyum, S. Mathavan and T. Zafara, Wood defects classification using laws texture energy measures and supervised learning approach, *Advanced Engineering Informatics*. 34 (2017) 125-135.
28. M. Sharif, M. A. Khan, Z. Iqbal, M. F. Azam, M. I. U. Lali and M. Y. Javed, Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection, *Computers and Electronics in Agriculture*. 150 (2018) 220-234.
29. T. Ojala, M. Pietikainen and D. Harwood, Performance evaluation of texture measures with classification based on Kullback discrimination of distributions, *Proceedings of 12th International Conference on Pattern Recognition*. (1994).
30. https://en.wikipedia.org/wiki/Ensemble_learning
31. Navneet Kaur, V. Devendran, Plant leaf disease detection using ensemble classification and feature extraction, *Turkish Journal of Computer and Mathematics Education*. 12(11) (2021) 2339-2352.
32. Navneet Kaur, V. Devendran, Novel plant leaf disease detection based on optimize segmentation and law mask feature extraction with SVM classifier, *Materials Today: Proceedings*.
33. Li, W., Chai, Y., Khan, F. *et al.* A Comprehensive Survey on Machine Learning-Based Big Data Analytics for IoT-Enabled Smart Healthcare System. *Mobile Network and Applications* 26, 234–252 (2021). <https://doi.org/10.1007/s11036-020-01700-6>.
34. Sowjanya Ramisetty, Kavita and Sahil Verma, “The Amalgamative Sharp WSN Routing and with Enhanced Machine Learning Journal of computational and theoretical nanoscience (JCTN), ASPBS publisher. Vol. 16 No. 9, 2019, pp. 3766–3769 , DOI: 10.1166/jctn.2019.8247 (Scopus)
35. S. More et al., "Security Assured CNN-Based Model for Reconstruction of Medical Images on the Internet of Healthcare Things," in *IEEE Access*, vol. 8, pp. 126333-126346, 2020, doi: 10.1109/ACCESS.2020.3006346.
36. Vijayalakshmi, B, Ramar, K, Jhanjhi, N, et al. An attention-based deep learning model for traffic flow prediction using spatiotemporal features towards sustainable smart city. *International Journal of Communication System*, Wiley 2021; 34, 4609. <https://doi.org/10.1002/dac.4609>.
37. Navneet Kaur, Sahil Verma, “Detection of Plant Leaf Diseases by Applying Image Processing Schemes” *Journal of computational and theoretical nanoscience (JCTN)*, ASPBS publisher. Vol. 16 No. 9, 2019, pp. 3728–3734, DOI: 10.1166/jctn.2019.8241