Journal of Mines, Metals and Fuels, *71(12C): 519-525; 2023. DOI: 10.18311/jmmf/2023/47275*

Print ISSN : 0022-2755

Journal of Mines, Metals and Fuels

Contents available at: *www.informaticsjournals.com/index.php/jmmf*

Studies on Detecting Plant Leaf Disease Using KNN Classifier and Severity Measurement

M. Sinthuja , S. Pravinth Raja[2], V. S. Akshaya[3] and M. Madhushree[2]

[1]Department of ISE, MS Ramaiah Institute of Technology, Bengaluru – 560054, Karnataka, India; sinthujamuthu@gmail.com

[2]Department of CSE, Presidency University, Bengaluru – 560064, Karnataka, India; pravinthraja@gmail.com

[3]Department of CSE, Sri Eshwar College of Engineering, Coimbatore – 641202, Tamil Nadu, India; akshaya.vs@sece.ac.in, shreemadhu440@gmail.com

Abstract

Food is one of the most basic needs of living beings. It is very essential to grow a sufficient quantity of crops to cope with the growing population. With time, it is seen that plants are being affected by various kinds of diseases. Different plant diseases are a major influence on agricultural output limitations, and they are generally difficult to control. As a result, early detection of plant diseases is crucial for avoiding output losses and improving the quality of agricultural goods. This research paper proposes a method to detect and classify the disease on plant leaves with a severity measurement for the grading system. The paper describes a machine learning-based detection and classification of diseases like early blight and late blight on "potato leaf" using the KNN classifier. Our research focuses on a rigorous examination of potato plant disease segmentation approaches.

Keywords: *Colour segmentation, Digital Photographs, Feature Extraction, Fuzzy Logic, Processing*

1.0 Introduction

Agriculture is the practice of cultivating soil, growing crops, and raising livestock. It feeds living creatures with nourishment and substance. It's also referred to as "farming." Agriculture, sometimes known as farming, is the most common form of employment in India. Agriculture is the practice of cultivating soil, growing crops, and raising livestock. It feeds living creatures with nourishment and substance. It's also referred to as "farming." Agriculture, sometimes known as farming, is the most common form of employment in India. "Potato" is the most widely produced crop in practically all of India's states. With the growing agriculture, various diseases are causing degradation in the quality of the soil as well as the plants. Excessive use of fertilizers, insecticides, and pesticides by farmers in the hope of making their crops grow healthy and quickly may result in the spread of illnesses that are damaging to the crops.

The two major diseases found on potato plant leaves are Late Blight (caused by the fungi *Phytophthora infestans*) and Early Blight (caused by *Alternaria solani*). Manual detection of disease on plant leaves requires time and may not be accurate. Machine learning techniques can be used to automate disease detection.

Image processing is the process of performing some operations on an image to get useful information from the image. Digital Image Processing is a technique that uses a computer system to process digital images through an algorithm.

The farmer captures the image of the leaf. After that, image preprocessing is done. After that, the disease can be detected by following a few steps, which help the farmer to take further action.

**Author for correspondence*

2.0 Literature Survey

For the past several years, several studies on automatic image processing and machine learning techniques have been conducted to detect plant disease. This section discusses some existing works.

Dhaware and Wanjale proposed a method for the classification of sunburn, grasshopper, and yellow mosaic leaf diseases using a Support Vector Machine Classifier (SVM). They have used a cluster-based technique for disease segmentation. The number of training data, testing data, and data cluster numbers of diseased parts was not defined[1].

Prakash *et al.* proposed a method to distinguish normal and abnormal leaves of citrus plants using an SVM Classifier. For disease portion segmentation, K-means clustering was applied. A classification accuracy of 0.09 is obtained by training and testing 60 images of citrus leaves[2].

Rajiv K. *et al.,* proposed a model that uses k-means segmentation for the separation of the diseased part and does feature extraction using GLCM. KNN was used as a classification algorithm. The achieved accuracy is 75.12% [3].

Padol and Yadav proposed a model to classify downy mildew and powdery mildew diseases found on grape leaves using a linear SVM classifier. Before segmentation, Gaussian filtering was applied before the K-means clustering algorithm was used for diseased part segmentation. 54 features were extracted from GLCM. The limitation of this model is that it considers too many features[4,8].

Chaudhary *et al.* transformed RGB images to different colour spaces such as YcbCr, HIS, and CIELAB during disease segmentation to compare performances. Dicot and monocot leaves were used. Mathur *et al.,* proposed a general method for plant disease detection which comprises four fundamental stages. Image preprocessing, segmentation of diseased parts using region-based segmentation, feature extraction using GLCM, and classification using KNN classifier are the final steps. The segmentation and classification processes were discussed in detail[6-10].

3.0 Proposed Methodology

In the proposed method, the images are captured by the farmers. Images can be taken with any kind of camera device. After obtaining the image, the images are sized using image preprocessing. In this preprocessing step, the transformation of the source image (the image which was captured by the farmer at first) into a new image is done. Somehow, the new image is similar to the source image. The difference that can be seen between the two images is that the new image has improved contrast, which means it results in changing the brightness of the individual image pixel[11], where the image is segmented into different pixels. We apply the GLCM algorithm for feature extraction. This step provides a clear picture of the disease in the plant leaf. In this step, we get to know about the disease. For disease classification, the KNN classifier is used.

This step will tell us if the leaf is affected or not. If it is affected, then what type of disease it is. In later steps, we find the severity of the disease detected and grading is done based on its value.

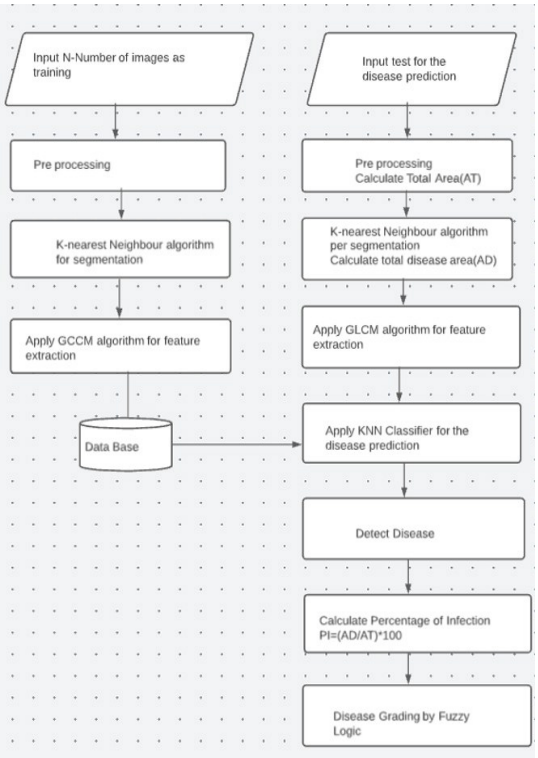


Figure 1. Block diagram of the proposed method.

3.1 Matlab Platform

3.1.1 Dataset Description

The pictures of the sick potato leaves were from the "Plant Village" website. This work uses a total of 200 photos. Of them, 150 photos are utilised to train the classifier, including 75 images of early- and late-stage blight-affected leaves. Figure 1. displays examples of each disease kind that can affect potato plants. 50 more random photos are utilised for testing as shown in Figure 2. The creation of





Figure 2. Sample images of potato plant disease : (a) Early Blight (b) Late Blight.

1) Mean, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

2) Standard Deviation, $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$

3) Energy = $\sum_{i=1}^{n}\sum_{j=1}^{n}(p(i,j))^2$

4) Contrast = $\sum_{i=1}^{n}\sum_{j=1}^{n}(i,j)^2 p(i,j)$

5) Homogeneity = $\sum_{i,j=1}^{n}\frac{p_{ij}}{1+(i-j)^2}$

6) Correlation = $\sum_{i,j=1}^{n} p_{ij}\frac{(i-\mu)(j-\mu)}{\sigma^2}$

Ground Truth Pictures requires the use of image editing software.

3.1.2 Image Acquisition

Image acquisition is the process of converting a physical image into a digital representation and storing it. The most important phase in every image processing job is image gathering. When the user selects the select image option, the gallery folder opens so that the user can select an image from the folder. On selection of the image from the folder, it is displayed on the GUI. Now, the user can click on the Upload Image option. The image, which is in RGB format, gets uploaded and stored for further analysis. Then the captured image is resized to 200*200 pixel resolution to reduce the computational burden and to efficiently utilise storage capacity.

3.1.3 Color Space Conversion

An image can be represented in different colour space formats. The acquired Rgb images of leaves are converted into Lab colour space as it provides more intuitive information about the image. The luminance layer is represented by "L*" in the L*a*b* model, and the chromaticity layer is represented by "a*b*." The "a*" and "b*" chromaticity layers, offer colour information along

the red-green and blue-yellow axes, respectively. This effort seeks to determine the average colour of each of three small sample sections from a given leaf picture in the 'a*b*' colour space, and then utilize Euclidian distance to assign each pixel to a different region. The main goal is to use the L*a*b* colour space to discern colours in a particular leaf image.

3.1.4 Color Segmentation

Image segmentation is the technique of dividing an image into various segments to make the representation of the picture more comprehensible and analytically applicable. At this stage, a picture is divided into three categories: green, infected, and background using the KNN[12].

Step 1: Obtain an Image

Step 2: Convert the input image to the L*a*b colour space.

Step 3: Use k-nearest neighbour to group the colours in the "a*b*" colour space. KNN is a well-known and straightforward supervised learning method that relies on instances and skips the learning phase. In reality, it's utilised to measure how similar or similar a pixel is to some of its closest neighbours, and it uses a distance matrix to compute the similarity. In this work, the Distance measure matrices are employed. From a given leaf picture, three sample areas are originally chosen, and after being transformed into LAB colour space, comparable pixels are allocated to the different regions based on Euclidian distance. The segmentation in this case uses a knn classifier with three neighbours.

Step 4: Among the three segmented parts, the diseased part is selected and sent for classification.

3.1.5 Feature Extraction

Following picture segmentation, certain characteristics are removed with the use of a matrix. The picture is used to create a matrix called a (GLCM), which takes the spatial connection between pixels into account. The sick leaf picture is used to extract a total of 6 characteristics that are then used to train the classifier. Energy, mean, homogeneity, standard deviation, correlation and contrast[13].

3.1.6 Classification

Classification is the process of labelling or categorising a new sample picture using the knowledge the classifier model learned during training. The KNN classifier is used to divide potato plant diseases into two categories in this work[14].

Nearest neighbour is a simple and effective classifier that may be utilised to resolve several categorization problems. It does not call on any prior expertise in machine learning methods, such as SVM. In other words, no additional training is needed if a new training pattern is included in an existing training set. When categorizing a new element vector, a comparison with the training dataset is made using Euclidean distance matrices. The element being classified is then given the class that appears the most frequently among its k-nearest neighbours after taking into account those neighbours. Choosing a parameter like "k," which stands for the quantity of neighbours to be considered, determines how well the model works and how accurate it is.

The classification procedure is divided into two sections. The first is the testing phase when the leaf pictures are left unlabeled and our algorithm creates a list of the k-nearest data points (training data) to label and categorise the unlabeled points. Is the training phase, when the leaf images are tagged with their classes.

3.1.7 Disease Severity Estimation and Grading Using FuzzyLogic

The binary image created from the actual scaled picture is changed so that the pixels matching to the leaf image are on. The total leaf area (T) is calculated from this picture. The total illness area is calculated using only the picture of the diseased component of the output image from colour image segmentation (D). Once T and D have

Table 1. Manual grading of diseases

Percent Infection	Disease Grade
1-20	0
20-40	1
40-60	2
60-80	3
80-100	4

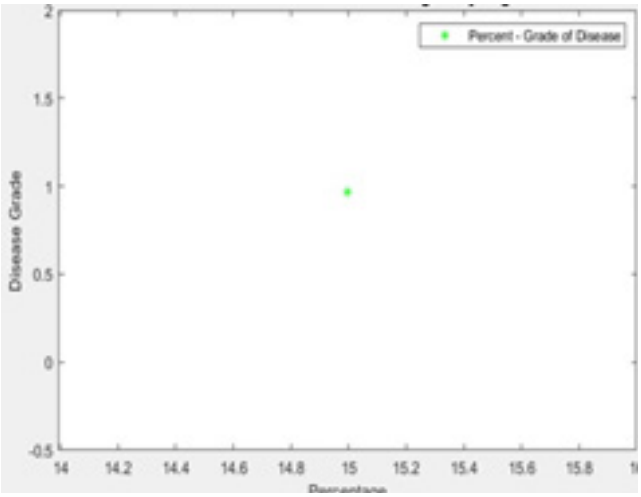


Figure 3. Plot of grading using fuzzy logic.

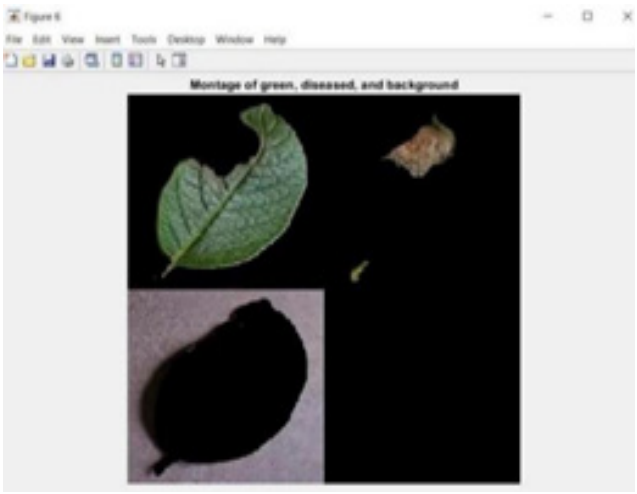


Figure 4. Segmented images.

been determined, a formula is used to get the infection percentage (P)[15].

P = (D/T) *100

A grade is assigned based on the percentage of infection. The problem that arises The grade stays 0 regardless of whether the infection rate is 5 or 15. Fuzzy logic, which was first introduced by Lotfi Zadeh (1965), is utilised in this study article to handle uncertainty, ambiguity and vagueness in a system. It is inculcated into agriculture to acquire the precise grade of the illnesses. A method for converting qualitative data into quantitative data is fuzzy logic (Figure 3).

A Fuzzy Inference System (FIS) is made to grade diseases using the disease score scale in Table 1. The input to the FIS is the percentage of infection, and the output is the grade assigned by it. The variables are specified using triangular membership functions, and the sickness is graded using six fuzzy criteria. The composition of a fuzzy rule-based model is straightforward, and it consists of four basic components:

A Fuzzification module that uses linguistic variables to convert inputs that are traditional measurements into fuzzy measurement[16].

- A set of conditioned fuzzy propositions that make up an if-then fuzzy rule basis.
- An inference method that, in essence, uses a fuzzy reasoning mechanism to obtain the outputs, *i.e.*, it computes using fuzzy rules.
- De-fuzzification.

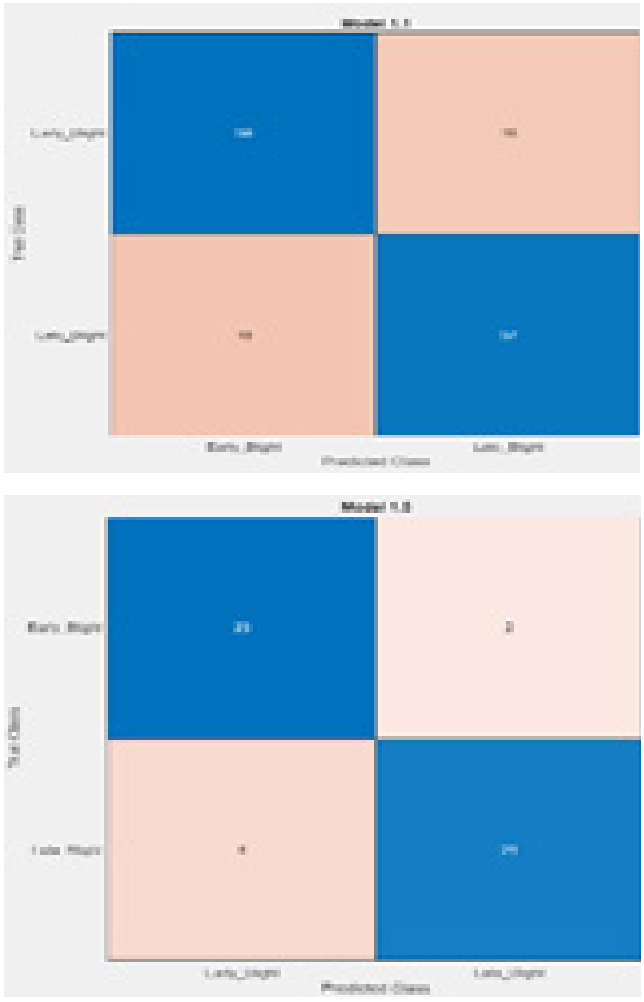
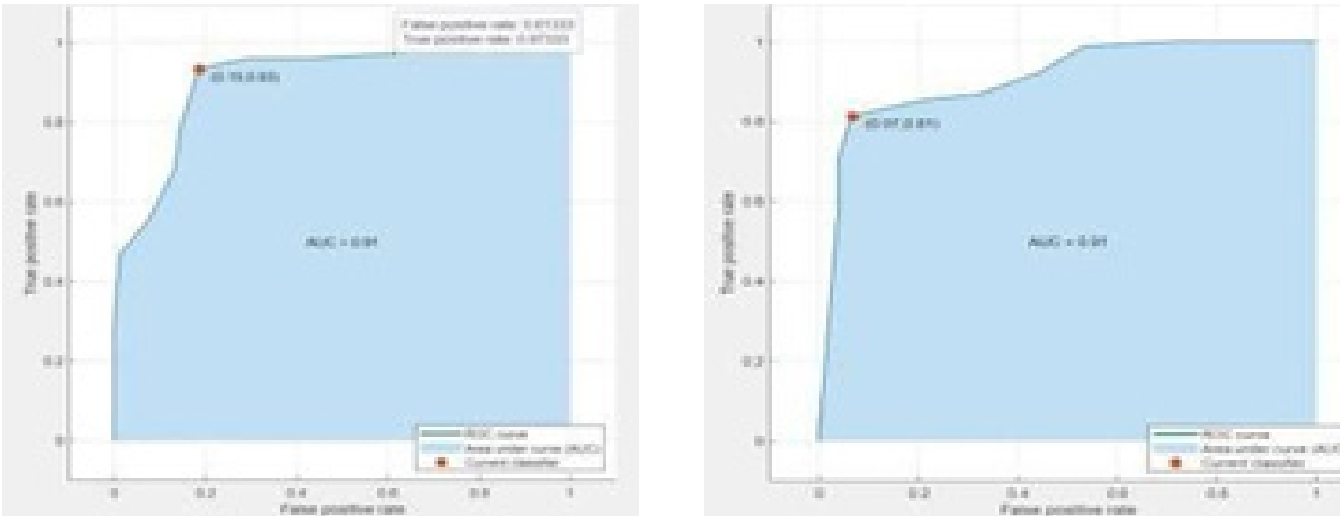


Figure 5. The trained KNN classifier's confusion matrix and the testing dataset.



(a) (b)

Figure 6. The trained classifier model's ROC curves. **(a)** Positive class: early blight **(b)** Positive class: late blight.

4.0 Results and Discussion

The output of leaf disease segmentation is shown in Figure 4, where each cluster highlights green, diseased, and background parts of the leaf image.

The output of the KNN classifier was compared to the manual segmentation result of the leaf image that was acquired using the image segmentation application in MATLAB. The confusion matrix and its parameters are used to calculate the performance difference between KNN and manual segmentation.

The dataset used in this proposed method consists of 200 potato plant leaf images, 150 images are considered for training and 50 images are considered for testing.

Both training and testing dataset has an equal mixture of early blight and late blight disease leaf images. Also in the feature extraction phase,6 features are considered. As a classification parameter Euclidian distance and 3 nearest neighbours are considered[17].

The effectiveness of the trained KNN classifier on the training dataset was estimated using the confusion matrix. The trained classifier model's confusion matrix is shown in Figure 5, with the bright orange region of the confusion matrix showing the wrong class and the blue-coloured area representing the illness class that was properly identified.

The confusion matrix and ROC curve both display numerical indications of the correct and wrong categorization.

The accuracy, false positive rate, and the rate of true positives were determined for the performance assessment and were found to be 88 to 92 per cent. The confusion matrix and ROC Curve were used to measure the performance of the classifier model on the testing dataset[18].

The confusion matrix and ROC Curve were used to measure the performance of the classifier model on the testing dataset as shown in Figure 6.

5.0 Conclusion

Techniques for automatically detecting plant diseases offer numerous benefits over manual ones. It requires less work, less time, and produces more precise results. The segmentation using the KNN classifier gives good results and improves the accuracy of disease detection. The performance of classification by KNN is found to give 87 to 90% accuracy, which is better compared with some of the already existing models.

Automatically grading the disease on plant leaves helps a lot. In this research paper, a method to grade diseases using a Fuzzy logic system is proposed, which

will help pathologists overcome the disadvantages of manual grading systems like time complexity.

In the future, the classification of more than two illnesses in different types of plants may be possible. Convolution neural networks can be used to improve classification accuracy and time complexity.

6.0 References

1. Dhaware CG, Wanjale KH. A modern approach for plant leaf disease classification which depends on leaf image processing. 2017 International Conference on Computer Communication and Informatics (ICCCI); 2017. p. 1-4. https://doi.org/10.1109/ICCCI.2017.8117733
2. Prakash RM, Saraswathy GP, Ramalakshmi G, Mangaleswari KH, Kaviya T. Detection of leaf diseases and classification using digital image processing. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS); 2017. p. 1-4. https://doi.org/10.1109/ICIIECS.2017.8275915
3. Rajiv K, Rajasekhar N, Lakshmi KP, Rao DS, Reddy PS. Accuracy evaluation of plant leaf disease detection and classification using GLCM and multiclass SVM classifier. Congress on Intelligent Systems, Singapore: Springer; 2020. p. 41-54. https://doi.org/10.1007/978-981-33-4582-9_4
4. Padol PB, Yadav AA. SVM classifier based grape leaf disease detection. 2016 Conference on Advances in Signal Processing (CASP); 2016. p. 175-9. https://doi.org/10.1109/CASP.2016.7746160
5. Chaudhary P, Chaudhari AK, Cheeran AN, Godara S. Color transform based approach for disease spot detection on plant leaf. Int J Comput Sci Telecommun. 2012; 3(6):65-70.
6. Mathur MA. Analysis of various plant disease detection techniques using KNN classifier. Int J Comput Sci Mobile Computing. 2019; 8(7):65-70.
7. Krithika N, Selvarani AG. An individual grape leaf disease identification using leaf skeletons and KNN classification. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS); 2017. p. 1-5. https://doi.org/10.1109/ICIIECS.2017.8275951
8. Amri I, Hamrouni L, Hanana M, Jamoussi B, Lebdi K. Essential oils as biological alternatives to protect date palm (*Phoenix dactylifera* L.) against *Ectomyelois ceratoniae* Zeller (Lepidoptera: Pyralidae). Chilean J Agric Res. 2014; 74(3):273-9. https://doi.org/10.4067/S0718-58392014000300004
9. Amarti F, Satrani B, Ghanmi M, Farah A. Chemical composition and antimicrobial activity of essential oils of *Thymus algeriensis* Boiss. and Reut. and *Thymus ciliatus* (Desf.) Benth from Morocco. Agron Soc Environ. 2010; 14(1):141-8.
10. Azelmat K, ElGarrouj D, Mouhib M, Sayah F. Irradiation of 'Boufeggous' dates: Effects on chemical composition during storage. Postharvest Biol Technol. 2006; 39(2):217-22. https://doi.org/10.1016/j.postharvbio.2005.10.008
11. Rani KSS, Raja SP, Sinthuja M, Banu BV, Sapna R, Dekeba K. Classification of EEG signals using neural network for predicting consumer choices. Comput Intel Neurosc. 2022. https://doi.org/10.1155/2022/5872401
12. Sudhan MB, Sinthuja M, Raja SP, Amutharaj J, Latha GCP, Rachel SS, *et al*. Segmentation and classification of glaucoma using U-net with deep learning model. J Healthc Eng. 2022. https://doi.org/10.1155/2022/1601354
13. Umamaheswari K. Multimodal biometrics for improving automatic teller machine security. Bonfring Int J Adv Image Process. 2011; 1(1):19-25. https://doi.org/10.9756/BIJAIP.1005
14. Raja SP, Rachel SS, Sapna R. Women's safety with a smart foot device. 2021 4th International Conference on Computing and Communications Technologies (ICCCT); 2021. p. 570-3). https://doi.org/10.1109/ICCCT53315.2021.9711778
15. Akshatha Y, Raja SP. Certain investigations on different mathematical models in machine learning and artificial intelligence. Simulation and Analysis of Mathematical Methods in Real-Time Engineering Applications; 2021. https://doi.org/10.1002/9781119785521.ch1
16. Raja SP, Umamaheshwari K, Soumiya B, Kokila B. Investigations on finger nail plates authentication. Int J Appl Eng Res. 2014; 9(21):5094-9.
17. Raja SP, Umamaheswari K, Kokila B, Soumiya B. Analysis on continuous authentication system using multimodal biometrics. Int J Appl Eng Res. 2014; 9(21):5082-7.
18. Sinthuja M, Evangeline D, Raja SP, Shanmugarathinam G. Frequent itemset mining algorithms-a literature survey. Intelligent Sustainable Systems: Proceedings of ICISS 2021; 2021. p. 159-66. https://doi.org/10.1007/978-981-16-2422-3_13