Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

In-Situ Event Localization for Pipeline Monitoring System Based Wireless Sensor Network Using K-Nearest Neighbors and Support Vector Machine

Ahmed H. Hadi [a] , [b] **Dr. Waleed F. Shareef**

[a] Control and Systems Engineering Department, University of Technology, Iraq. Email: eng.ahmed.altajer@gmail.com

[b] Control and Systems Engineering Department, University of Technology, Iraq. Email: 60026@uotechnology.edu.iq

ARTICLE INFO

Article history:
Received: 18 /02/2019
Rrevised form: / /
Accepted : 05 /03/2020
Available online: 15 /09/2020

Keywords:

WSN, In-Situ, Pipeline Monitoring, Localization, PCA, SVM and KNN Classification.

ABSTRACT

Pipeline Monitoring Systems (PMS) benefits the most of recent developments in wireless remote monitoring since each pipeline would span for long distances which make conventional methods unsuitable. Precise monitoring and detection of damaging events requires moving large amounts of data between sensor nodes and base station for processing which require high bandwidth communication protocol. To overcome this problem, In-Situ processing can be practiced by processing the collected data locally at each node instead of the base station. However, this introduce a challenge to the limited resources available on the nodes. In this paper, a Wireless Sensor Network (WSN) was implemented for In-Situ Pipeline Monitoring System with proposed algorithms for event location estimation. The proposed algorithms include feature extraction (using ANOVA), dimensionality reduction using statistical procedure that is (Principle Component Analysis PCA) and data classification using supervised learning K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). The proposed system was tested on pipelines in Al-Mussaib Gas Turbine Power Plant. During test, knocking events are applied at several distances relative to the nodes locations. Data collected at each node is filtered and processed locally in real time in each two adjacent nodes. The results of the estimation is then sent to the supervisor at base-station for display. The results show the proposed system ability to estimate the location of knocking event.

MSC :

https://doi.org/10.29304/jqcm.2020.12.3.705

1. Introduction

Transportation of resources over pipelines is so common around the world on different regions duo to simple scheme, execution, and low functioning costs [1]. Oil and Gas pipelines are considerable work of technology that require wide level of maintenance in order to operate them with a long life. Any failure of pipelines during operation leads to accidental and huge cost. The damage, erosion, and metal lose are generally distinguished by visual examination, ultrasonic testing, radiography, thermography, acoustic emission, and in-line assessment tool. These processes of testing usually very sluggish, high cost, and time taking. Novel methods in Structural Health Monitoring

Corresponding author : *Ahmed H. Hadi* , *Dr. Waleed F. Shareef*

Email addresses: *ahmed.altajer@gmail.com*

Communicated by ***Alaa Hussein Hamadi***

(SHM) are necessary based maintenance can be accomplished professionally. These techniques involve response of structure in healthy condition as well as damaged condition in order to identify the location and harshness of damage [2].

2. Literature Review

There are many ways to accomplish Pipeline Monitoring System (PMS) depending on sensing methods like Visual inspection, Ultrasound, Radiographic, Thermography, Pressure, Chemical and Magnetic. Recently, many improvement projects introduced for PMS depends on Satellite, Smart Pigging, Drones, Smart Robots, and Dogs [3]. PipeNet [4] project for leak detection and localization was studied. Several signal processing algorithm were implemented such as wavelet transform, cross correlation, recognition of pattern; a sensor node hardware built by Intel mote, ARM7 core, and Bluetooth communication.
VibNet is introduced based on WSN for vibration monitoring. Triple axis vibration sensor ADXL335 used for acceleration measurements, sensing method depend on the received signal strength indicator (RSSI) algorithm while the communication section implemented by Zigbee transceiver [5].
LAN server and Raspberry Pi used in [6] for intelligent garbage monitoring method by using gas sensor, ultrasonic sensor, load cell, and moisture sensor. A similar design to detect and localize the leakage in water pipeline
In-situ WSN based learning techniques used to monitor interesting field by collecting a specific information, analyze it locally and transmit the result of classification to base-station for monitoring. More than one wired or wireless technologies have used on side of data collection. Many researchers preferred using wireless sensor nodes to avoid wiring and range problems. Data collected from pipelines may be pressure, acoustic, or water quality to take the advantages of WSN technology. The main advantages of WSNs are real-time monitoring and detecting leakages to prevent resources loss [1].
Supervised learning in one of the important data processing applications in machine learning depending on provide sets of input (datasets with labels), train the system and then find the output. Supervised learning can be categorized into classification and regression. Classification can be logic-based statistical learning (SVM) and instance-based (KNN) algorithms [7].
Machine learning gets more popularity in the recent years because of their swiftness to unexperienced scenarios and ability to solve complicated jobs, which are difficult to solve by using mathematical model [8]. The statistical features extraction are very important for classification or recognition process and the supervised learning K-Nearest Neighbors (KNN) has been used for classification and compare the results with Support Vector machine (SVM) [9].
Figure 1 shows the taxonomy of the techniques of machine learning which are four types of learning methods, KNN and SVM works as classification supervised learning machine while PCA work as dimensionality reduction that discussed in test the proposed system.

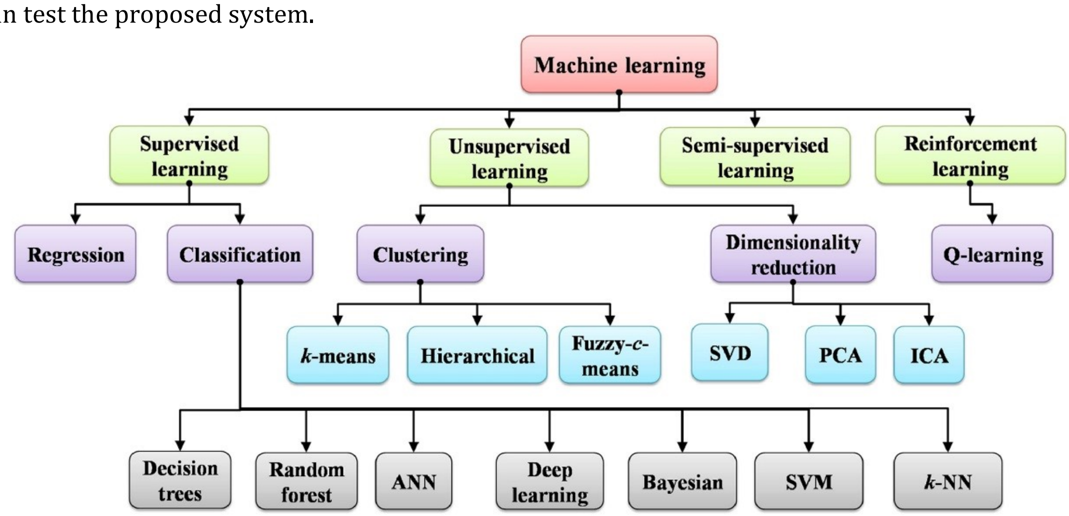

Fig.1 Taxonomy of Machine Learning Techniques [7]

This paper presents a design and implementation of In-situ structural health state monitoring of oil pipeline. Section 3 presents the wireless node hardware design, and then Section 4 describes the features extraction and dimensionality reduction. Section 5 presents description and results of the two methods for vibrational signal classification, while Section 7 concludes the paper.

3. Architecture of the Proposed System

The proposed system consists of two sensor nodes attached at the end of 24-meter carbon steel, plain ends oil pipe. Node No. 1 attached at 1 meter from right end beside of the pipe by Duct tape while Node No. 2 attached to the other end at 24 meter as shown in Figure 2. There are eight concrete stands installed every 6 meter to support the pipe. Table 1 shows the specification of the selected oil pipe, each part in the table should effect on the natural frequency that selected on the proposed system [10].

Table 1. Oil Pipeline Specifications

Pipeline Product	Crude oil	*Design Temperature*	85°C
Ends Type	Plain Ends	*Length*	24 m
Type of Pipe	Carbon Steel	*Height*	8 inch
Wall Thickness	6.3 mm	*Weight*	16.07 kg/m
Designed Corrosion rate	0.1 mm/year	*Max Operation Pressure*	1035 PSI

$$\mathrm{Fn} = \frac{1}{2\pi} . 22.4 . \sqrt{\frac{EI}{\mu L^4}} \qquad \text{(Eq.1)}$$

Where:
Fn: natural frequency of the pipe in (Hz)
E: Young's modulus of elasticity (200GPa or 30E6psi for steels)
I: fourth polar moment of inertia (0.049*[OD^4-ID^4]). Where OD and ID is Outer and Inner Diameter.
μ: pipe mass per unit length lbs. /inch or kg/m.
L: the intervals between pipe upholding.

Referring to the information in Table 1 and the Eq.1 [11], the natural frequency of the tested pipe in this experiment is approximately 130 Hz.
In site experiment, the scenario of Event of One Knock per Second of 1m From Node 1. The Scenario have five cases for five different distances. Each case have six set of data that collected to ensure that the proposed system gives same results. Hammering activity done at (1, 6, 12, 18, 24 meter) from the right end of the pipe (Node 1) as shown in figure 3. Hammering event selected from a group of on-ground pipe activities as (drilling, knocking, forklift working, etc.). The proposed system focus on finding the location of event. Each node collect data from a certain period of live vibration readings resulting from hammering activities on different distances, analyze it locally, then send the results of the activity and its location to the supervisor at base-station through WiFi since each node plays as an access point. A damaging event of short periods of continuous knocking using a hand held hammer is applied to the pipe during the tests
The output data rate for the accelerometer is adjusted to 400 Hz to cover the natural frequency of the tested pipe. Normal mode (10 bit) resolution selected with +/- 16 G (gravity) and normal mode for power, disable auto sleep to ensure that a continuous operation. The ADXL345 connected to the Raspberry Pi 3 via I2C protocol. The results of triple-axis (X, Y, and Z) of the scenario illustrated in the next section.

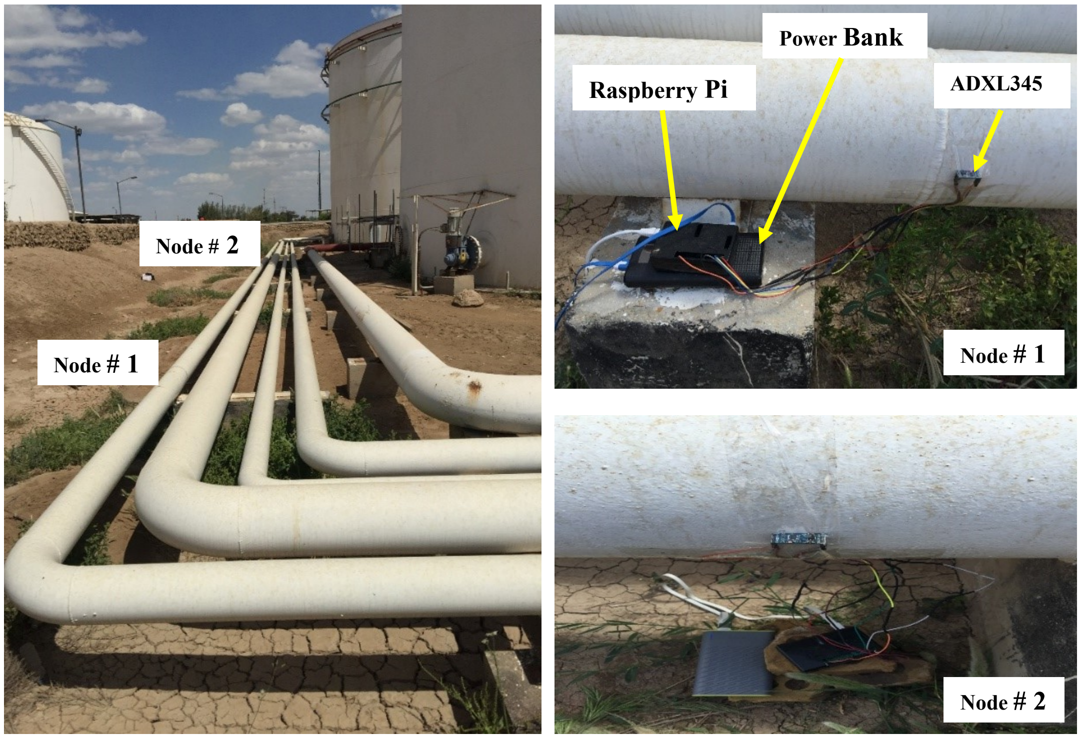

Fig. 2 Accelerometer Layout on the Pipe at Al-Mussaib Gas Turbine Power Plant

4. Feature Extraction, Dimensionality Reduction

The Scenario was implemented by knocking the pipe from the top, vertically, by handheld hammer. Six knocks were applied at a rate of one knock per second in five cases. In each case, different location on the pipeline selected as illustrated in Table 2

Table 2 Knocking Locations of Scenario

Case Number	Distance (*De)*
Case # 1	1 m
Case # 2	6 m
Case # 3	12 m
Case # 4	18 m
Case # 5	24 m

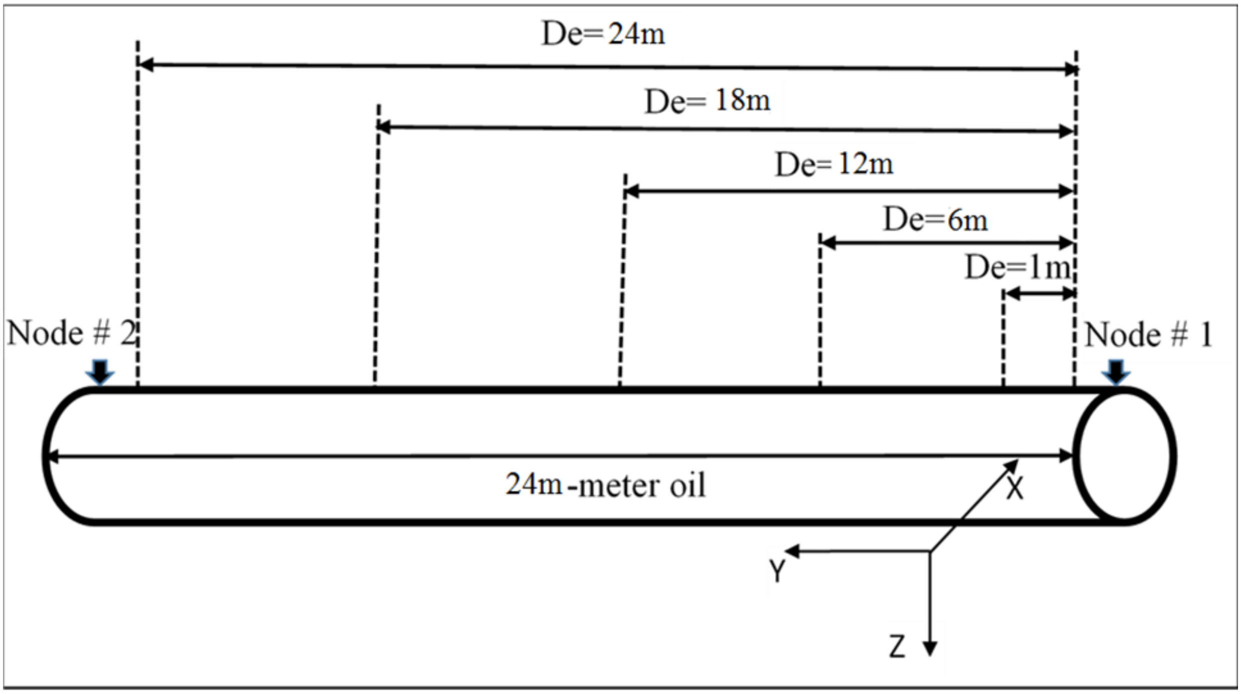

Fig. 3 Distance Selected for Several Distance Event (*De*)

The larger *De* is the larger the attenuation applied to the signal. By decreasing average distance between the WSN nodes, the signal to noise ratio (SNR) increased [12]. In addition, when the distance between two adjacent nodes increased it require more transmission power, which affect the monitoring system privacy. So when design of WSNs, short-range transmission should be considered to reduce eavesdropping and attenuation as well as minimize power consumption. Hence, it is recommended to splitting up large distances between nodes into several shorter distances. Theoretically, the number of nodes attached on the pipe can be increased more than two nodes, but the limited coverage of WiFi could restrict using more nodes.
The steps of data collection, features extraction, features selection, and event classification as shown in Figure 4 - flow chart of In-situ monitoring algorithm.

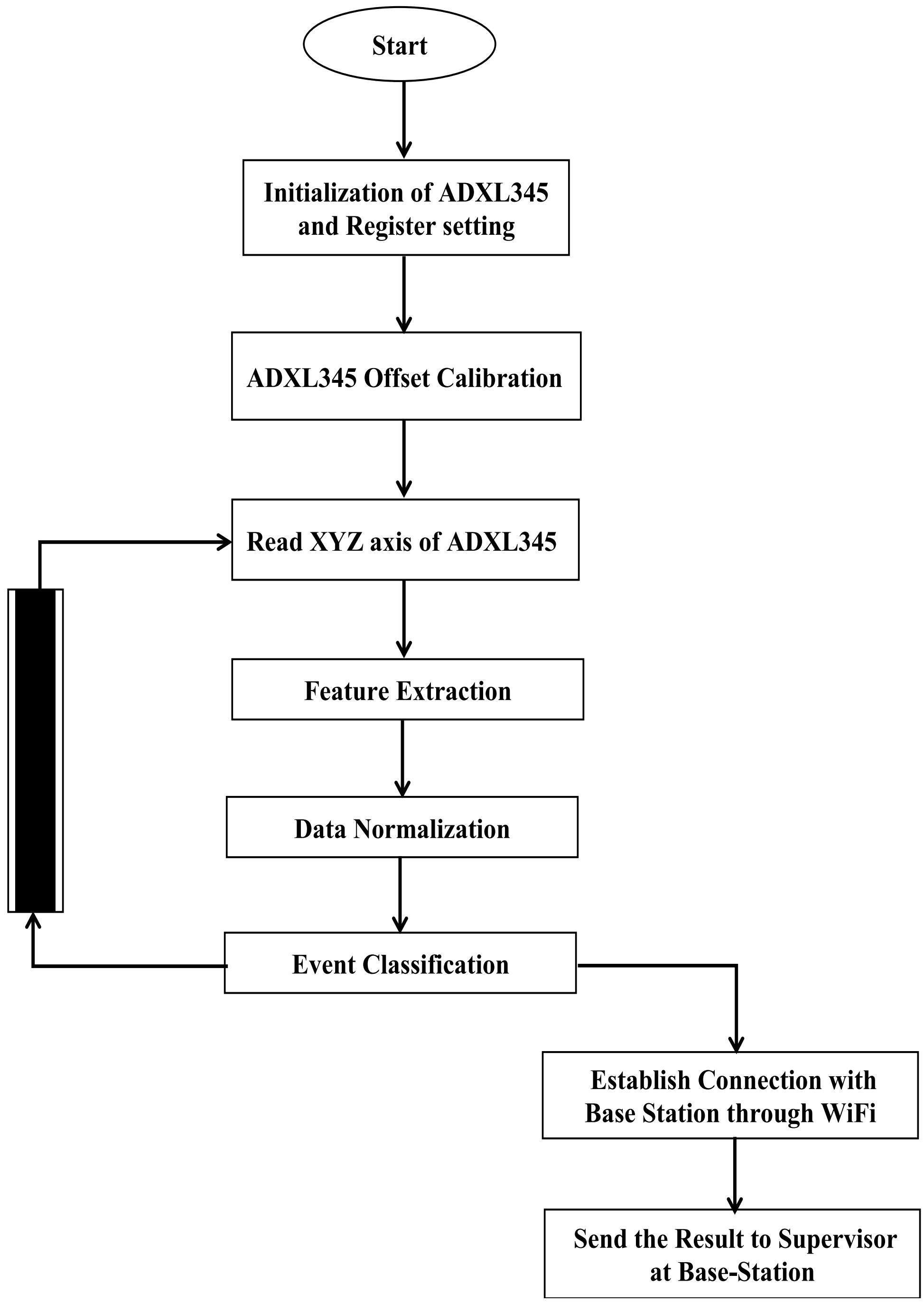

Fig. 4 Flow Chart of In-Situ Monitoring System Algorithm

4.1 Data Collection

Data collected from the accelerometer in digital format. Each readings contain Three values that are the acceleration on the three axis: X, Y, and Z. each measurements. Each value is stored in two bytes with values range between -512 to +511 in 2'complment format. Then it converted to the designated range (either ±2 or ±16). Therefore, each reading contain six bytes. Then the six bytes are read in one burst through I2C protocol as shown in Figure 5, and then stored in one line text or excel file on the Raspberry Pi board as shown in Figure 6.

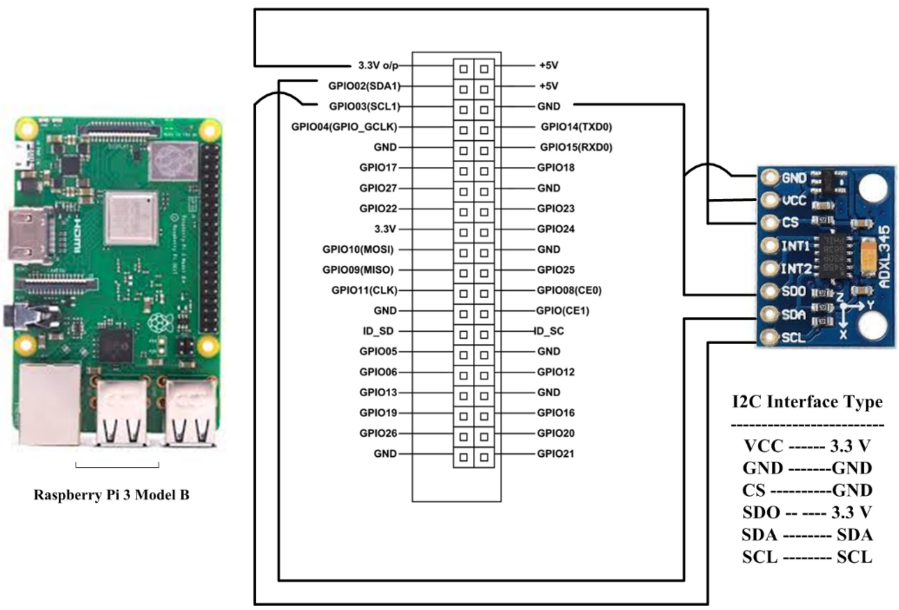

Fig. 5 Accelerometer Interface with Raspberry Pi 3

[redacted]

[redacted]

[redacted]

[redacted]

```
x=22   y=22     z=1       Time=2019-01-08 00:25:20
x=27   y=23     z=1       Time=2019-01-08 00:25:20
x=25   y=21     z=1       Time=2019-01-08 00:25:20
x=-512   y=511     z=1       Time=2019-01-08 00:25:20
x=-236   y=-10     z=1       Time=2019-01-08 00:25:20
x=69   y=10     z=1       Time=2019-01-08 00:25:20
x=-399   y=99     z=1       Time=2019-01-08 00:25:20
x=-71   y=38     z=1       Time=2019-01-08 00:25:20
x=110   y=46     z=1       Time=2019-01-08 00:25:20
x=-17   y=20     z=0       Time=2019-01-08 00:25:20
x=377   y=-61     z=1       Time=2019-01-08 00:25:20
x=58   y=20     z=1       Time=2019-01-08 00:25:20
x=277   y=-106     z=0       Time=2019-01-08 00:25:20
x=21   y=21     z=1       Time=2019-01-08 00:25:20
x=226   y=12     z=1       Time=2019-01-08 00:25:20
x=511   y=408     z=1       Time=2019-01-08 00:25:20
x=31   y=27     z=1       Time=2019-01-08 00:25:20
x=42   y=44     z=1       Time=2019-01-08 00:25:20
x=51   y=17     z=1       Time=2019-01-08 00:25:20
x=47   y=21     z=1       Time=2019-01-08 00:25:20
```

Fig. 6 Sample of Log file that store the accelerometer 3-axis Measurements with the corresponding timestamp

4.2 Feature Extraction and Selection:

Feature extraction step designed to get non-redundant, compact, data meaningful and representation, so it is established by redundant and irrelevant information removal form the gathered data.

The efficiency of any classifier would rely directly on the choice of feature extraction and feature selection method that applied on data. This part assumed that the classifier used smaller and relevant features would achieve higher and better accuracy as well as require less memory. In addition, features extraction also improve the computational speed of the classifier. The selection of classifiers would use in the proposed system depends on the accuracy that calculated by MATLAB – Classifier Learner for each classifier depending on features selected [13]. Figure 7 shows an example of accuracy of many types of classifiers by MATLAB version R2018b.

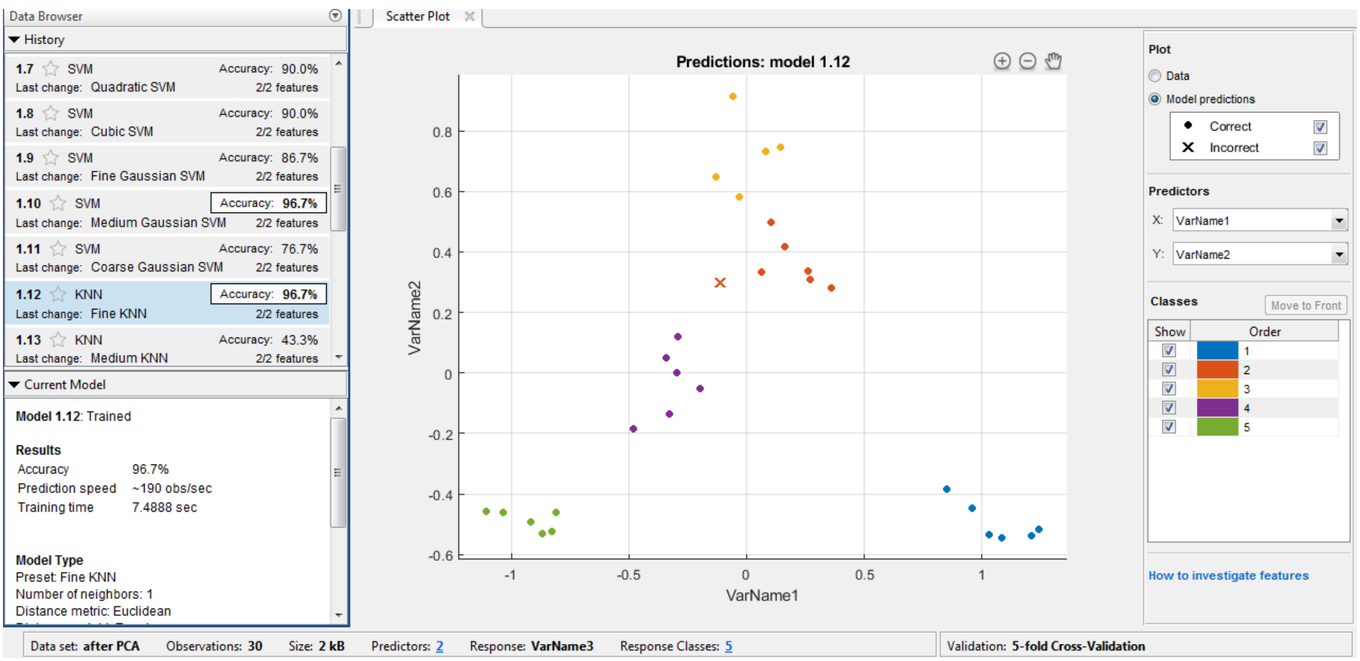

Fig. 7 Classifier Learner by MATLAB Showing the Accuracy of Each One

There are many types of statistics referring to the sample value, that means there is first, second,… Kth order of statistics. In this paper, first order statistics features like (mean, median, mode, standard deviation (STD), Variance, Standard Error, Kurtosis, Skewness, Minimum, and Maximum) calculated for the accelerometer-collected data.

One-way Analysis of Variance (ANOVA) is one of the statistical technique used for data analysis by assessing the discriminatory power of each sample in the features vector and test whether the means of multiple groups are significantly different. It assumes that all samples are normally distributed with equal variance and all samples mutually independent.

The ANOVA implemented by using IBM SPSS Statistics software Version 20, ANOVA Table 3 shows calculations of F-statistic. The table also include P-Value that represent important factor to test the features.

When F-Statistic is HIGH (as high as possible) and P-Value ≈ 0.000, the feature is selected. So according ANOVA table mentioned below, there are 7 numbers selected which related to four features for a good classification and the other features unselected because it have lower values of F-statistic which led to bad classification [14].

Table 3 ANOVA Table for All Statistic Features Examination

Feature Type	F - Statistic	P - Value	Selected
N1 Mean	0.136	0.968	No
N2 Mean	2.034	0.120	No
N1 STD	128.790	0.000	Yes
N2 STD	153.692	0.000	Yes
N1 median	2.055	0.117	No
N2 median	6.369	0.001	No
N1 variance	141.819	0.000	Yes
N2 Variance	248.203	0.000	Yes
N1 STD err	128.790	0.000	Yes
N2 STD err	153.692	0.000	Yes
N1 kurtosis	67.100	0.000	Yes
N2 kurtosis	8.216	0.000	No
N1 skewness	9.974	0.000	No
N2 skewness	1.875	0.146	No
N1 min	0.611	0.658	No
N2 min	9.175	0.000	No
N1 max	1.940	0.135	No
N2 max	9.603	0.000	No
N1 p2p	0.806	0.533	No
N2 p2p	17.174	0.000	No
N1 sum	0.136	0.968	No
N2 sum	2.034	0.120	No

Table 3 shows (11 features examination) of two nodes (one and two), each feature node have F-Statistic and P-Value calculation, the highlighted selected features is chosen when F-Statistic is high as possible and P-Value reaches to zero.

4.3 Dimensionality Reduction by PCA:

The need of Dimensionality Reduction is to avoid using too many predictors in high dimensional data. Using large amount of data (several features) for classification presents many challenges as follow:
a. Memory requirements
b. CPU requirements (time taken for machine learning algorithm)
c. Correlation between predictors
d. Over fitting
e. Some algorithms do not work correctly with too many predictors.
In this matter, special technique called Principle Component Analysis (PCA) has been used in this Paper. This way allow reducing high number of dimensions of a system and keeping all information for the characterization of the various points. PCA is an unsupervised feature selection technique that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables [15]. The new dimensions generated called principal components PCs, it is a linear combination of the original features with different coefficients associated to each original features, and they are orthogonal to each other to maintain most of the variability of the features. Moreover, PCA is widely used for different other purposes such as image compression, orientation detection, face detection, and finding the most relevant variables. Table 4 shows two Principle Components with each pre-defined class that related with event location.

Table 4 Two-Principle Components after PCA Reduction

Principle Component 1	Principle Component 2	Class	*De*
0.959939	-0.444981	1	1 meter
0.85185	-0.384465	1	
1.210965	-0.539363	1	
1.243671	-0.514933	1	
1.084057	-0.543532	1	
1.030201	-0.533262	1	
0.161241	0.416288	2	6 meter
0.261969	0.336211	2	
0.106103	0.497073	2	
0.064782	0.332679	2	
0.358523	0.280374	2	
0.272489	0.307203	2	
-0.057313	0.912546	3	12 meter
-0.029058	0.580808	3	
0.079449	0.733085	3	
-0.128701	0.648878	3	
0.145619	0.744943	3	
-0.113724	0.297377	3	
-0.481167	-0.183657	4	**18 meter**
-0.29162	0.11989	4	
-0.195344	-0.052651	4	
-0.298741	-0.000543	4	
-0.327863	-0.134584	4	
-0.342697	0.051195	4	
-1.107941	-0.457024	5	**24 meter**
-1.036423	-0.459603	5	
-0.868947	-0.531851	5	
-0.827158	-0.523636	5	
-0.808797	-0.461928	5	
-0.915364	-0.492537	5	

5. Events Classification of the Vibrational Signals

The last part covers discussion of classification of the selected features depending on that predefined classes of different event locations. There are many types of classification methods used upon different types of available features, According to these selected features, the classifier is selected depending on the percentage of performance. Confusion Matrix is very useful way to get the performance of any classifier. A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class, this is the key to the confusion matrix Therefore, the confusion matrix includes the following four terms; TP (true positive), FP (false positive), TN (true negative) and FN (false negative) as shown in Table 5 [16].

Table 5 Confusion Matrix

(Rows describe results of predictions for corresponding class, Columns represent actual class)

	Class 1 - Predicted	Class 2 - Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

The most common methods in this manner are supervised classifiers like Linear Discriminant (LDA), K Nearest Neighbor (KNN), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). KNN and SVM were chosen for the proposed system because KNN has the ability to learn from small datasets, easy of interpretation, fast and low calculation time While SVM used to enhance the ability of anti-interference, accurately recognize the state of a pipeline [17] In addition, this classifier is simple in terms of training and testing. The Support Vector Classifier is characterized by fast convergence to the global optimization, excellent generalization capability and immunity to high-dimensional data [18]. So it is selected to compare the results of KNN with it.

5.1 Classification by KNN

The two principle components resulted after PCA reduction for used to find KNN classification percentage and calculate the Confusion Matrix that mentioned in previous part.

Figure 8 shows the plot of KNN classification for the proposed system, this plot a result of executing Python code in Raspberry Pi 3. Each six dots group with the same color represent one event location that classified correctly by KNN classifier, so there are five different colored areas separated by curves for five locations.

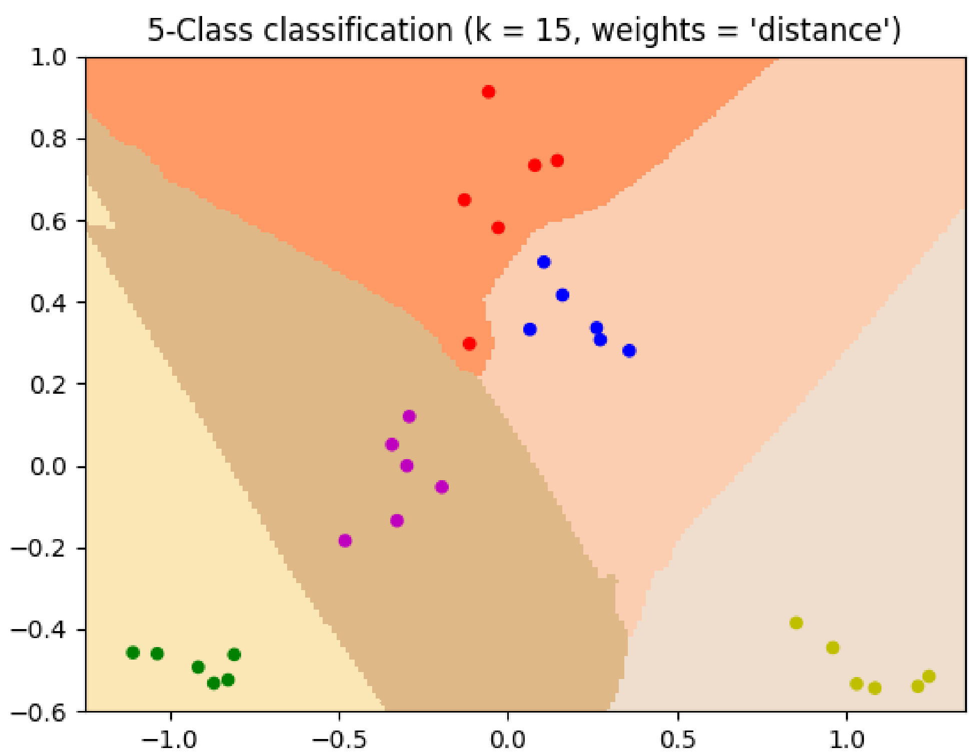

Fig. 8 KNN classification for 30 instance and 5 locations by Python version 3

K in KNN is the number of nearest neighbors used to classify test samples [19]. In this case, we choose K = 15 from the trial way because there is no physical or biological way to determine the best value for “K”, so we have to try out a few values before settling on one [20], small number of K may be selected that led to noisy and missed classification.

It is clear that almost 30 instances were separated correctly and each class have its color to distinguish between them.

Figure 9 shows KNN Confusion Matrix where 5 predicted class versus 5 true class was examined and each class have 6 data group, that shows the percentage of KNN classifier is 100%.

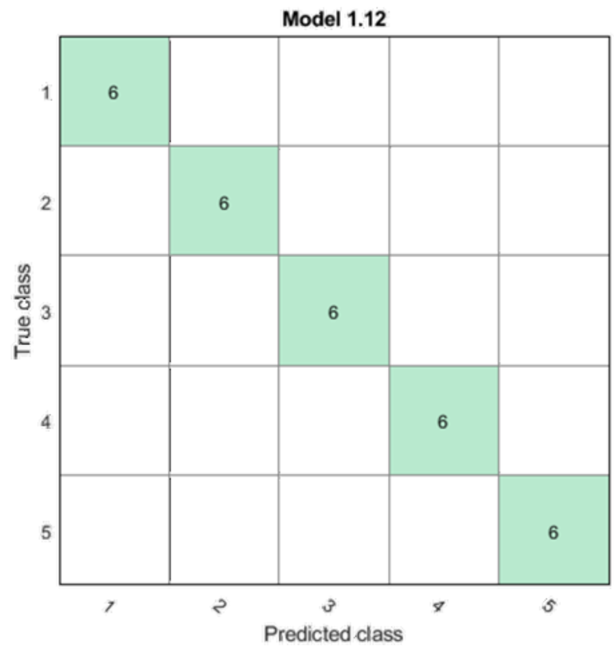

Fig. 9 Confusion Matrix by MATLAB - KNN classification

5.2 Classification by SVM

The results of two Principle Components PCs from PCA reduction used for SVM classification. SVM works by calculate Support Vectors and it depends on these vectors to estimate the decisional area between two or more class.

Figure 10 shows the separation of five classes as the test of the proposed system, The Regulation Parameter denoted by C tells us how margin between classes selected. When C large, the smaller hyperplane is chosen while when C small will cause the optimizer to look for a larger margin separation hyperplane.

Another parameter used in this type of classification that is Kernel Function. There are four types of kernel are: Linear, Polynomial, Radial Basic Function (RBF), and Sigmoid. Kernel represent a dot product of input data points mapped into the higher dimensional feature space by transformation [21]

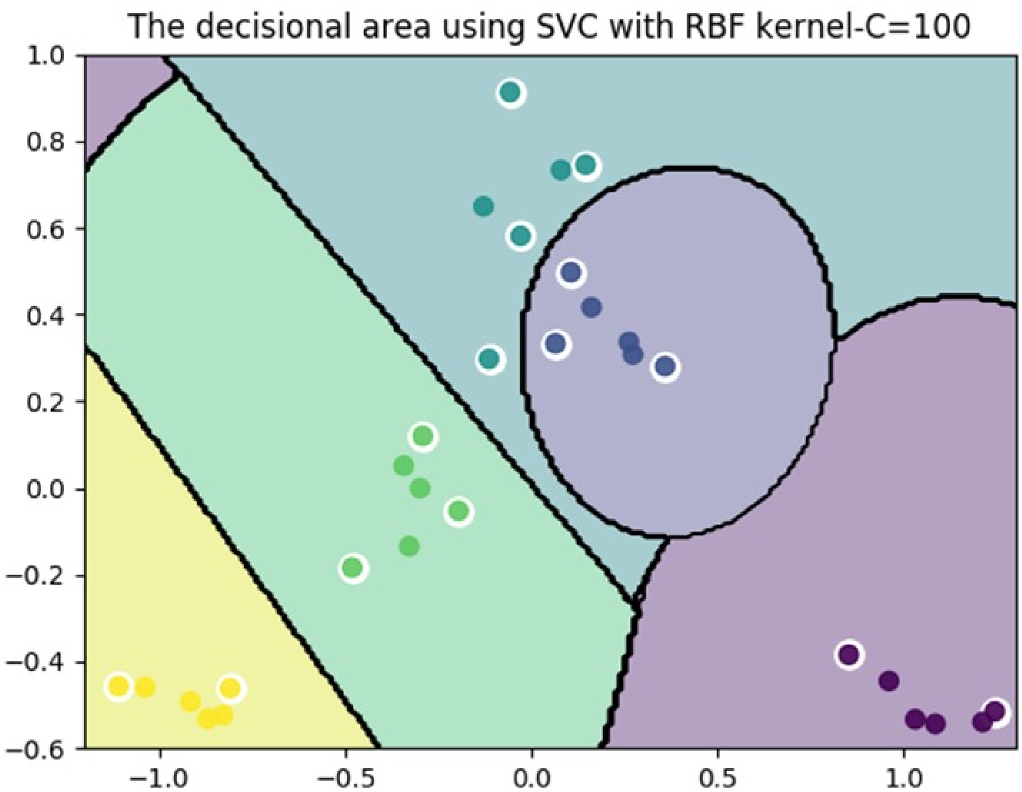

Fig. 10 SVM Classification by Python version 3

Figure 11 shows SVM Confusion Matrix where 5 predicted class versus 5 true class was examined and each class have 6 data group, that shows the percentage of classifier is 96.66 % because 1 instance did not classified correctly from the total instances that are 30 instances. Each six dots group with the same color represent one event location that classified correctly by KNN classifier, so there are five different colored areas separated by lines or curves for five locations.

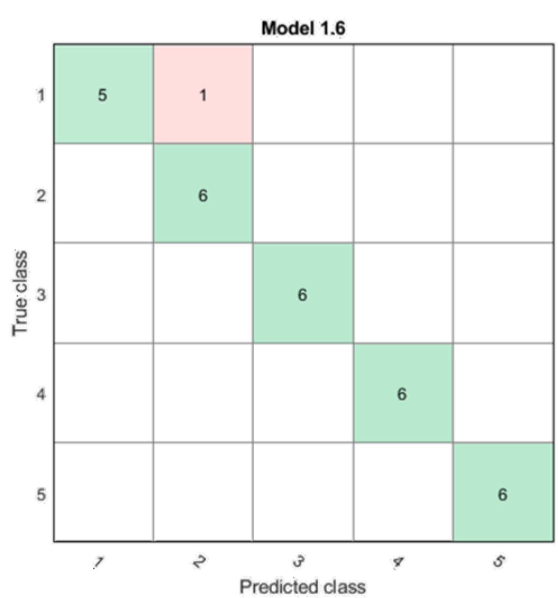

Fig. 11 Confusion Matrix by MATLAB - SVM classification

6. COMPARRATIVE ANALYSIS

Table 6 presents a comparative keynote of selected previously developed systems and the proposed system.

Name	Type of Sense	Sensor name	Protocol & Communication used	Data Analysis
VibNet [5]	Vibration \ event detection	ADXL335 – max acceleration up to +/- 3G	UART / Zigbee	Off-line
Mossa N. [10]	Vibration \ event identification	ADXL345 – max acceleration up to +/- 16G	I^2C / Zigbee	Off-line
Proposed system	Vibration \ event localization	ADXL345 – max acceleration up to +/- 16G	I^2C / WiFi	On-line

7. CONCLUSIONS

An intelligent In-Situ WSN based real time pipeline monitoring system aimed to detect and locate the event location has been presented in this paper. Wireless accelerometer sensor and advanced microprocessor board (Raspberry Pi 3) were used for each node to perform In-Situ processing of the collected data to avoid high bandwidth required for data transmission between nodes and the base station. The board consists of 1.2GHz Quad-Core ARM Cortex-A53 (64Bit) and 3-axis accelerometer. The proposed system is able to collect vibration signal from the attached accelerometer, filter the signal, calculate and select the statistical features, and finally classify it according to the pre-defined class for each event location. By implementing all the processing inside the same node the node save power and transition bandwidth that make the system more reliable for harsh environment. The system is applied to pipeline in Al-Mussaib GT power plant.

The results showed that the system is able to differentiate between several knocks at different locations within 1 to 24 meters. This achieved by processing a set of selected features from the vibration signal using PCA dimensionality reduction and two supervised classification methods (KNN and SVM). Both KNN and SVM classification methods show promising results to reliably classify damaging events of several distances. However, KNN takes relatively short time to classify different events and KNN percentage of classification is greater than percentage of SVM.

The results obtained in this Paper provide useful guidelines on the design of reliable In-Situ Pipeline Monitoring System based on WSN.

REFRENCES

1. Kayaalp, F., et al., *Leakage detection and localization on water transportation pipelines: a multi-label classification approach.* Neural Computing and Applications, 2017. 28(10): p. 2905-2914.
2. Zajam, S., T. Joshi, and B. Bhattacharya, *Application of wavelet analysis and machine learning on vibration data from gas pipelines for structural health monitoring.* Procedia Structural Integrity, 2019. 14: p. 712-719.
3. Journal, P.a.G., *https://pgjonline.com/magazine/2018/September-2018-vol-245-no-9/features/leak-detection-going-to-the-dogs last accessed 28 November, 2019.*
4. Obeid, A.M., et al., *Towards realisation of wireless sensor network-based water pipeline monitoring systems: a comprehensive review of techniques and platforms.* IET science, measurement & technology, 2016. 10(5): p. 420-426.
5. Hossain, N.I., S. Reza, and M. Ali. *VibNet: Application of Wireless Sensor Network for Vibration Monitoring Using ARM*. in *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. 2019. IEEE.
6. Nageswari, S. and T. Saravanan, *Intelligent Garbage Monitoring Method Using LAN Server and Raspberry PI.* framework. 14(15): p. 16.
7. Kumar, D.P., T. Amgoth, and C.S.R. Annavarapu, *Machine learning algorithms for wireless sensor networks: A survey.* Information Fusion, 2019. 49: p. 1-25.
8. Paudyal, S., M.S.A. Atique, and C.X. Yang. *Local Maximum Acceleration Based Rotating Machinery Fault Classification Using KNN*. in *2019 IEEE International Conference on Electro Information Technology (EIT)*. 2019. IEEE.
9. Raktrakulthum, P.-a. and C. Netramai. *Vehicle classification in congested traffic based on 3D point cloud using SVM and KNN*. in *2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE)*. 2017. IEEE.
10. Mossa, N.F., W.F. Shareef, and F.F. Shareef, *Design of Oil Pipeline Monitoring System based on Wireless Sensor Network.* IRAQI JOURNAL OF COMPUTERS, COMMUNICATION AND CONTROL & SYSTEMS ENGINEERING, 2018. 18(2): p. 53-62.
11. Harris, C.M. and A.G. Piersol, *Harris' shock and vibration handbook*. Vol. 5. 2002: McGraw-Hill New York.
12. Hawi, R., *Wireless Sensor Networks--Sensor Node Architecture and Design Challenges.* International Journal of Advanced Research in Computer Science, 2014. 5(1).
13. *https://ww2.mathworks.cn/help/stats/classification-learner-app.html?searchHighlight=app%20learner&s_tid=doc_srchtitle*. 28-11-2019.
14. Chen, D., et al., *Selecting genes by test statistics.* BioMed Research International, 2005. 2005(2): p. 132-138.
15. Verma, C.V. and S. Ghosh. *Dimensionality Reduction Using PCA Algorithm for Improving Accuracy in Prediction of Cardiac Ailments in Diabetic Patients*. in *Proceedings of International Conference on Wireless Communication*. 2020. Springer.
16. Dubitzky, W., M. Granzow, and D.P. Berrar, *Fundamentals of data mining in genomics and proteomics*. 2007: Springer Science & Business Media.
17. Wan, J., et al., *Hierarchical leak detection and localization method in natural gas pipeline monitoring sensor networks.* Sensors, 2012. 12(1): p. 189-214.
18. J Cui, Y.W.-M.a.M.S., 2010 - degruyter.com, *Novel Approach of Analog Fault Classification Using a Support Vector Machines Classifier.* 2010.
19. *http://abhijitannaldas.com/ml/kmeans-vs-knn-in-machine-learning.html*. 2017 6/12/2019.
20. *https://stackoverflow.com/questions/11568897/value-of-k-in-k-nearest-neighbor-algorithm*. 2019 6/12/2019.
21. Han, J., J. Pei, and M. Kamber, *Data mining: concepts and techniques*. 2011: Elsevier.