PSO Trained Hybrid Intelligent Classifier Using Wavelet and Statistical Features for Pipeline Leak Classification

Hanan A. R. Akkar[1], Wael A. H.[2], Ibraheem H. M. Al-Dosari[3]

[1]Electrical Engineering Department, University of Technology Baghdad, Iraq.
[2]Communications Engineering Department, University of Technology, Baghdad, Iraq.
[3]Computer Communications Department Al-Rafidain University College Baghdad, Iraq.
dr_hananuot@yahoo.com, wail_alsaadi@yahoo.com, ibraheemdoser77@gmail.

Abstract— One of the main problems in the oil industrial field is the leakage in transporting pipelines due to its effect on human society , environment, and money loss. Therefore, the bottleneck for most researches in this subject is to minimize false alarm rate (FAR) for the adopted leak detection method. Although some recent methods succeed in classifying the existence or absence of the leak as a binary classification problem. But this paper proposed a novel leak detection technique which predicts the leak location and estimates its size within certain pre-defined ranges. In order to simulate the environmental conditions for real-time operating oil pipeline, accurate simulator known as OLGA program creates the oil physical parameters. Various methods for features extraction are considered such as statistical and wavelet techniques which are implemented to get the features from the fluid simulated waveforms. These features are organized and fed to an ANN classifier trained by PSO algorithm to determine the leak class out of 10 suggested classes. The proposed leak detection technique is used to simulate 18 kilometers belonging to Iraqi crude oil pipelines company operated in Baghdad. The achieved results of the true positive rate (TPR) for the proposed applied method for the leak detection and classification of different leak classes in terms of their positions and magnitudes were about 97%.

Index Terms— Particle Swarm Optimization (PSO) , Leakage Detection , Artificial Neural Network (ANN), Hybrid Classifier.

I. INTRODUCTION

Due to the importance of the energy generation for the human being life and since oil and gas play an important role in this energy foundation many methods are used to protect these products and transport them safely via pipelines that almost faced some faults and cracks that occur due to many causes such as earthquake or bad planning and installation for pipelines. Different methods have been suggested for oil pipelines leak detection. Literatures classified these methods into two major classes, hardware and software. In hardware-based technique usually a physical component such as transducers or sensors are used to measure the physical parameter and its variation to detect the occurrence of the leak in the pipeline. Infrared camera, optical fiber cable, acoustic and ultrasonic sensors are used as hardware based methods when they are installed along the pipeline. On the other hand, Software-based methods sometimes use the sensing elements to supervise the fluid in the pipeline and use some models to investigate for the leak occurrence based on some features and measurements such as Real Time Transient Modeling (RTTM), Mass balance, Negative Pressure Wave (NPW) and pressure point analysis (PPA) [1]. In recent years, many leak detection systems are proposed and developed by researchers to protect the oil pipelines from faults, which in turns avoid a loss of money and prevent

environmental hazards. The concept behind the NPW method is a sudden reduction in pressure appeared at the leakage location. Meanwhile a negative pressure signal is emerged and this signal is propagated from the leakage occurrence position to both ends of the pipeline terminals which will change both inlet and outlet pressures. and this degradation in pressure could be measured by a pressure sensor and analyzed to identify the leak position and magnitude using wavelet based multi-resolution transform analysis [2].Unfortunately sometimes only one feature such as fluid pressure is not absolutely enough to detect the leak and its location, so an additional feature is augmented in the measurement process, such as flow rate or temperature profile. Artificial neural network and machine learning are the most powerful tool that are used for leak detection and magnitude classification using different learning algorithms and kernels for both supervised and unsupervised methods. Leakage in the pipeline can be considered as a classification problem with both assumptions, binary either (leak or non-leak) and multi-class (leak, location, size, etc.). If the data set is available with some input features and the required target to train the ANN, we can get the system model that can be used as an intelligent classifier for the new real input data in the future [3]. Some simulators succeed in simulating the behavior of the fluid in the pipeline with the all-internal conditions, OLGA software can be considered as a simulator that will provide statistical features, which will be used as predictors for a designed classifier to identify the leakage. In related past works, various classifiers based machine learning (linear and nonlinear) were studied. The TPR for most models were calculated. The comparison evaluation results reached to true positive rate (TPR) of about 96% for the binary classification problem. However, the main drawbacks for these models were their failure in detecting the leak position and its magnitude. However, getting higher precise classifier achievements is proposed by using new augmented methods for combining different classifiers. This technique classifies the model to data level, discrimination level, and finalized by classification level. At the data level, the information is gathered from various transducers to establish an augmented waveform with adequate properties better than the input data. At the discrimination level, various features are created and combined together prior to classification process feeding. Finally, at the classification level, the obtained results are combined from different classifiers using adequate techniques yielding an outcome class. Some researchers adopted feature extraction for both time-based and spectral-based features from NPW signal and constructed a new feature vector that will be fed to the classifier such as MLP-ANN to yield a true positive rate (TPR) of 98%, which is better than using either time-based features or spectral based features individually [4]. In this work, a combination of machine learning, ANN, and wavelet analysis is used to solve the classification problem for the supervised learning process, in cooperation with OLGA simulator different scenarios are generated with various assumptions to simulate the real-time conditions for different sized leaks distributed along various sections along the pipeline. Time and wavelet based features extraction methods are used to minimize false alarm rate using MLP-ANN as a core for the intelligent classifier. Wavelet based denoioising method with a novel thresholding method is used to shrink the coefficients and reduce the noise as a pre-processing for the classification problem. The obtained results emphasis the effectiveness for the proposed work in identifying the leak occurrence and diagnosing its location and size.

II. BASIC THEORY

A- WAVELET TRANSFORM

The term "wavelet" is just a small wave. This small wave should have a rapid decay to zero and a minimum alternator oscillation, in negative and positive directions with different amplitudes. Wavelets are a good mathematical tool that can be used to extract information from different signals, which include images and audio signals. Mathematically, wavelet is a function with zero mean, which has energy that's potent in time. Many families of wavelets have been constructed and used in extracting

frequency and time information for different applications [2]. Wavelet based denoising technique is important for it's wide usages. It significantly reduces the noise and gives better localization analysis of one dimensional waveforms (or 2D matrices). Wavelet tool can be utilized for different applications such as wavelet denoising and compression, where different thresholding methods are proposed for reducing noise and shrinking the coefficients in different topologies. To extract the information from 1D waveform, the waveform is analyzed using filter banks theory, where two passes of the original data are made through two filters in parallel one LPF to extract the approximations coefficients and HPF to extract the details coefficients, after that the two filters outputs are decimated. Sometimes this procedure is followed 2 or more times to get more resolution and extract many features from the signal. However, these used filters have many special characteristics and features mentioned in wavelet theory literatures such as their finite impulse response and anti-imaging feature. Decomposing the signal into different levels can be done on details coefficients only, which is known as discrete wavelet transform (DWT). On the other hand, if both details and approximations coefficients are decomposed, this topology is known as discrete wavelet packet transforms (DWPT). In de-noising principle, different thresholding methods can be implemented such as (soft, hard, firm, non-negative garrote, Qian, custom, square-chebyshev, and sawtooth) [5].

B- ARTIFICIAL NEURAL NETWOTRK

Artificial neural network (ANN) plays an important role in many applications. Many problems emerged from surrounding life need an implementation of clustering, recognition, regression and classification, to solve these problems. Artificial intelligence and its sub-classes have outstanding magic performance in wide fields that need smart decision beginning from small assignments in binary classification passing through large clustering problems up to large recognition assignments. Industrial problems play an important role in the development of the artificial intelligence theory, such as leak detection and localization in oil pipeline [6]. Further, due to facilities and different architectures for the neural network, it found a successful application in localization of pipeline leaks and estimating its size by using Matlab software with its neural network toolbox and proper dataset to train the classifier model and validate it with a suitable test data [7].

C- BACKPROPAGATION (BP)

The backpropagation (BP) algorithm was suggested at the 80th of the past century for updating different neural network weights and validating the network for further test. The BP algorithm propagates the error between the target and the network output backward through the network. Learning rules are responsible for the weight adaptation for the artificial neural network. In addition, there exist many rules used for this learning process. All of these rules tend to minimize the error between the network output and the target value. Generalized delta learning rule is the most common usage for the BP algorithm with feed forward neural network and other supervised architectures. The accuracy for the network is measured by some factors like; the time consumed for training, number of epochs until the training stopped, the memory requirement for storing the ANN and the achieved MSE [8].

D- PERFORMANCE MEASURES

A confusion matrix is used to evaluate the adopted model performance in certain classification problems. Moreover, it discriminates regions of richly and poorly classifier performance. The matrix is organized as follows: each row represents the actual class, while each column represents the estimated class, such that the matrix diagonal will represent the crossing between them. If the diagonal tends to unity then this is an indication for the matching between the actual and estimated classes. In addition, colored based logic is applied in this matrix such that, green squared values indicate the good

performance and correct classification for the classifier model. However, if the squares are red then this is incorrect class or a misclassification is occurred. The values inside the cells represent the observation number. True Positive Rates (TPR), False Negative Rates (FNR) are additional performance measure for the classification process. For illustration, a false negative rate (FNR) of 0.1 means that the present model has 10% misclassification out of total samples .While a true positive rate of 0.8 means that 80% of the total samples is classified properly with this model [9].

III. METHOLODOGY

The proposed leakage detection system has several steps that will be explained in this section. Starting with data acquisition, different leak classes are simulated in the OLGA software; after that, AWGN is augmented to under study data to simulate the real-time process. Secondly, the statistical and wavelet based features are individually extracted from the artificial data augmented by noise. However, these features are combined and entered to an intelligent classifier consisting of MLPNN trained by PSO algorithm. Finally, the performance measures and TPR scores are calculated for further comparison study.

A- PRE-PROCESSING

Most of problem solving techniques try to process the signal as a pre-conditioning way to ensure the ability for further processing to achieve the requirements. Sensing for any physical parameters cannot be done without an inherent noise. There are different types of noise occupied by different bands and have different characteristics. So, wavelet transform tool can be used for cancellation or at least reduction for the noise to enhance SNR for the signal under study. There are many wavelet mother functions and thresholding methods to denoise the signal. Coiflet with hard threshloding technique is used for denoising and generally minimizing MSE.

B- FEATURE SELECTION

Features can be classified into two domains: time domain and transformed (frequency, wavelet, and other transforms) domain features. Time domain features are based on characteristics for the signal that can be extracted from its behavior in time domain, such as average, standard deviation, moment or other statistical features. On the other hand transformed domain features consist of power spectral density PSD, energy, and entropy or other features. In this work, four features have been selected to classify the leaks in oil pipelines, which are mean, variance, PSD, and entropy [10].

C- PSO ALGORITHM

Particle swarm is based on the algorithm described by Kennedy and Eberhart , some modifications are suggested by Mezura-Montes and Coello. Also an improvement was suggested by Pedersen [11]. The particle swarm algorithm starts by asigning prilimenary population particles and an average speed for each of them, then the algorthem calculates the cost function and tries to minimize it based on certain criteria. PSO algorithm searching for the best particle which has the mninum error and its location is the nearest to the solution. This best particale location and velocity are used to improve the whole swarm and update their locations and velocities accordinglly [12].

In each iteration, any particle is updated by two "best" values. The first value is the best solution (fitness) it has achieved so far, which is called pbest. The second value that is tracked by the PSO is the best value, obtained so far by any particle in the population, which is known as the global best. After calculating the two best values, each particle updates its position and velocity with equations (1) and (2).

$$v[t]=v[t-1]+c1*rand(t)*(pbest[t]-present[t-1])+c2*rand(t-1)*(gbest[t]-present[t-1]) \qquad (1)$$

$$present[t]=present[t-1]+v[t] \qquad (2)$$

where : v[t] is the particle velocity, present[t-1] is the current particle (solution). pbest[] and gbest[] are the best position of particle and swarm, respectively, rand (t) is a random number between (0,1). c1 and c2 are learning constants. usually c1 and c2 equal 2.

The pseudo code of the PSO algorithm is summarized hereunder

-Start
-Initialize initial population of particles;
-Evaluate the fitness value;
-Repeat
-Change velocity of particles
 -Limit the velocity in [vmin, vmax];
-Produce new solutions/positions for the particles
-Apply the greedy selection process;
-Memorize the best solution achieved so far;
Until the stopping criteria

IV. RESULTS AND DISCUSSION

ANN used for leak detection where 10 leakage scenarios are adopted including the occurrence of a leak for veroius combinations of position and magnitude listed in tables 1,2 and 3. Extracted features are fed to the classifier. Where the input vector combined both statistical and wavelet based features having samples for all the classes mentioned in Table 3.

Recent scintific researchs introduced a leakage detection system generally depend on some features obtained from the physical parameters for the fliud flowing inside the pipeline. A novel approach try to use the concept of multi-classifier or combination for different classifier to improve the classification accuracy and TPR values .

FAR is a measure of how the classifier decline from the true value, so the adopted classifer for this work try to combine two domain features and improve the reslut for the confusion matrices as compared with indiviusaul domain based features. Table 4 illustrate a comparison for different classifiers adopted in this work and recent related works .Refering to table 4 it is abouvuise that 64% TPR for statistical based method. Transformed doamin method use wavelet coefficient as features for the classification purposes and hence improve the TPR to about 84%. And the confusion matrecies for both classifiers are shown in figures 1 and 2.

The hybrid classifier, suggested in this paper, significintly performed best among the three adopted classifiers in this research. although from figures 1 and 2, the statistical feature-basedclassifier showed good performance in class (1 and 3) .but the wavelet feature-based classifier showed better performance in all other leak classesas compared with statistical feature based classifier see figure 2 .

PSO algorithm was used to train the combined or hybrid classfier,simulation results showd a TPR of 97% for the new proposed classifer. The confusion matrix . The simulations for the hibrid classifier showd a TPR of about 97%. The prediction for 10 classes is presented in figure 3.

Figures 3 showd the classification accuracy of the hybrid features based classifier suggested in this paper. The diagonal sqaures tend to unity ,hence the pridection class matched the true class . Refering to figures 1,2 and 3 , although the overall accuracy for the proposed classier is better than the old classifiers. however, hybrid based PSO classifier (which is proposed in this paper) enhance the classification accuracy for classes 2,3,4 and 7.

The ANN clasifer has single hidden layer, with 10 neurons in the hidden and output layers. The activation function applied in hidden layer is logsigmoid function as shown in figure 4. About 75% of the total denoised data set (as shown in figure 5) is selected to train the MLPNN classifier based PSO algorithm. Once the classifier is trained, a part from the data is applied to test and validate the leak detection system for its ability to classify the leak size as in figure 6. The confusion matrix is then evaluated and the corresponding TPR value is calculated [13].

TABLE 1 DIFFERENT LEAK SIZES

Leak size	Size (in.)
Small (S)	Size <0.25
Intermediate (I)	0.25<Size<0.5
Large (L)	0.5< Size <1

TABLE 2 DIFFERENT LEAK CLASSES

Class no.	Case condition (location , size)
1	No Leak
2	S - B
3	S - M
4	S - E
5	I - B
6	I – M
7	I – E
8	L – B
9	L – M
10	L - E

TABLE 3 DIFFERENT LEAK LOCATIONS

Leak location	Loca*tion measured from beginning (km)
Beginning of the pipeline (B)	L<7
Middle of the pipeline (M)	7<L<13
End of the pipeline (E)	13<L<18

TABLE 4 PERFORMANCE COMPARISON FOR DIFFERENT INTELLIGENT CLASSIFIERS

Algorithms	Feature	TPR (%)
[14] PSO	Statistical	86.54
[14] EPSO	Statistical	86.54
[14] ABC	Statistical	95.19
[15] SVM	hybrid of statistical and wavelet	90.48
[15] LS-SVM	hybrid of statistical and wavelet	90.48
[15] LSTSVM	hybrid of statistical and wavelet	94.44

[16] ANN	Statistical	96.33
[16] SVM	Statistical	95.67
[16] Decision tree	Statistical	94.67
[16] Bayesian net.	Statistical	94.33
Our work PSO-ANN	Statistical	64
Our work PSO-ANN	Wavelet	84
Our work PSO-ANN	hybrid of statistical and wavelet	97

class	1	2	3	4	5	6	7	8	9	10
1	0.81	0	0.03	0.16	0	0	0	0	0	0
2	0	0.84	0.12	0.04	0	0	0	0	0	0
3	0.04	0.15	0.45	0.13	0	0	0.25	0	0	0
4	0.16	0.01	0.09	0.64	0.01	0	0.05	0	0	0
5	0	0	0.31	0	0.99	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0.04	0	0	0.7	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

FIG. 1 CONFUSION MATRIX FOR CLASSIFIER WITH WAVELET FEATURES BASED TECHNIQUE

class	1	2	3	4	5	6	7	8	9	10
1	0.96	0	0.03	0.01	0	0	0	0	0	0
2	0	0.77	0.17	0.06	0	0	0	0	0	0
3	0	0.14	0.53	0.09	0	0.02	0.23	0	0	0
4	0.04	0.1	0.14	0.63	0	0	0.05	0.05	0	0
5	0	0	0.11	0.1	0.72	0.07		0	0	0
6	0	0	0	0.09	0	0.82	0.09	0	0	0
7	0	0	0.03	0.03	0	0.09	0.63	0.24	0	0
8	0	0	0	0	0.28	0	0	0.08	0.64	0
9	0	0	0	0	0	0	0	0.64	0.26	0
10	0	0	0	0	0	0	0	0	0.11	1

FIG. 2 CONFUSION MATRIX FOR CLASSIFIER WITH STATISTICAL FEATURES BASED TECHNIQUE

class	1	2	3	4	5	6	7	8	9	10
1	0.9	0	0	0.01	0	0	0	0	0	0
2	0	0.95	0	0.01	0	0	0	0	0	0
3	0.04	0.03	0.95	0.03	0	0	0	0	0	0
4	0.07	0.01	0.01	0.96	0.01	0	0.03	0	0	0
5	0	0	0.04	0	0.98	0	0	0	0	0
6	0	0	0	0	0.01	1	0	0	0	0
7	0	0	0	0.01	0	0	0.97	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

FIG. 3 CONFUSION MATRIX FOR HYBRID CLASSIFIER WITH PSO BASED TECHNIQUE
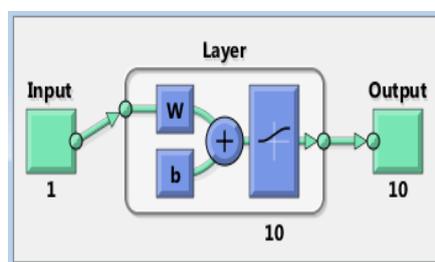
FIG. 4 NEURAL NETWORK ARCHITECTURE USED FOR CLASSIFICATION
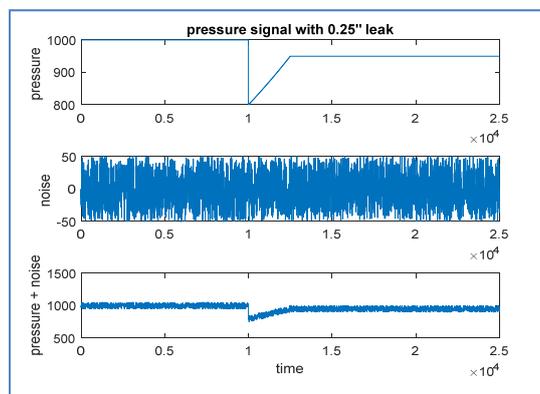
FIG. 5 PRESSURE SIGNAL FOR 0.25" LEAK
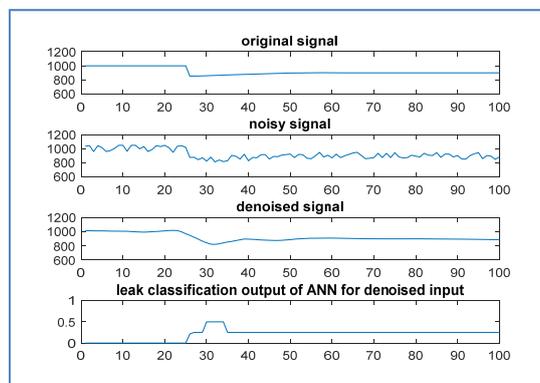
FIG. 6 DETECTION FOR 0.25" LEAK

V. CONCLUSIONS

Leakage reperents a major problem in the oil industrial field, due to its catastrophic effects on environmental and economical life. In this work, a novel leak detection method was suggested which significiently decreases the false negative rate (FNR). Artificial intelligence has achieved better resluts in leak detection solving problems. So, due to the difeculties of getting real-time data for leaked pipeline, simulators are required to create a proper model and dataset for this problem. This study is implemented on part of the iraqi oil pipeline simulated by the OLGA software. After that this dataset is used to train the classifier model and a hybrid classifier shows better performance than the classifier based on single domain features type. The PSO trained classifier outperformed other classifiers from TPR and the classification accurecy was read from the confusion matrix.

REFERENCES

[1] H. Saeed, S. Ali, S. Rashid, S. Qaisar, and E. Felemban, "Reliable monitoring of oil and gas pipelines using wireless sensor network (WSN)," in Proceedings of the IEEE 9th International System of Systems Engineering Conference, Pages 230-235, 2014.

[2] S. Rashid, S. Qaisar, H. Saeed, and E. Felemban, "Performance analysis of leak detection algorithm in long range pipeline networks using transform analysis," in Proceedings of the IEEE Conference on Systems, Process and Control (ICSPC '13), pp. 10–15, 2013.

[3] Akkar H. , Hadi W. , Al-Dosari I., "Pipelines Leak Classification Based on Artificial Intelligence", International Journal of Engineering Research And Management (IJERM) Vol. 5, Issue-07,pp.53-61 July 2018.

[4] Zhang, T., Yufei, T., Zhang, T., Zhao, J.,"A novel hybrid technique for leak detection and location in straight pipelines". J. Loss Prev. Process Ind. 35, pp.157–168, 2015.

[5] Akkar H., Jasim F., "Intelligent Training Algorithm for Artificial Neural Network EEG Classifications", International Journal of Intelligent Systems and Applications 10 (5), pp.33-41, May 2018.

[6] Akkar H., AL-Samarrie A., Saeed A., " Simulation Design of Back propagation Neural System of Sensor Network Trained by Particle Swarm Optimization", International Journal of Scientific & Engineering Research (IJSER), Volume 7, Issue 4, April-2016.

[7] Akkar H., Hadi W. , Al-Dosari I., "Design and Implementation of Intelligent Classifier and Size Estimator for Leakage in Oil Pipelines", International Journal of New Technology and Research (IJNTR), Vol.4, Issue-7, July, Pages 34-47, 2018.

[8] Akkar H., Mahdi F., "Training Artificial Neural Networks by PSO to Perform Digital Circuits Using Xilinx FPGA", Engineering and Technology Journal 29 (7), pp.1329-1344, 2011.

[9] Dang, T. L. and Hoshino, Y., "An FPGA based classification system by using a neural network and an improved particle swarm optimization algorithm," Soft Computing and Intelligent Systems, pp. 97-102, 2016.

[10] Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A. P., Bishop, R. and Rueckert, D., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," The IEEE Conference of Computer Vision and Pattern Recognition, pp. 1874-1883, 2016.

[11] Mohammadi, N. and Mirabedini, S. J., "Comparison of particle swarm optimization and back propagation algorithms for training feed forward neural network," Journal of Mathematics and Computer Science, no. 12, pp. 113-123, 2014.

[12] Koohi, I. and Groza, V., "Optimizing particle swarm optimization algorithm," Canadian Conference on Electrical and Computer Engineering, vol. 27, pp.1-5, 2014.

[13] Akkar H. , Hadi W. , Al-Dosari I., "Comparison Evaluation for Different Learning Algorithms in Classification Problem Based on ANN", International Journal of Engineering and Techniques (IJET), Vol. 4, Issue 4,pp.143-157, July 2018.

[14] S. Mandal , F. Chan, M. Tiwari," Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM", Expert Systems with Applications 39, pp.3071–3080, 2012.

[15] X. LANG, P. LI, Z. HU, H. REN, Y. LI," Leak Detection and Location of Pipelines based on LMD and Least Squares Twin Support Vector Machine ",IEEE Access Vol. 5, pp. 8659-8666, 2017.

[16] K. Mollazade, M. Omid a, A. Arefi," Comparing data mining classifiers for grading raisins based on visual features", Computers and Electronics in Agriculture 84, pp.124–131, 2012.