



ISSN: 0067-2904

PFDINN: Comparison between Three Back-propagation Algorithms for Pear Fruit Disease Identification

Samar Amil Qassir

Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq

Received: 20/2/2020

Accepted: 23/7/2020

Abstract

The diseases presence in various species of fruits are the crucial parameter of economic composition and degradation of the cultivation industry around the world. The proposed pear fruit disease identification neural network (PFDINN) frame-work to identify three types of pear diseases was presented in this work. The major phases of the presented frame-work were as the following: (1) the infected area in the pear fruit was detected by using the algorithm of K-means clustering. (2) hybrid statistical features were computed over the segmented pear image and combined to form one descriptor. (3) Feed forward neural network (FFNN), which depends on three learning algorithms of back propagation (BP) training, namely Scaled conjugate gradient (SCG-BP), Resilient (R-BP) and Bayesian regularization (BR-BP), was used in the identification process. Pear fruit was taken as the experiment case during this work with three classifications of diseases, namely fire blight, pear scab, and sooty blotch, as compared to healthy pears. PFDINN framework was trained and tested using 2D pear fruit images collected from the Fruit Crops Diseases Database (FCDD). The presented framework achieved 94.6%, 97.3%, and 96.3% efficiency for SCG-BP, R-BP, and BR-BP, respectively. An accuracy value of 100% was achieved when the R-BP learning algorithm was trained for identification.

Keywords: Pear fruit diseases, k-means clustering, First and second order statistical features, Scaled conjugate gradient backpropagation SCG-BP, Resilient backpropagation R-BP, Bayesian regularization backpropagation BR-BP.

مقارنة بين ثلاثة خوارزميات للانتشار العكسي لتحديد أمراض فاكهة الكمثرى

سمر اميل

قسم الحاسبات، كلية العلوم، الجامعة المستنصرية، بغداد، العراق

الخلاصة

إن وجود الأمراض في العديد من أنواع الفاكهة هو معيار مهم للتكوين الاقتصادي وتدهور صناعة الزراعة في جميع أنحاء العالم. يعرض هذا العمل المقترح تشخيص لأمراض فاكهة الكمثرى باستخدام إطار عمل الشبكة العصبية (PFDINN) لتحديد ثلاثة أنواع من أمراض الكمثرى. تم الكشف عن المنطقة المصابة في فاكهة الكمثرى باستخدام خوارزمية التجميع K-means، يتم حساب الميزات الإحصائية الهجينة من صورة الكمثرى المجزأة ودمجها لتشكيل واصف واحد، ثم تطبيق الشبكة العصبية الأمامية لإجراء عملية التصنيف (FFNN) الذي يعتمد على ثلاث خوارزميات للانتشار الخلفي: (Resilient R-BP، SCG-BP)، Bayesian Regulation BR-BP) في عملية تحديد نوع المرض. تم أخذ فاكهة الكمثرى كحالة للتجربة خلال هذا العمل مع ثلاثة تصنيفات للأمراض: اللقحة النارية، والكمثرى المعدي والبقع السخيفة بالإضافة إلى

الكمثرى الصحية . تم تدريب إطار العمل (PFDINN) واختباره باستخدام صور فاكهة الكمثرى ثنائية الأبعاد، التي تم جمعها من قاعدة بيانات أمراض محاصيل الفاكهة (FCDD). يحقق إطار العمل المقدم كفاءة 94.6% و 97.3% و 96.3% لـ SCG-BP و R-BP و BR-BP بالتسلسل ؛ و 100% ، عندما تم تدريب خوارزمية التعلم (R-BP).

1 Introduction

An identification system can be defined as a one-to-many matching system of computer vision to obtain the closest levels to human recognition [1]. In agricultural sciences, one important source of information and data are images [2]. The analysis of digital images and the technology of image processing are the fundamental tools to obtain identification systems in the computers and micro-electronics. These tools support developing images of microscopic to telescopic visual range, then suggest a domain for the appropriate analysis [3]. The PFDINN framework has been used for implementing power-driven systems of cultivation procedures by utilizing images from pears fields. Plant diseases can manifest as spots on fruits and, if not treated on the exact time, they would cause serious damages. Immoderate uses of the pesticides for a fruit disease recovery expands the risk of poisonous level of residue on a cultivation products and was specified as an important instrument to the contaminant of the ground water. Pesticides can be also regarded as one of the most influencing factors on production cost and, thus, their utilization must be kept to a minimum [4]. Therefore, there are many attempts to get a method which can be identified the diseases within the fruits the moment that they provide their marks on the resurgent fruits such as an appropriate administration application which can be applied [4] . A new framework is presented to find and specify pear diseases from images. The objective of this PFDINN framework is to detect and identify three types of pear fruit diseases, namely fire blight, pear scab, and sooty blotch. The first type is caused by a common bacteria, called *Erwinia amylovora* (E.A.). The second type is caused by the fungus *Venturia pirina* (V. P.) and manifests as circular spots. The last type is caused by many different fungi [5, 6]. In this paper, k-means algorithm was used in the first phase to detect the infected areas of pear fruit. In the second phase, hybrid statistical features were extracted from the infected areas. In the final phase, FFNN with three learning algorithms were used. The rest of the paper consists of the following sections. In section 2, related work is reviewed. Section 3 provides a background of methods used in this work. Section 4 describes the proposed PFDINN framework. In section 5, the results of the experiments are presented. Section 6 discusses conclusions and future works.

2 Related Work

Some related works of fruits, vegetables, and grain galleries have been previously published. Sandika *et al.* described a process of potato disease classification using decorrelation stretching technique to optimize the differences of color in the loaded images and then applying a fuzzy C-mean clustering to the segment which has been affected by the disease. Finally, they suggested to utilize an approach based on neural networks to configure areas based on the similarity of color textured background. The suggested algorithm achieved an accuracy of 93% from 27 images captured in various light sensitivity values, with various distances, and at various positions [3]. Pawan *et al.* presented a method to detect and diagnose cotton diseases to prevent heavy loss of crops. This kind of diseases can lead to a severe reduction in the production process. In this research, we first applied a preprocessing step to images from the histogram quantization to increase the contrast of images with low contrast level. K-means clustering algorithm was used for segmentation, which classifies objects based on a set of features into a K number of classes and finally classifies their performance using neural network [7].

3 Background on Methods

The main approach adopted in this work of detection and identification of pear diseases depends on k-means clustering for the segmentation phase, whereas hybrid statistical features are extracted in the second phase, and FFNN with three learning algorithms are used for the identification phase. Each of these methodologies will be explained in the following sections. Figure-1 shows the three types of diseases exhibited in the infected pear images utilized in this work.

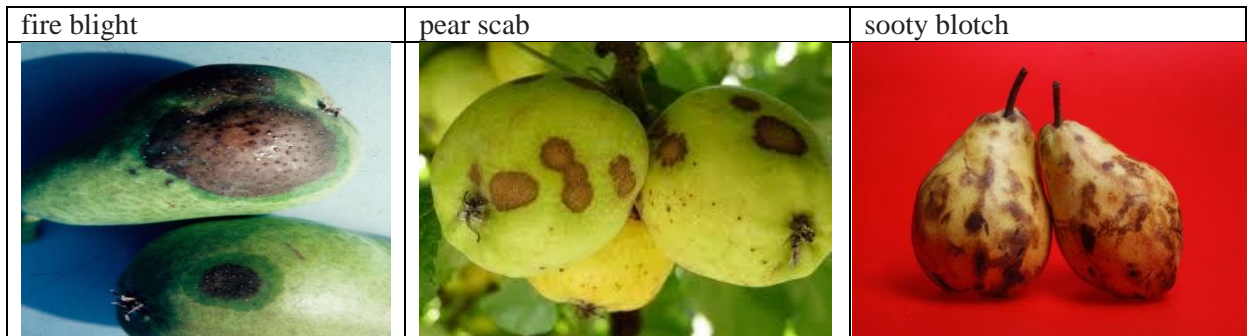


Figure 1- The three types of Pear fruit disease: the first is fire blight, the second is pear scab and lastly is sooty blotch. The first type is caused by common bacterium (E. A.). The second type is caused by the fungus (V. P.) where its shape likes circular spots. The final type is caused by various fungi

3.1 K- Means Clustering Algorithm

The disease detection is the extraction of the affected area by disease out of the fruit image. This helps to extract the affected segment out the whole healthy fruit portion . The segmentation divides the image into significant partitions for better analysis and comprehending of the image. Thresholding and growing areas can be described as two fundamental methods of segmentation. It is the most facilitated whereas the gray image can be segmented according to the threshold value [8]. This method does not supply an effective segmentation of the colored image and then limits the categorization of the different items of the image if it includes multiple areas or colors. Furthermore, it decreases the proper detection of the desired area [9]. K-Means clustering is a simple, fast, and iterative method [10]. This method has been utilized due to: firstly, it is almost all the time mathematically faster than the algorithm of hierarchical clustering, if k is kept little. Secondly, tighter clusters than hierarchical clustering [10], [11]. The flowchart for K means clustering is shown in Figure-2. It is worked by using “centroids”, K various randomly points as the initial values in the data, and it gives every data-point to the closest centroid. After every point assignment, the centroid is proceeded to the average of all the points which have been given to it [12]. It is described in the equation (1):

$$J = \sum_{p=1}^k \sum_{q=1}^n \|a_p - c_q\|^2 \quad (1)$$

Where j is stands for the objective function; the P is stands for the number of clusters; the Q is the number of cases; A represents the case q and finally the C is stands for the centroid for cluster p.

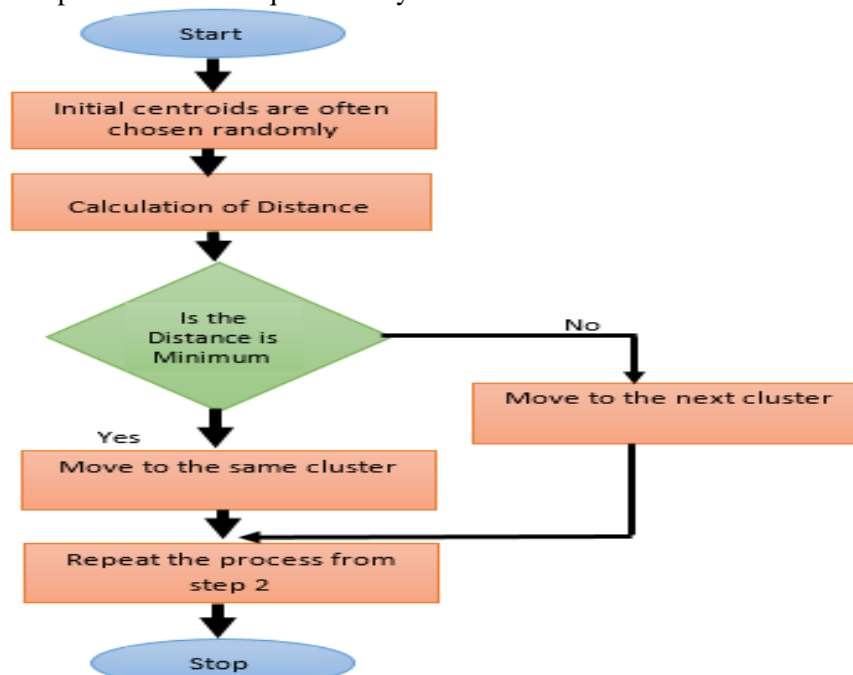


Figure 2- flowchart for k-mean clustering algorithm

3.2 Statistical Features

In colour and texture feature extraction analysis, the image texture shows information about the locative positioning of a colouring image. Two kinds of statistical features are used in the proposed PFDINN framework for extracted features: Colour Co-occurrence Method (CCM method) and Features Based on First and Second Order Statistics.

3.2.1 Colour Co-occurrence Method (CCM method)

A colour is extensively utilized for visual attribute and also it is mostly utilized to make a comparison between images [13]. This approach extracts the texture and colour out of any image. In CCM; The angular moment (H), as explained in equation (2), it is utilized to detect homogeneity out of the image, where N is the whole number of strength and $f(x, y)$ represents the image. The intensity of covariance (C) matrix is given in equation (3). The entropies in expressions of sum (S) and Difference (D) are given in equations (4 and 5) where (a, b) are the referral pixel at image location and $k=I-x$ or $k=I-y$ where I is strength of the image [14].

$$H = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} [f(x, y)]^2 \quad (2)$$

$$C = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (x - l_2)(y - l_2) \times f(x, y) \quad (3)$$

$$S = \sum_{x=0}^{N-1} f_{a+b}(k) \ln f_{a+b}(k) \quad (4)$$

$$D = \sum_{x=0}^{N-1} f_{a-b}(k) \ln f_{a-b}(k) \quad (5)$$

3.2.2 Features Based on First and Second Order Statistics

The features are described to quantify properties of an image region by exploiting space relations underlying the gray level distribution of a given image [15]. Figure-3 explains the statistical features classification.

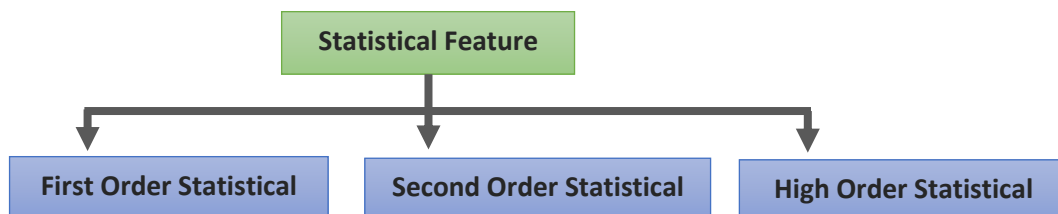


Figure 3- Statistical based features

The first order histogram is used to obtain various statistical properties that depend on individual pixel values to depict the region of interest. In this statistical classification, the intensity mean (IM), standard deviation (Std), which is the number used to tell how measurements of a group are widespread from the average. The skewness (Skn) of an allocation is the number that indicates the degree of asymmetry of the distribution. The kurtosis (kt) of a distribution is the number that indicates the sharpness of a peak distribution in relative to the normal distribution, which can be extracted from the diseased area of pear fruit since these features give good indication for the intensity distribution [16]. Each of these measures is illustrated in equations (6), (7), (8), and (9).

$$IM = \frac{1}{N} * \sum_x \sum_y f(x, y) \quad (6)$$

$$Std = \sqrt{\frac{\sum_x \sum_y (f(x, y) - IM)^2}{N}} \quad (7)$$

$$Skn = \frac{1}{N} * \sqrt{\frac{\sum_x \sum_y (f(x,y) - IM)^3}{Std^3}} \quad (8)$$

$$Kt = \frac{1}{N} * \sqrt{\frac{\sum_x \sum_y (f(x,y) - IM)^4}{Std^4}} \quad (9)$$

Where $f(x,y)$ is symbolize the intensities of diseased area and N is the number of pixels in the diseased area. In second order statistics; the spatial grey level dependence matrix algorithm is easy to implement, which provides information on image homogeneity and can be computed in equation (10). Entropy is the measure of the disorder of the spatial gray level dependence matrix it measures the disorder of the spatial gray level dependence matrix. It is computed by equation (11). The correlation is the measure of the level of grey linear reliance between pixels at the determined locations, which can be computed in equation (12). Local homogeneity is computed in equation (13), and inertia is quantifying local variations presented in the image. It is computed in equation (14) [15].

$$ASM = \sum_{i,j} P(I_1, I_2)^2 \quad (10)$$

$$Contrast = \sum_{I_1, I_2} |I_1 - I_2|^2 \log P(I_1, I_2) \quad (11)$$

$$Correlation = \sum_{I_1, I_2} \frac{P(I_1 - \mu_1)P(I_2 - \mu_2)P(I_1, I_2)}{\sigma_1 \sigma_2} \quad (12)$$

$$Homogeneity = \sum_{I_1, I_2} \frac{P(I_1, I_2)}{1 + |I_1 - I_2|^2} \quad (13)$$

$$Entropy = - \sum_{I_1, I_2} P(I_1, I_2) \log P(I_1, I_2) \quad (14)$$

3.3 Neural Network Method

For identification process, there are various methodologies used in researches nowadays, like artificial neural networks, support vector machines and fuzzy logic etc. The back propagation training is a familiar method for practicing artificial neural network [17]. The first level is representing network training while the second level represents testing, So that, choosing the input points and weighting operations are major factors. This model of neural network is a multilayer feed forward passes in two kinds of phases, which are known as propagation and weight upgrade. Three algorithms of back propagation training utilized in this work; the first is depends on Scaled conjugate gradient SCG-BP [19], the second is Resilient R-BP [19] and the third is Bayesian regularization BR-BP [20], [21]. In this work; normalization based on min- max method was utilized as explain in equation below:

$$U_i = \frac{(X_i - X_{Min})}{(X_{Max} - X_{Min})} * (r - t) + t \quad (15)$$

Where U_i : The feature input from 1 to 13 value; t, r are the normalization range; X_{Min} , X_{Max} are the upper and lower values.

4 The Presented PFDINN framework

The goal of this work can be described by detecting and identifying three types of pear fruit diseases; which where fire blight, pear scab and sooty blotch. The PFDINN framework pass through three main phases. In the first phase; the K-Means clustering algorithm is used to locate the infected region out of the pear image; hybrid texture features are extracted in second phase. While in third phase the FFNN using three learning algorithms was utilized for disease identification process, as shown in the Figure-4.

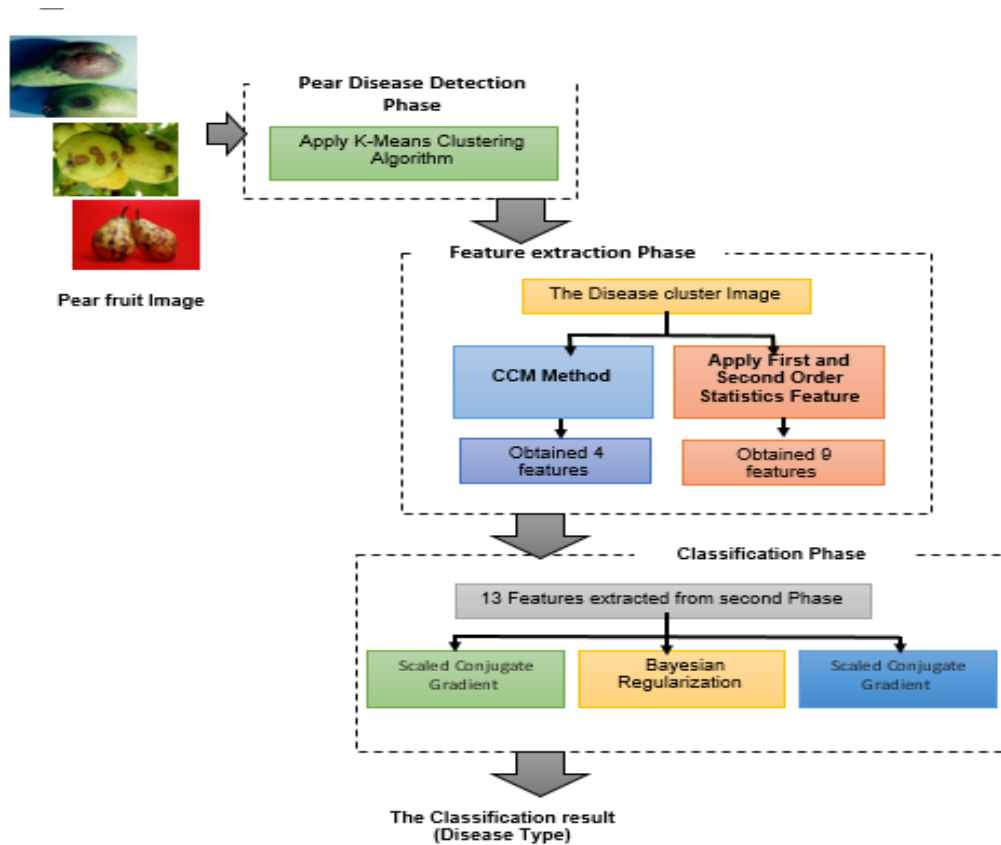


Figure 4- Presented PFDINN framework diagram

4.1 Image Acquisition stage

It is the first step of any identification system, in the presented framework of the pear fruit 2D image was collected from Fruit Crops Diseases Database (FCDD). FCDD is entirely searchable and maintains comprehensive text search. The primary focus of the FCDD is on supplying a possible information of fruit harvest diseases.

4.2 Pear Disease Detection Phase

The first phase of proposed PFDINN framework aims to extract the infected area from pear fruit images based on utilizing k-means clustering approach. Figure-5 shows some results of this phase. The value of K that used in this work was 3.

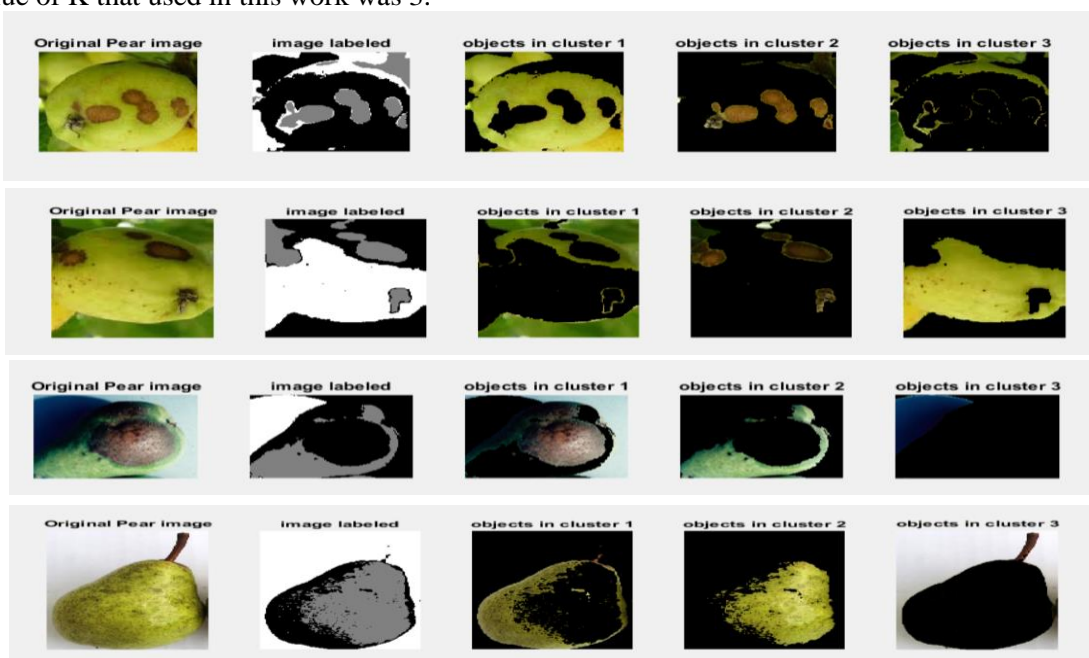


Figure 5-Pear Disease Detection Phase

4.3 Feature extraction Phase

In presented algorithm thirteen significant texture features are extracted from the infected area of pear fruit image. In given PFDINN framework, two types of statistical features calculations were used: Color Co-occurrence Method (CCM method) based features, hybrid with the features which are based on the First and the second order statistics: four texture features extracted from the first method and nine texture features extracted form the second method, as explained in section (3.2).

4.4 Identification Phase

The Identification process uses three network training algorithms: (SCG-BP, BR-BP, and R-BP). To generalize nonlinear relationships of inputs and outputs dataset; SCG-BP, BR-BP, R-BP were used, the neural network uses two layers of feed forward (FFNN), which are the weight and bias values were updated according to each learning algorithm. The measure of error utilized in this work was cross entropy error (CE error). The sigmoid trigger function was utilized for operating nerve cells in the hidden layer. This supervised learning topology of back propagation has a priority because of its implementation simplicity, as well as the availability of adequate “input – target” database for training.

4.4.1 Scaling and Training Step

In this work; a min-max scaling was essentially utilized to prevent larger numbers from overriding smaller ones, as indicated in section (3.3). Pear-Inputs matrix and Pear-Targets matrix; represent the (Input-Target) layers for the FFNN, the Pear-Inputs matrix represents the texture features extracted from the infected area of pear fruit images. 178 input image dataset were obtained from FCDD dataset of fresh pear fruit productions which contains the three types of pear diseases. The input matrix size was 13×178 ; (13) hybrid texture feature values extracted from each pear image of (178) image, the targets matrix size was 3×178 where each column indicates a disease category with a one in either element 1, 2 and 3 or zero for the rest elements, using binary codes [1 0 0]; [0 1 0]; [0 0 1]. The accessible image dataset was arranged as follows: Training image-set: 124 elements pear images; Validation image-set: 27 pear images and Testing image-set: 27 pear images. As the validation error starts increasing; the training process has been ended. After terminating the training procedure, the trained network was formed.

4.4.2 Testing Step

The execution results of R-BP can be shown in Figure-6; and the comparison with other two training algorithms explains in table (1). It is obvious that the minimum squared mistake in case of BR-BP approach is obtained at 23 intervals, its gradient was $9.1143e-08$ and in case of SCG-BP approach it has required 18 intervals, its gradient was 0.00062873 . In case of R-BP, its gradient was 0.0005197 and achieved at 14 epochs.

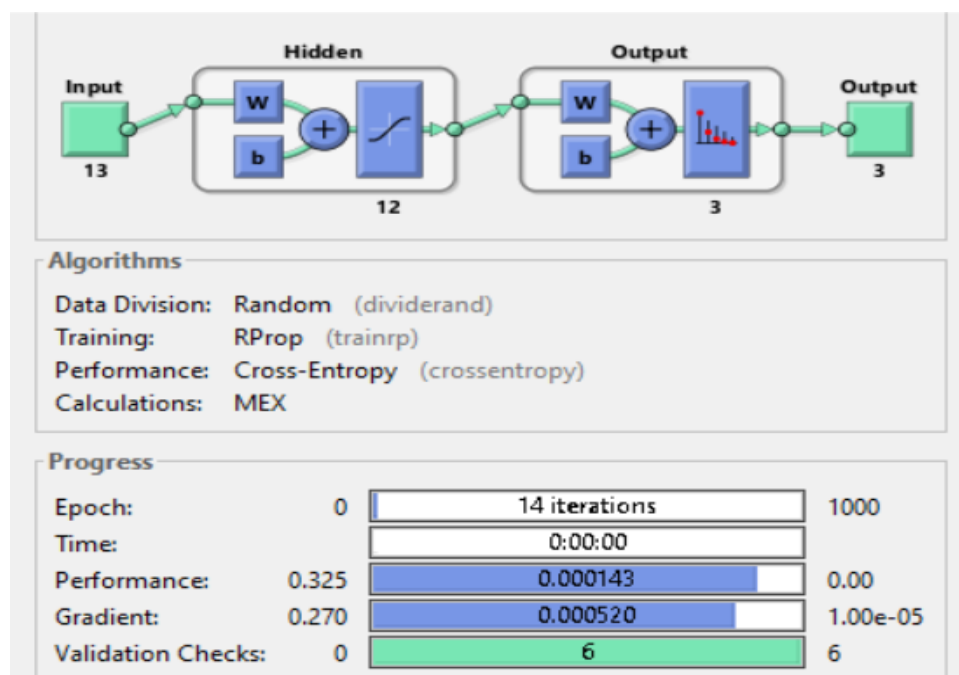


Figure 6- the Result of FFNN using R-BP with 12 hidden nodes in hidden layer

Table 1-The performance of training and testing of SCG-BP, BR-BP and R-BP in sequence

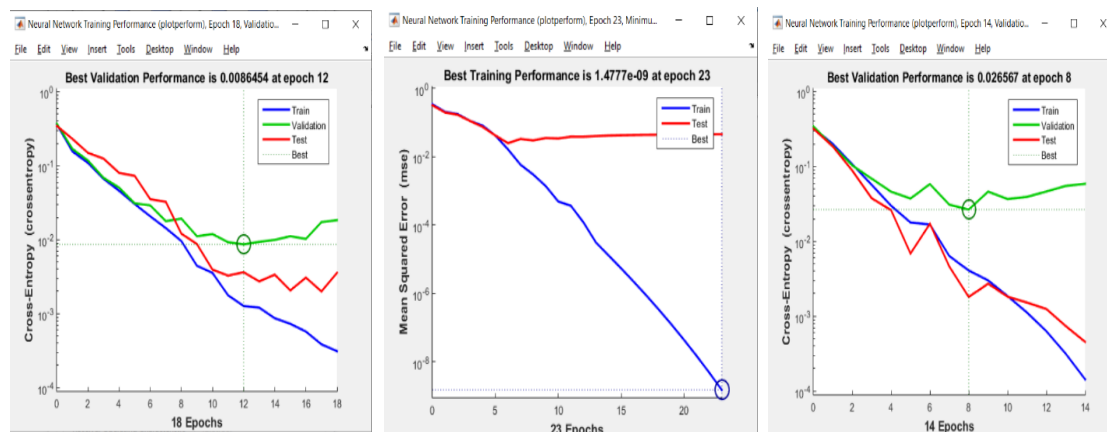
| NN elements | SCG-BP | R-BP | BR-BP |
|--------------------|-----------|----------|----------|
| No. of Hidden Node | 12 | 12 | 12 |
| No of Iteration | 18 | 14 | 23 |
| Performance | 0.0086454 | 0.026567 | 1.48E-09 |
| Training | 100% | 100% | 100% |
| Validation | 100% | 96.3% | NaN |
| Testing | 94.6% | 97.3% | 96.3% |

5. Experimental Results

The whole framework was implemented using Matlab software tool box. 178 pear fruit image samples were considered from dataset FCDD. 124 pear images were used for training step; 40 samples of fire blight disease, 40 samples of pear scab disease and 44 samples of sooty blotch disease. For identification of pear image samples using three BP learning algorithms (SCG-BP, R-BP and BR-BP), the network model in each case was instructed with 70% of the images of pear for each disease kind, and the residual 15% pear images can be utilized for each of testing and validation. The accuracy percentage can be defined as the proportion of the correctly recognized pear image samples according to the number of tested samples of pear image. The percentage of accuracy was obtained by the equation (16), and best execution of each neural learning model plot is drawn in Figure-7.

$$A = \frac{\text{correctly identify pear image samples}}{\text{the total number of test pear images}} \times 100 \quad (16)$$

The Neural model was implemented using 13 neurons for input layer, 12 for the hidden layer and 3 for the output layer. Table (1) summarizes the result obtained by each model, so it is obvious that R-BP (resilient back-propagation) needs less epochs than SCG-BP algorithm and BR-BP for the same iteration. Thus; the procedure of training was high performance in case of resilient approach than the other two approaches. The training execution and confirmation schemes showed that the R-BP approaches were faster training methods than (SCG-BP and BR-BP) methods. The transfer function in every network plays a prime role to determine the output values. The best Performance and Error Histogram, the latest shows how accurately the trained model fits the dataset of each neural learning model and is drawn in Figures- (7) and (8) respectively.

**Figure 7**-the Best Performance of FFNN using SCG-BP, R-BP, and BR-BP in sequence

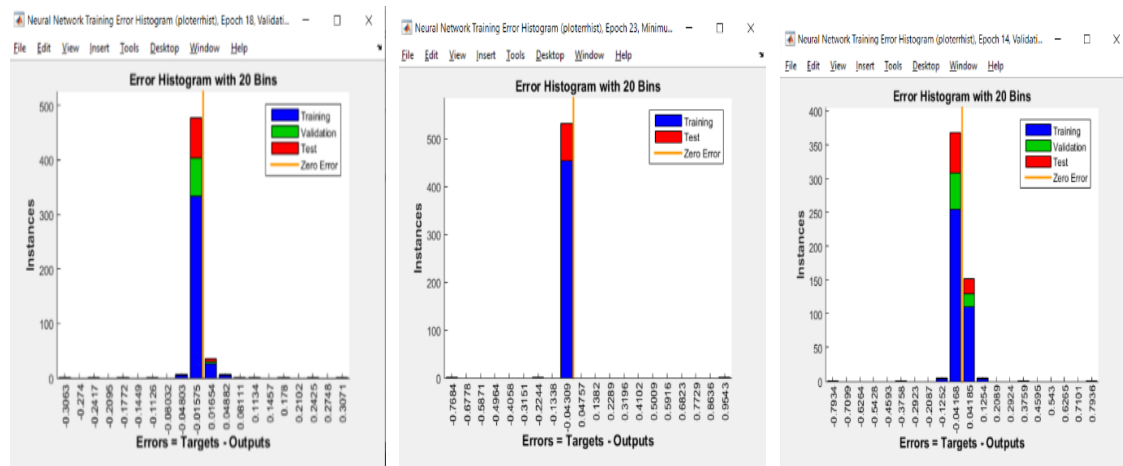


Figure 8- the Error Histogram of FFNN using SCG-BP, R-BP, and BR-BP in sequence

6. Conclusion

In the presented PFDINN framework, it seems very significant to have a demonstrated approach for categorizing the glitches on the pear fruit spontaneously. This work was pass with three main phases: Pear Disease detection, hybrid statistical features extraction and identification process. It was essential utilized for automatically uncovering the pear disease area in addition to identifying these disease types. This work has used K-means clustering algorithm for partitioning of pear fruit image to divide the healthy area, disease region and back-ground region of the inserted pear fruit image according to extracted hybrid statistical features, then fed these features to BP neural model for identification process. The presented framework qualified the identifications of three kinds of pear fruit diseases with 94.6%, 97.3% and 96.3% efficiency for SCG-BP, R-BP, and BR-BP respectively. The robustness in this work was offering an access for effective discrimination to these diseases. This system was utilized to take the place of the instructions pear identification methods, which can be used by cultivation proficient in recognizing the correct pesticide and its amount to get the problem under control in an organized and successful manner. In future work, to increase accuracy rate by using large number of pear fruit samples and the work implemented more diseases kinds by utilizing this approach.

ACKNOWLEDGMENTS

The authors would like to be a gratitude for Mustansiriyah University (www.uomustansiriyah.edu.iq) Baghdad-Iraq for its appreciated support in the present work.

References

1. Jagadeesh D. and Rajesh Y. **2015.** "Image Processing Based Detection of Fungal Diseases in Plants", *Procedia Computer Science*, **46**: 1802-1808, Published by Elsevier in 2015.
2. Monzurul Islam and Pankaj Bhowmik. **2017.** "Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine", 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE).
3. Sandika Biswas and Bhushan Jagyasi. **2014.** "Severity Identification of Potato Late Blight Disease from Crop Images Captured under Uncontrolled Environment", 2014 IEEE Canada International Humanitarian Technology Conference - (IHTC).
4. Milos I. and Petar S. **2015.** "Data mining model for early fruit diseases detection", 23rd Telecommunications Forum Telfor (TELFOR), Publisher:IEEE, Published in: 2015.
5. Ashwini Awate and Damini Deshmankar. **2015.** "Fruit disease detection using color, texture analysis and ANN", International Conference on Green Computing and Internet of Things (ICGCIoT), Published in 2015.
6. Rashmi Pawar and Ambaji Jadhav. **2017.** "Pomogranite disease detection and classification", International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Publisher: IEEE, in 2017.
7. Pawan P. Warne and Dr. S. R. Ganorkar. **2015.** "Detection of Diseases on Cotton Leaves Using K Mean Clustering Method", in 2015.

8. Vippon P. and Sakshi A. **2019**. "Fruit Disease Detection Using Rule-Based Classification", Springer, in 2019.
9. Trimi N. and Sushma K. **2017**. "Detection of plant disease using threshold, k-mean cluster and ANN algorithm", 2nd International Conference for Convergence in Technology (I2CT), Publisher: IEEE, Publisher: IEEE, Published in: 2017.
10. Usha K. and Jeevan P. **2019**. "Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN", 3rd International Conference on Computing Methodologies and Communication (ICCMC), in 2019.
11. Singh and Chetia. **2018**. "Diseased Leaf Detection Using K-Mean Clustering And Texture Features", *International Journal of Recent Scientific Research*, **9**(2(I)): 24427-24429, in February 2018.
12. Todkar V. **2019**. "Green Leaf Disease detection Using Raspberry pi", 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Publisher by IEEE, in 2019.
13. Manisha B. and H.A. Hingoliwalab. **2015**. "Smart Farming: Pomegranate Disease Detection Using ImageProcessing", *Procedia Computer Science*, **58**: 280 – 288, Elsevier, in 2015.
14. Shiv R. and Anand S. **2016**. "Apple disease classification using color, texture and shape features from images", Published in Signal, Image and Video, springer, in 2016.
15. Haris K. and Kohsuke N. **2015**. "Characterization of First and Second Order Statistics of Large Scale Fading Using Vehicular Sensors", ISSN: 1550-2252, *IEEE Xplore*, in 2015.
16. Jainy S. and Vinod K. **2013**. "Segmentation, Feature Extraction and Multiclass Brain Tumor Classification", *Journal of Digital Imaging*, springer in 2013.
17. Zafer C. and Adnan F. **2017**. "A Study of Artificial Neural Network Training Algorithms for Classification of Cardiocography Signals", *Bitlis Eren University Journal Of Science And Technology*, **7**(2).
18. Roseli G. **2019**. "Coffee Leaf Disease Recognition Based on Deep Learning and Texture Attributes", 23rd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Published by Elsevier, in 2019.
19. Ariel C. and Chandrasurya K. **2015**. "The Notation Scanner Systems Using Resilient Backpropagation Method", *Procedia Computer Science*, **59**: 98-105, International Conference on Computer Science and Computational Intelligence (ICCS CI 2015), Published by Elsevier in 2015.
20. H.R.Ansari and M.J.Zarei. **2018**. "A new comprehensive model for relative viscosity of various nanofluids using feed-forward back-propagation MLP neural networks", *International Communications in Heat and Mass Transfer*, **91**: 158-164, Published by Elsevier, in 2018.
21. Murat K. **2016**. "Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks", Department of Computer Engineering, Mu,s Alparslan University, in 2016.