The effect of linkages in the Hierarchical Clustering of Auto-Regressive algorithm for Defect Identification in Heat Exchanger Tubes

Zakiah Abd Halim, Nordin Jamaludin, Azma Putra

Abstract*: Pattern recognition approach based on Auto-Regressive (AR) algorithm is an alternative way to provide a more accurate defect identification from stress wave propagated along ASTM A179 heat exchanger tubes. The AR algorithm characterizes the shape of the stress wave signals by AR coefficients and clustered using 'centroid' linkages. However, the increase of number of stress waves limiting the function of clustering into meaningful groups. This paper proposes the 'ward' linkages as an improved hierarchical clustering method to define the defect features from the reference tube signals and those from the artificially induced defective tubes. The clustering results from the 'ward' linkages were represented via a dendrogram showing the hidden pattern between clusters. The defect in the heat exchanger tubes are easily interpreted from the dendrogram and can be successfully identified from Maximum Group Distance Criteria (MGDC). The pattern recognition approach using 'ward' linkages in AR algorithm has been shown to effectively identify the defects in the heat exchanger tubes.

***Keywords* : Auto-Regressive, Defect Identification, Hierarchical Clustering, Pattern Recognition**

I. INTRODUCTION

Tubes are a well-known method of transferring fluids between equipment and at the same time function to separate the fluids inside the application such as heat exchanger and boilers. High temperature fluids circulated at both sides of the tubes and make them susceptible to metal degradations such as pitting, corrosion and crack. These discontinuities are major concerned in the industries as they affect the tube integrity as well as raising multiple operation, finance, health and safety issues to the plant operators [1]–[3]. To keep the heat exchanger to perform at its best condition, inspection of heat exchanger tubes is carried out periodically to detect abnormalities in the tubes after being exposed to service.

Conventional methods identify the presence of defect manually from the time series signals. It is found to be cumbersome to plant inspector as the defect identification is dependent on the expertise and proficiencies of the personnel [4]. To have a highly skilled and experienced personnel requires high resources, in terms of training, time and cost. A more accurate data analysis technique leading to automate defect identification for the heat exchanger tubes is therefore of interest.

Pattern recognition provides a wide range and practical overview of the significant pattern in the data set, which enable for a specific trend to be identified and aids the classification of data. The concept of pattern recognition to study the trend of the stress wave signals has been studied in many applications [5]–[11]. Choosing the right classification algorithm is a crucial in ensuring the successful and meaningful hidden pattern in the big data analysis. Previous research has successfully differentiated severity of defects in low speed journal bearings [12], [13]. The research utilized Levinson's forward linear prediction to represent shape of the signals according to AR algorithm.

Recent research also embarks on stress wave propagation pattern in pipelines and tubes [14]–[16]. A new pattern recognition approach has been developed for quick and easy defect identification of defect in ASTM A179 seamless heat exchanger steel tubes. Previous attempt of Vibration Impact Acoustic Emission (VIAE) technique has successfully differentiated the preliminary stress wave signals [14]. The problem is formulated by matching two sets of signals, from a reference tube and a defective tube using a linkage 'centroid'. The stress wave signals were naturally grouped into two main clusters and were shown via dendrogram plot. The Euclidean distance between these two sets of signals is used to recognize the presence of static defect. The dendrogram plot assist visual inspection for defect identification. This AR algorithm method was efficient in separating acoustic emission signals and permitted quick and easy defect recognition in heat exchanger tubes.

The preliminary results shown that AR algorithm is capable to identify the defect in the defective tubes. In previous study [14], only 10 stress wave signals were used in each analysis. However, as the VIAE method becomes more established, more signals are required in the pattern recognition analysis.

Manuscript published on November 30, 2019.
* Correspondence Author

Zakiah Abd Halim*, Fakulti Kejuruteraan Mekanikal, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Melaka, Malaysia. Email: zakiahh@utem.edu.my

Nordin Jamaludin, Fakulti Kejuruteraan dan Seni Bina, Universiti Kebangsaan Malaysia, Selangor, Malaysia. Email: nurdin@ukm.edu.my

Azma Putra, Fakulti Kejuruteraan Mekanikal, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Melaka, Malaysia. Email: azma.putra@utem.edu.my

As the number of the stress wave signals increasing, the linkage between each signals becomes non-monotonous. The clusters between each type of tube conditions disappear and it is difficult to identify the Maximum Group Distance Criteria (MGDC) between clusters.

In this paper, an improvised Auto Regressive (AR) algorithm is presented as the pattern recognition algorithm due the ability of the approach to classify unique pattern based on the shape of the signals. This paper focuses on the linkage 'ward' in hierarchal clustering which enables the stress wave signals to be clustered into the correct cluster and provides significant MGDC which is important for defect decision.

II. STRESS WAVE GENERATION

A. Heat Exchanger Tubes specifications

The experiment involved four defective heat exchanger tubes and a tube without defect used as the reference tube. The heat exchanger tubes were ASTM A179 seamless cold-drawn steels having dimensions of 19.05 mm outer diameter, 2.11 mm tube thickness, and 1 m tube length. Each defective tube has one artificially induced through-hole defect. The locations of the defect, Y relative to the sensor were prepared based on details in Table I. The schematic diagram is shown in Fig. 1.

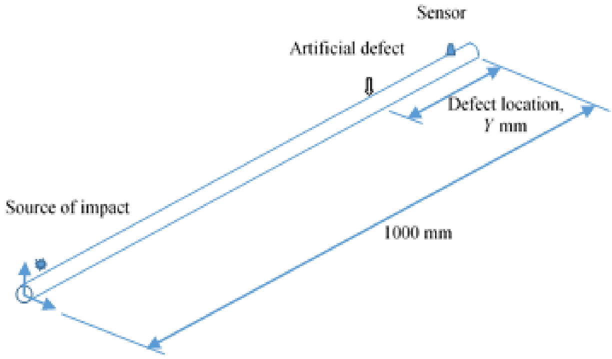


Fig. 1. Schematic diagram of heat exchanger tube showing the location of the defect relative to the sensor location.

Table- I: Details of defect location in the tested tubes

Type of the tube	Label of the signals	Defect location relative to the sensor, Y (mm)
Reference tube	R1-R30	-
Tube 1	A1-A30	100
Tube 2	B1-B30	200
Tube 3	C1-C30	300
Tube 4	D1-D30	400

B. Instrumentations

A simple and feasible approach called Vibration Impact Acoustic Emission (VIAE) is used to acquire the stress wave signals. The stress wave in the heat exchanger tube was excited externally by using an instrumented impact hammer at one end of the tube. The impacted load was recorded at 37 N. The stress wave generated from the impact were captured by using an acoustic emission (AE) sensor located at the opposite of the tube end. Fig. 2 shows the schematic of the investigation set up. The location of the impact excitation in the testing tube are shown in Fig. 3.

For each tube listed in Table I, the data were collected for 30 times impact excitations. This number of impacts is the minimum requirement for analysis of pattern recognition. The stress wave propagations were recorded in time domain. The acquired stress wave signals have sample length of 524,288 µs and sampled at 10 MHz. For convenience of analysis, each stress wave signals for the reference tube was labelled as R1 to R30, whereas stress wave signals from the defective Tube 1, Tube 2, Tube 3 and Tube 4 were labelled as A1-A30, B1-B30, C1-C30 and D1-D30, respectively. See again Table I.

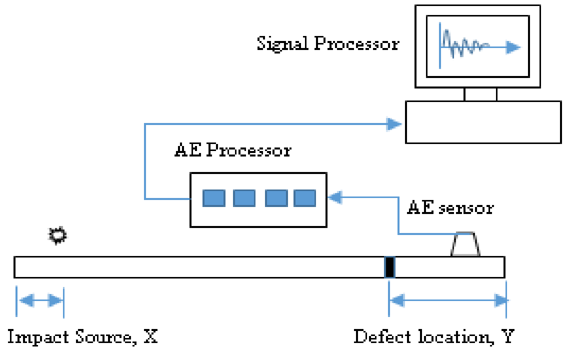


Fig. 2. Schematic diagram of the experimental setup for VIAE.

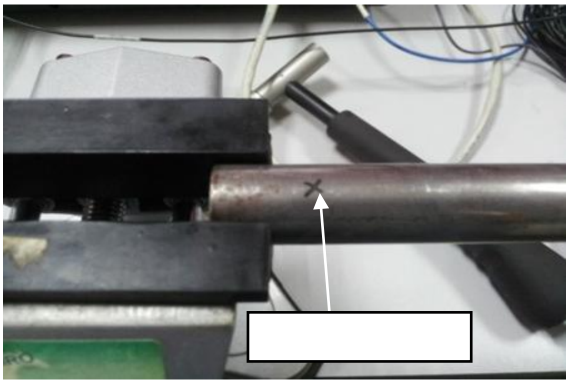


Fig. 3. Location of impact excitation using an impact hammer.

III. PATTERN RECOGNITION

The data were analyzed both in time series and pattern recognition analysis. The stress wave signals were classified using AR algorithm. There are 5 basic steps involved in the pattern recognition analysis using AR algorithm. Every stress wave signals were extracted at 0.01s, which resulted only 100 000 point data will be represented in unit vector. The size of the matrix [AEsignals] in this study is X × 100 000, where X is the number of signals involved in the pattern recognition analysis.

The AR algorithm describes discrete-time stochastic process, x(n) with a definite corresponding forward prediction error (FPE) given by;

$$x(n) = \sum_{k=1}^{M} h(k)x(n-k) + W(n) \qquad (1)$$

in which W(n) is a white noise signal, h(k) are the AR coefficients and M is the model order. The W(n) has zero mean and variance σ2 serves as the input in the AR algorithm. At sufficiently high M, a set of AR coefficients signify the signal's shape. Hence, the correct model order for any AR algorithm is of a significant decision. The Levinson-Durbin algorithm was used to calculate M due to its efficiency in solving recursive solution. The value of M is chosen when the σ^2 has converged to a constant value.

The AR coefficients were clustered using unsupervised hierarchical clustering. Clustering was formed by allocating all set of AR coefficients in a single cluster. Euclidean distance was selected to measure the similarity between each AR coefficient. At each of the iteration, the Euclidean distance compares a pair of the signals. Two signals that has the least Euclidean distance was clustered in the same cluster. When a new signal was added into the group, a new distance was calculated, and a new group was formed. The most heterogeneous signal was grouped into another cluster. This steps were repeated until all the signals were added into the pattern recognition analysis.

This research presents the linkage 'ward' as defect identification feature. Linkage 'ward' calculates the AR coefficients based on the minimum increase of sum of squares between the signals. The algorithm was calculated using MATLAB software. The results of the algorithm were presented via dendrogram, which is a proactive tool capable of extracting the relevant hidden pattern from the measured signals. In the dendrogram, signals with similar characteristics are clustered into similar cluster, whereas signals with different characteristics are group in different cluster. Finally, the identification of defect can be analysed according to the label of signals in each main cluster in the dendrogram. The MGDC value between the main clusters is recorded as the representation of degree of dissimilarity between the patterns of the measured signals. The flowchart of the pattern recognition method was illustrated in Fig. 4.

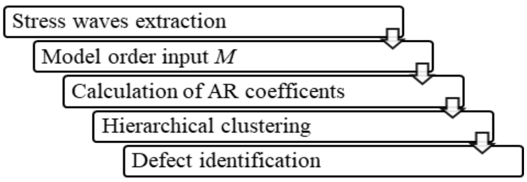


Fig. 4. The flowchart of the pattern recognition method using Auto-Regressive (AR) algorithm.

IV. RESULT AND DISCUSSION

A. Time series

Fig. 5 shows the typical of stress wave signal captured for reference tube and defectives tubes when impacted with external impact load. The stress wave shows a burst type signal, i.e. a typical resemblance of a Hsu-Nielson signal. It can be observed that each tube condition produces different stress wave pattern. Each stress wave signals contain multiple, repetitive series of burst signal. The series of burst signals are due to the type of stress wave propagated in the tube structure, which consist of surface wave and longitudinal stress wave. These two type of stress waves are generated simultaneously as the impact force is exerted to the tube structure.

However, due to location of the AE sensor at the end of opposite impact location, the surface waves were captured earlier than the longitudinal waves even though the longitudinal waves have higher velocity compared to the surface waves. The longitudinal waves have longer travel path in the tube structure compared to surface waves as illustrated in Fig. 6. The tube thickness becomes the limiting factor that guide the longitudinal wave (in x-direction) to propagate along the tube structure (in y-direction). The signals are reflected at either ends of the tube.

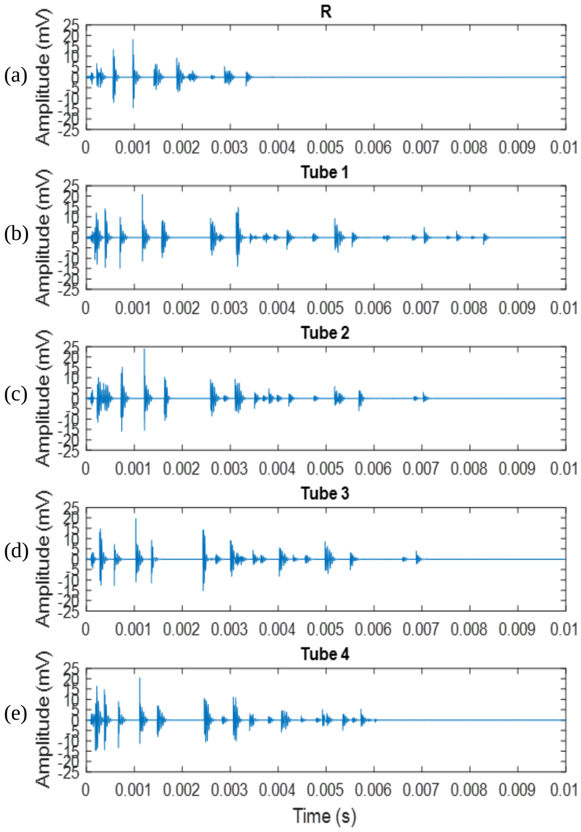
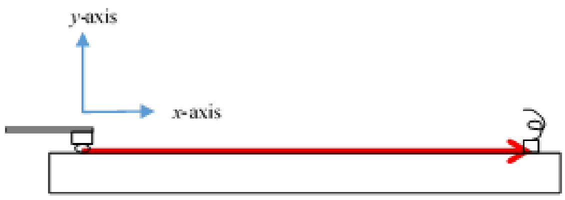
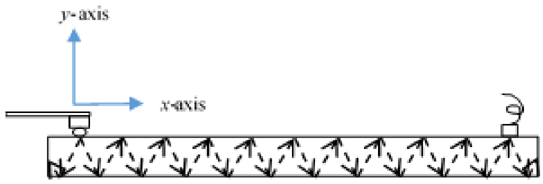


Fig. 5. Examples of the stress waves measured from the VIAE method for: (a) Reference tube; (b) Tube 1; Y=100 mm; (c) Tube 2; Y=200 mm; (d) Tube 3; Y=300 mm; and (e) Tube 4; Y=400 mm.



(a)



(b)

Fig. 6. Illustration of stress wave propagated in the tube structure: (a) Surface waves; (b) Longitudinal waves.

The location of the defect also affects the unique pattern of stress propagation. More repetitive small burst signals can be captured when the defect is located close to the AE sensor as seen in Fig. 5(b) for Tube 1.

The small burst signals extend the travel path of the stress wave propagation in the tube.

The stress wave propagated in the reference tube was recorded at 3777 μs (this is shown in Fig. 5(a) for the last burst signal in the graph). Fig. 5(b) to Fig. 5(d) also show that the duration of the stress wave signals decreases as the defect location becomes further away from the sensor. The average duration of the stress wave propagated in each tubes were recorded in Table II.

Table- II: Duration of the stress wave propagation in the tested tubes

Type of the tube	Defect location relative to the sensor, Y (mm)	Duration, t (μs)
Reference tube	-	3777
Tube 1	100	8761
Tube 2	200	7422
Tube 3	300	7293
Tube 4	400	6184

However, it is difficult to differentiate the stress wave propagation pattern visually especially as the number of the small burst signals vary for each impact. Hence it is desirable to have better defect identification method which provides more accurate identification mathematically. The next section discusses the proposed better linkages between the stress waves based on the features of the measured signals.

B. Pattern recognition

In the pattern recognition analysis, all the 30 stress wave signals measured in the experiment for each defective tube are compared to the 30 stress wave signals for the reference tube. The FPE analysis (Eq. (1)) indicates that a model order of M = 40 is sufficient to represent the shape of the stress wave signals propagated in heat exchanger tubes.

Before proceeding to the results using the linkage ‘ward’, the results from the implementation of linkage 'centroid' are first discussed here, which have been presented in the previous works [14], [17] to be successful for less amount of measured data. Fig. 7 presents the dendrogram plots of pattern recognition analysis using AR algorithm and linkage 'centroid' for each pair of the reference and defective tubes.

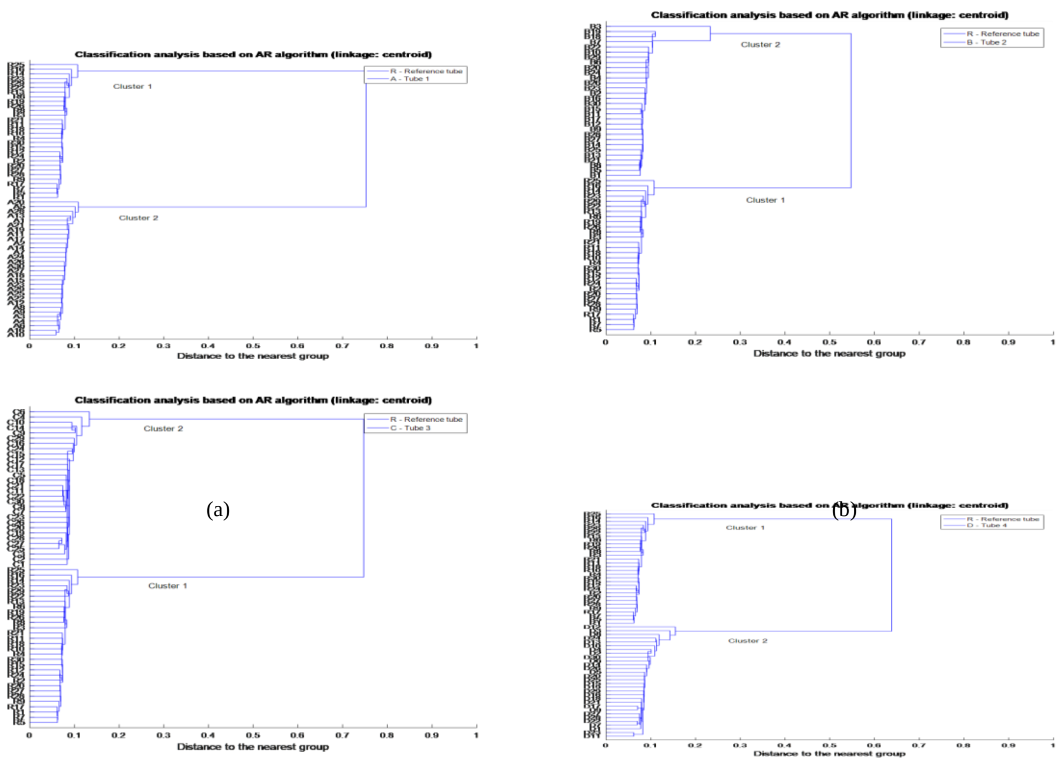


Fig. 7. Dendrogram plot from linkage 'centroid' for: (a) Reference tube-Tube 1; (b) Reference tube-Tube 2; (c) Reference tube-Tube 3; (d) Reference tube-Tube 4.

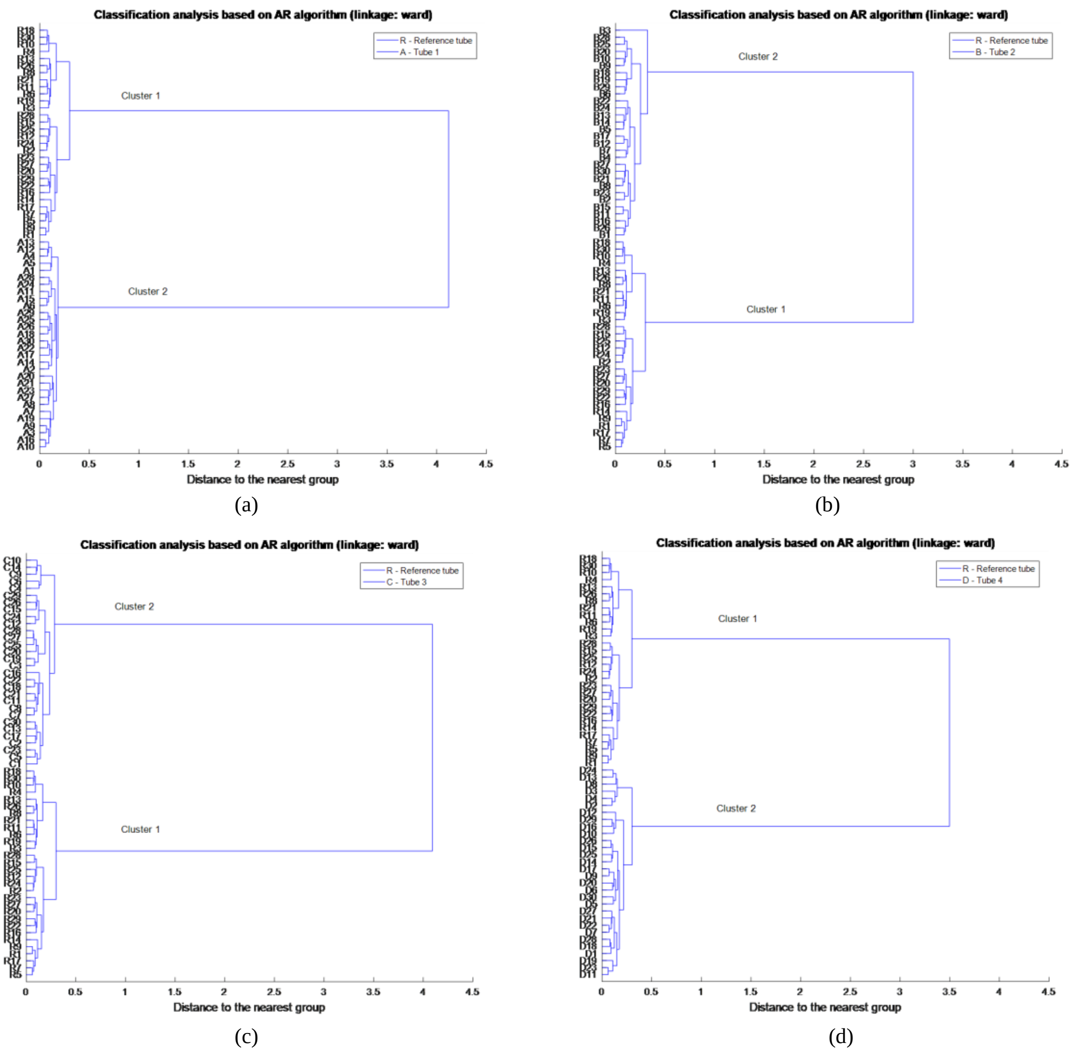


Fig. 8. Dendrogram plot from linkage 'ward' for: (a) Reference tube-Tube 1; (b) Reference tube-Tube 2; (c) Reference tube-Tube 3; (d) Reference tube-Tube 4.

The x-axis denotes the Euclidean distance between each signal value while the y-axis indicates the stress wave signals represented by the corresponding labels (see Table II). The linkage 'centroid' is often used due its convenient computation to represent the center of the shape predicted using AR coefficients [17], [18].

The dendrogram illustrated in Fig. 7 shows that the stress wave signals were differentiated into two main groups. Cluster 1 contains 30 stress wave signals propagated in the reference tube, whereas Cluster 2 contains 30 stress wave signals propagated in either defective tube. Signals in Cluster 1 indicates the baseline signals in this analysis.

It is observed that from the linkage 'centroid' method, the inversion between each signal condition which produces non-monotonic chaining is apparent within both clusters. The linkage 'centroid' tends to attach signal one by one to the cluster thus demonstrates the relatively small growth of curve.

The non-monotonic chaining is a reflection that the shapes of the stress wave signals have almost similar statistical patterns between one another, hence it is difficult to differentiate the 'centroid' of the stress wave signals and this affects the MGDC values in the dendrogram. The results indicate as if no defects are presence in the heat exchanger tubes.

These results are totally different from the previous study in [19] as the number of stress wave signals involved in the pattern recognition analysis in the previous work was only 10 signals. Recent study suggested that minimum number of signals to represents the pattern of a training set is 30 signals in any pattern recognition analysis [20]. It is found here that linkage 'centroid' is only suitable for a small number of signals. As the number of signals used in the pattern recognition analysis increases, the linkage between signals becomes chaining with each other and produces no apparent cluster representing the pattern of stress wave propagation in the heat exchanger tubes. Hence, the linkage 'centroid' between each signal is considered as a weak statistical representation of the signals.

Fig. 8 depicts the dendrogram plots of pattern recognition analysis using AR algorithm and linkage 'ward' for each pair of the reference and defective tubes. In contrary to non-monotonic chaining signals as in Fig. 7 from the linkage 'centroid', the linkage 'ward' produces monotonous linkages between each signal within the cluster.

There are also few mini clusters in the Cluster 1 and Cluster 2 due to small proximity of signals between the two clusters. Linkage 'ward' calculates the Euclidean distance between two signals using the error of sum squared. The pair that has smaller summed squared is grouped together in a cluster. When the summed square of the new pair signal is larger, a new cluster is formed. The size of clusters increase until all the signals are linked together in the hierarchical tree.

As the linkage 'ward' involves an increase of sum of squared error between two signals at the beginning, hence the distance between two signals becomes dilating from very first iterations. The significant jump between Cluster 1 and Cluster 2 in Fig. 8 indicates well separated clusters. The separation between clusters is measured from the dendrogram plot. The MGDC value is the maximum distance between the two main clusters. The MGDC value presents how significant the dissimilarity of the signals from each other. The degree of dissimilarity indicates that the shape of stress waves in the defective tube as computed by AR algorithm is not similar with the shape of the stress waves in the reference tube. The MGDC value between the two main clusters are tabulated in Table III. The MGDC values from dendrogram plots in Fig. 7 ('centroid') and Fig. 8 ('ward') vary approximately from 0.50 to 0.75 and 3.0 to 4.2, respectively.

Table- III: Duration of the stress wave propagation in the tested tubes

Case	Label of the signals	MGDC	
		'centroid'	'ward'
Reference tube-Tube 1	R-A	0.7522	4.120
Reference tube-Tube 2	R-B	0.5477	4.101
Reference tube-Tube 3	R-C	0.7472	4.093
Reference tube-Tube 4	R-D	0.6386	3.498

From Table III, the MDGC value of reference tube-Tube 1 has the highest distance which is 4.120 for linkage 'ward'. It means that the signals in Tube 1 have the most significant different characteristics from those of the reference tube compared to the other defective tubes. MGDC values more than 2 indicate that defect is presence in the heat exchanger tube. It can also be observed from Table III that the MGDC values reduce as the defect location becomes further away from the AE sensor. The sensor–source distance determines the duration for stress wave propagation, as well as the shape of the signals.

Currently, there is no standard benchmarking in setting the similarity index between the stress waves signals, but previous research suggested that MDGC < 0.3 between clusters indicates that the stress waves signals have similar characteristics or similar pattern, whereas MDGC > 2 indicates significant dissimilarity of the signal patterns between clusters [21].

V. CONCLUSION

This study presents the effects of linkages in hierarchical clustering in AR algorithm for defect identification in heat exchanger tubes. The AR algorithm successfully allocates the correct signals into their natural groups. The dendrogram plot illustrates the capability of AR algorithm in discriminating the type of stress wave signals propagating in the tube structure. It is found that the linkage 'ward' distinctly showed the clustering of defect signals from the reference signals. The results indicate that linkage 'ward' can be implemented for modelling large number of VIAE signals. It successfully increases the homogeneity of each signals in the main cluster and increases the MGDC values between the clusters. These results demonstrate that pattern recognition approach for defect detection is feasible for fast and potentially useful to identify the presence of defects in heat exchanger tubes and other similar structure. This study can be extended for more complex defects and to investigate the robustness of the method for defect identification.

ACKNOWLEDGMENT

The heat exchanger tubes were sponsored by TENAGA TIUB SDN. BHD. Zakiah Abd Halim would like to thank Universiti Teknikal Malaysia Melaka (UTeM) and the Ministry of Higher Education, Malaysia for her PhD sponsorship and project expansion through short term grant PJP/2019/FKM(8B)/S01680.

REFERENCES

1. J. H. Bulloch, A. G. Callagy, S. Scully, and A. Greene, “A failure analysis and remnant life assessment of boiler evaporator tubes in two 250MW boilers,” *Eng. Fail. Anal*., vol. 6, 2009, pp. 775–793.
2. S. Cicero, R. Lacalle, R. Cicero, and J. García, “Failure analysis of a steam generator superheater drain tube used in a dump,” *Eng. Fail. Anal*., vol. 17, 2010, pp. 301–312.
3. S.R. Allahkaram, P. Zakersafaee, and S. A. M. Haghgoo, “Failure analysis of heat exchanger tubes of four gas coolers,” *Eng. Fail. Anal*., vol. 18, 2011, pp. 108–1114.
4. N. J. Carino, “Training: Often the missing link in using NDT methods,” *Constr. Build Mater*., vol. 38, 2013, pp. 1316–1329.
5. M. Arredondo, D. A. Tibaduiza, M. McGugan, H. Toftegaard, K. K. Borum, L. E. Mujica, J. Rodellar, and C. P. Fritzen, “Multivariate data-driven modelling and pattern recognition for damage detection and identification for acoustic emission and acousto-ultrasonics,” *Smart Mater. Struct*., vol. 22. 2013, pp. 105023.
6. X. Wang, and V. Makis, “Autoregressive model-based gear shaft fault diagnosis using the Kolmogorov–Smirnov test,” *J. Sound Vib*., vol. 327, 2009, pp. 413–423.
7. R. Yao, and S. N. Pakzad, “Autoregressive statistical pattern recognition algorithms for damage detection in civil structures,” *Mech. Syst. Signal Process*., vol. 31, 2012, pp. 355–368.
8. L. P. Dickinson, and N. H. Fletcher, “Acoustic detection of invisible damage in aircraft composite panels,” *Appl. Acoust*., vol. 70, 2009, pp. 110–119.
9. J. L. Ferrando, V. Kappatos, W. Balachandran, and T. H. Gan, “A novel approach for incipient defect detection in rolling bearings using acoustic emission technique,” *Appl. Acoust*., vol. 89, 2015 pp. 88–100.
10. A. Glowacz, “Acoustic based fault diagnosis of three-phase induction motor,” *Appl. Acoust*., vol. 137, 2018, pp. 82–89.
11. K. Yu, T. R. Lin, and J. W. Tan, “A bearing fault diagnosis technique based on singular values of EEMD spatial condition matrix and Gath-Geva clustering,” *Appl. Acoust*., vol. 121, 2017, pp. 33–45.
12. N. Jamaludin, and D. Mba, “Monitoring extremely slow rolling element bearings: Part I,” *NDT E Int*., vol. 35, 2002, pp. 349–358.
13. N. Jamaludin, and D. Mba, “Monitoring extremely slow rolling element bearings: Part I,” *NDT E Int*., vol. 35, 2002, pp. 359–366.
14. Z. A. Halim, N. Jamaludin, S. Junaidi, and S. Y. S. Yahya, “Vibration impact acoustic emission technique for identification and analysis of defects in carbon steel tubes: Part B Cluster analysis,” *J. Mech. Sci. Technol*., vol. 29, 2015, pp. 1559–1570.

15. S. Lalot, and H. Pálsson, "Detection of fouling in a cross-flow heat exchanger using a neural network based technique," *Int. J. Therm. Sci*., vol. 49, 2010, pp. 675–679.
16. S. Li, Y. Song, G. Zhou, "Leak detection of water distribution pipeline subject to failure of socket joint based on acoustic emission and pattern recognition," *Meas. J. Int. Meas. Confed*., vol. 115, 2018 pp. 39–44.
17. Wang, L. M., and Shao, Y. M., "Crack fault classification for planery gearbox based on feature selection technique and K-means clustering method," *Chinese Journal of Mechanical Engineering*, vol. 31, 2018, pp. 1-11.
18. Zhou, Q., Shen, H. Zho, J., Liu, X., and Xiong, X., "Degradation state recognition of rolling bearing based on K-means and CNN Algorithm," *Shock and Vibration*, 2019. https://doi.org/10.1155/2019/8471732
19. Z. A. Halim, N. Jamaludin, S. Junaidi, and S. Y. S. Yahya, "Detection of tube defect using the autoregressive algorithm," *Steel Compos. Struct*., vol. 1, 2015, pp. 131–152.
20. L. Yu, S. Z. Li, "Acoustic emission (AE) based small leak detection of galvanized steel pipe due to loosening of screw thread connection," *Appl. Acoust*., vol. 120, 2017, pp. 85–89.
21. N. Jamaludin, "Monitoring low speed rolling element bearing using stress waves technique," PhD Thesis, Cranfield University, 2000, unpublished.

AUTHORS PROFILE

Zakiah Abd Halim received her Ph.D from Universiti Kebangsaan Malaysia (UKM) in 2018. She is currently a senior lecturer in Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka (UTeM). Her current research is in the area of structural integrity, non-destructive testing method such as acoustic emission and evaluation of defect behavior using signal processing analysis. Dr. Zakiah is a member of Malaysian Society of Non Destructive Testing (MSNT) and has Malaysian Skills Certificate as Ultrasonic Tester Level 1 and Level 2.

Nordin Jamaludin earned his PhD. degree in 2000 in the area of condition monitoring at Cranfield University. He continued his career as a lecturer in Universiti Kebangsaan Malaysia (UKM) until promoted to his current position as Associate Professor. He has more than 13 years of experience in the condition monitoring field. Area of research are investigations on material damage mechanism in structural and machinery components using acoustic emission and metal magnetic memory technology as part of integrity monitoring analysis and assessment. Assoc. Prof. Dr. Jamaludin is a fellow in Welding Institue of Malaysia, Chairman of Malaysian Acoustic Emission Working Group as well as member of Malaysian Society of Non Destructive Testing (MSNT).

Azma Putra received his Ph.D in 2008 from the Institute of Sound and Vibration Research (ISVR), University of Southampton, United Kingdom. He is now an Associate Professor in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. His current interests are engineering acoustics and noise control, vibro-acoustics and structural dynamics. Dr. Putra is a member of Institute of Noise Control Engineering (INCE), USA, International Institute of Acoustics and Vibration (IIAV), USA and the Society of Vibration and Acoustics Malaysia (SVAM).