

클러스터링 알고리즘을 이용한 배관의 부식 상태 분류

천강민*, 신건호*, 허장욱*,#

*금오공과대학교 기계공학과(항공기계전자융합공학전공)

State Classification of the Corrosion of Pipes Using a Clustering Algorithm

Kang-Min Cheon*, Geon-Ho Shin*, Jang-Wook Hur*#

(Received 13 April 2022; received in revised form 06 June 2022; accepted 12 June 2022)

*Department of Mechanical Engineering (Department of Aeronautics, Mechanical and Electronic Convergence Engineering of Mechanical Engineering), Kumoh National Institute of Technology

(Received 13 April 2022; received in revised form 07 May 2022; accepted 12 June 2022)

ABSTRACT

Pipes transport and supply fuel in various categories; however, corrosion occurs because of the external environment, impurities are mixed in the fuel, and substances leak to the outside, which can lead to serious accidents. Therefore, in this study, inspection equipment using a laser scanner was manufactured to classify conditions according to the degree of corrosion of the outer wall of the pipe, and the corrosion height and maximum value of the pipe were obtained from the surface information. Using the k-means method, it was classified into four states, and the standard of the average height and maximum height of corrosion for each state was derived.

Keywords : State Classification(상태 분류), Clustering Algorithm(클러스터링 알고리즘), Laser Displacement Sensor(레이저 변위 센서), Corrosion Height(부식 높이), K-Means(K평균)

1. 서론

배관은 건물 외벽, 해양 및 지하 등에 설치되어 에너지를 필요로 하는 작은 산업용 기계에서부터 발전소까지 물이나 연료 등을 공급 및 이송하는 역할을 수행한다^[1,2]. 그러나 외부 환경에 장기간 노출되어 배관의 내면, 외면에 발생한 부식으로 인해

내용물에 불순물이 섞여 장비에 공급될 수 있으며, 배관 내 물질이 외부에 유출되어 큰 사고로 이어질 수 있다^[3].

이로 인해 발전소와 같이 위험한 작업 현장에서 배관의 내부 상태를 판단하기 위해 적외선열화상 카메라, 자기누설탐상, 초음파탐상 등 비파괴검사의 요구가 증대되고 있으나, 배관의 외부에 대한 검사는 내부에 비해 활발히 이루어지지 않는 실정이다^[4-6]. 비파괴 검사 장비 중 레이저 변위센서는 배관과 비접촉 측정으로 표면 정보를 획득할 수

Corresponding Author : hhjw88@kumoh.ac.kr

Tel: +82-54-478-7399, Fax: +82-54-478-7399

있으며, 작업자의 현장 투입없이 원격으로 검사할 수 있는 장점이 있으나, 레이저 변위 센서를 사용하여 부식 높이 또는 크랙을 탐지하는 연구는 이루어지지 않고 있다. 또한, 배관 부식 진단 측정 방법 중 육안관찰을 통해 표면 진단시 간단하게 부식 상태를 판정할 수 있지만, 작업자마다 주관적이고 정성적이라는 적용한계가 있다. 그러나, 레이저 변위 센서는 육안으로도 식별하기 어려운 배관의 표면을 검사함으로써 주관적이라는 한계에 영향을 받지 않는다^[7].

따라서 본 연구에서는 육안으로 식별하기 어려운 배관의 외벽 표면 정보를 획득하기 위해 레이저 변위 센서를 이용하여 제작한 측정 장비로 배관의 부식 높이를 측정하였으며, 상태 분류를 위한 파라미터는 부식의 높이로 선정하였다^[1]. 또한, 배관의 표면 데이터를 이용하여 배관의 상태를 분류하는 인공지능 방법론 적용을 위해 클러스터링 알고리즘 기법 중 최적의 상태 분류 알고리즘을 제시하였다.

2. 부식 높이 측정

2.1 측정센서 및 장비

배관의 부식 높이 측정을 위해 이용한 레이저 변위 센서는 대상 물체와 물리적인 접촉이 없이 표면의 정보를 획득할 수 있는 장치이며, Riftek사의 RF627을 사용하였다. 센서는 레이저를 방사하여 연속적으로 라인을 스캔하여 2차원 평면 데이터를 획득하며, 측정 정밀도는 1 μ m의 매우 정밀한 장치로 센서의 형상과 측정 범위를 Fig. 1에 나타내었다^[8].

측정장비는 측정센서의 높은 정밀도로 인해 센서의 진동을 최소화하기 위해 볼 스크류와 장비의 양쪽에 리니어 가이드를 부착하였으며, 배관 측정 위치에 롤러를 설치함으로써 회전을 통해 다른 면을 측정이 가능토록 Fig. 2와 같이 설계하여 제작하였다. 전원공급 및 제어장치로 측정거리와 이동속도를 조절하며, 실시간 측정되는 형상을 확인하고, 데이터를 저장하기 위해 컴퓨터와 연결하였다.

배관의 측정 조건으로 폭 방향은 21mm(센서의 측정범위), 축 방향은 150mm로 데이터를 획득하였다.

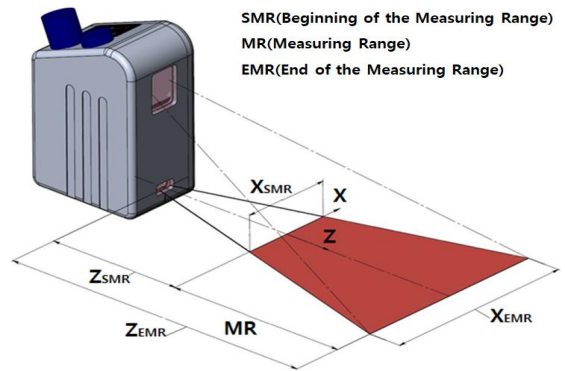


Fig. 1 The sensor shape and measurement range

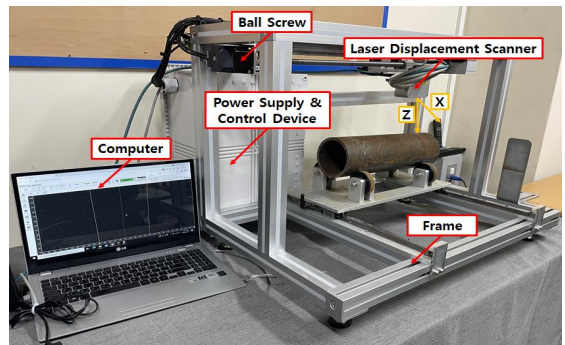


Fig. 2 Measuring equipment



Fig. 3 Pipe shape according to the degree of corrosion

데이터 용량과 스캔 속도를 고려하여 프레임 수는 284frame/s, 센서 이송속도는 30mm/s로 스캔하였다.

2.2 측정대상

산업에서 사용하는 배관 중 Fig. 3과 같이 고압가스 공급용 강관을 선정하였으며, 표면이 깨끗한 강관부터 모든 면이 부식된 것까지 다양한 형상의 강관을 무작위로 스캔하여 라벨링 되지 않은 표면 데이터를 총 228개 확보하였다.

3. 상태분류 알고리즘

3.1 클러스터링

클러스터링은 데이터 집합에서 유사한 특징을 가진 집단의 중심점을 기준으로 군집화시키는 과정이며, 사람이 수행하는 것이 아닌 컴퓨터가 수행함으로 알려지지 않은 군집을 발견할 수 있다⁹⁾. 클러스터링 기법은 Table 1과 같이 다양한 기법이 있으며, 동일한 데이터 집합을 상대가 다른 클러스터로 구분하기 위해서는 적절한 기법 선정이 중요한데, 일반적으로 클러스터링 기법 중 Partitioning methods의 k-Means와 Density-based methods의 DBSCAN (Density-Based Spatial Clustering of Applications with Noise) 알고리즘이 많이 사용되고 있다¹⁰⁾.

3.2 k-Means 알고리즘

k-Means는 클러스터에 대한 정보가 없을 때 유사한 데이터들을 하나의 클러스터로 만드는 대표적인 분할 기법으로, 데이터들과 데이터가 포함된 클러스터 중심과의 거리를 최소화하며, 서로 다른 클러스터 중심의 거리를 최대화하는 알고리즘이다. k-Means는 사용자가 사전에 클러스터 k개수를 설정하면 k개의 중심점(Centroid)을 무작위로 위치시키고, 중심점과 데이터와의 거리로 특정 클러스터에 데이터를 배정한다. 이때 중심점과 각 데이터들의 거리는 유클리디안 거리를 사용하며, 배정된 결과를 기준으로 각 클러스터 별로 평균값을 계산하여 중심점을 갱신하

고, 각 클러스터와의 거리 계산과 중심점 갱신 과정은 중심점의 변화가 충분히 작아질 때까지 반복한다. 이와 같은 k-Means 알고리즘은 데이터 세트의 모든 객체에 대한 제곱 오차의 합을 나타내는 식 (1)을 최소화하는 것으로 정의할 수 있다¹¹⁾.

$$E = \sum_{i=1}^k \sum_{p \in C_i} dist(p, c_i)^2 \quad (1)$$

여기서, C_i 는 클러스터를 표시하며, $dist(p, c_i)$ 는 C_i 클러스터에 속하는 데이터인 p 와 C_i 클러스터의 중심점인 c_i 의 유클리드 거리를 의미한다¹⁰⁾.

3.3 DBSCAN 알고리즘

DBSCAN은 초기 데이터로부터 같은 군집 내의 데이터들은 밀도가 높게 형성될 것이라는 가정에 밀집도가 높은 데이터들을 하나의 클러스터로 형성하는 알고리즘이다. DBSCAN은 클러스터의 개수를 사전에 지정하지 않고, 각 데이터들이 서로에게 이웃인지 판단하기 위한 최소 거리를 나타내는 Eps(Epsilon)와 클러스터로 인정하기 위한 Eps내의 최소 데이터 개수 MinPts(Minimum Points)를 지정해야 한다¹¹⁾.

DBSCAN은 임의의 데이터를 선택하여 Eps 반경 내에 MinPts 이상의 데이터가 있으면 군집을 형성하며, 클러스터링 과정을 Fig. 4에 나타내었다. P_1 은 클러스터의 중심점(Core point)이며, Eps내의 이웃

Table 1 Overview of clustering methods

Method	General Characteristics
Partitioning methods	<ul style="list-style-type: none"> - May use mean or medoid (etc.) to represent cluster center(Distance-based) - Effective for small- to medium-size data sets
Hierarchical methods	<ul style="list-style-type: none"> - Clustering is a hierarchical decomposition (i.e., multiple levels) - Cannot correct erroneous merges or splits
Density-based methods	<ul style="list-style-type: none"> - Can find arbitrarily shaped clusters - Clusters are dense regions of objects in space that are separated by low-density regions, Each point must have a minimum number of points within its "neighborhood"
Grid-based methods	<ul style="list-style-type: none"> - Use a multiresolution grid data structure - Fast processing time (typically independent of the number of data objects, yet dependent on grid size)

데이터를 중심으로 동일한 검사를 실시한다. 그러나, Eps내에 존재하지만 MinPts를 만족하지 못하여 클러스터를 형성하지 못하는 P_2 를 경계점(Border point)이라고 하며, 임의의 데이터가 이웃 데이터의 Eps반경 외에 있고, MinPts를 만족하지 못하는 데이터는 노이즈로 판단한다.

4. 인공지능 모델 학습 및 결과

4.1 데이터 전처리

센서는 레이저 라인을 방사하여 배관의 표면을 연속적으로 스캔하고, 한개의 라인마다 원주 방향으로 x, z 데이터를 645개씩 획득한다. 이때, x는 배관의 원주 방향 변위, z는 레이저 변위 센서와 배관 사이의 높이를 의미하며, 각 배관마다 동일한 면적의 부식 정도를 비교하기 위해 z축 데이터만 추출

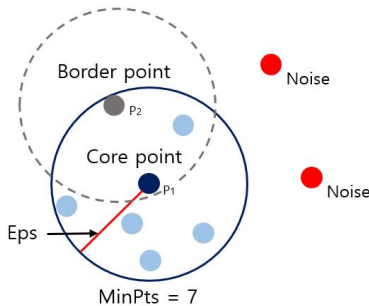


Fig. 4 DBSCAN

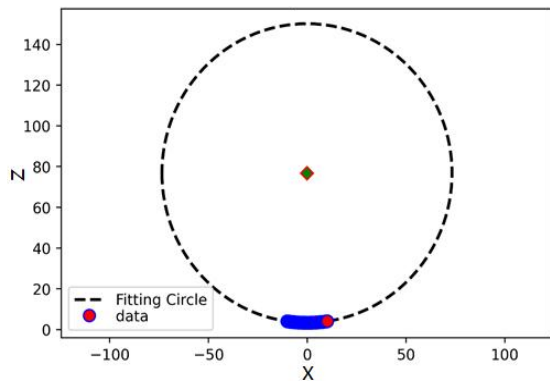
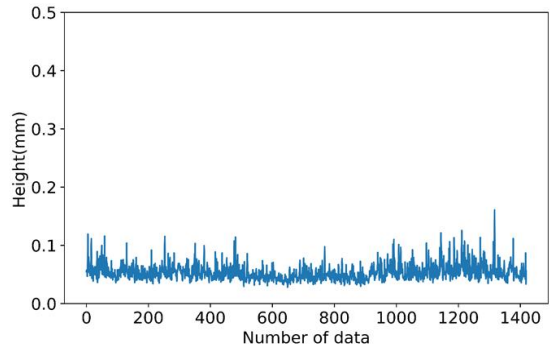


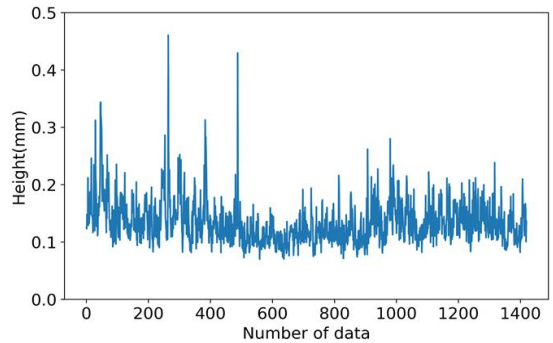
Fig. 5 Suitable circle from scan data

하고, 데이터 개수를 일치시켰다^[1]. 또한, 표면에서 빛의 반사로 인한 결측치는 원주 방향으로 선형 보간하여 보완하였다^[1].

또한, 검사 장비의 수평 오차로 인해 측정 시 센서에서 배관의 중심까지 편차가 발생하는 문제로 인해



(a) Average corrosion height



(b) Maximum corrosion height

Fig. 6 Corrosion height from the circumference direction

Table 2 Mean and max value of corrosion height

Pipe	Mean	Max
1	0.035280	0.071384
2	0.036549	0.073715
3	0.062327	0.162710
4	0.035017	0.071532
5	0.053354	0.134569
6	0.033834	0.0668618
7	0.060731	0.157388
8	0.035574	0.073704
9	0.061008	0.154202
10	0.066175	0.168117

표면의 부식 높이를 정확하게 비교하기가 곤란하였다^[1]. 이를 해결하기 위해 측정 데이터로부터 원의 방정식을 사용하여 Fig. 5와 같이 배관의 원주를 그렸으며, 이때, 표면으로부터 원주방향의 부식 높이 평균은 Fig. 6(a), 최대 부식 높이는 Fig. 6(b)에 나타내었다. 총 측정 면적의 평균 부식 높이와 최대 부식 높이의 평균값을 계산하였으며, 획득한 228개의 데이터 중 일부를 Table 2에 나타내었다.

4.2 상태 분류

k-Means 알고리즘은 사용자가 사전에 지정하는 클러스터의 개수에 따라 결과가 다르게 나타나며, 최적의 클러스터 개수를 도출하기 위해 엘보우 방법을 사용하였다. 엘보우 방법은 클러스터 개수를 변화시키며 SSE(Sum of Squared Errors) 값을 그래프로 나타내었을 때 기울기가 급격하게 변하는 지점을 최적의 클러스터 개수로 정하는 방법이다. 클러스터 개수에 대한 SSE 값을 Fig. 7과 같이 나타내었으며, 클러스터 개수가 3과 4에서 기울기가 변하는 지점으로 나타났지만, 급격하게 변하는 개수를 명확하게 판단할 수 없으므로 다른 평가 척도를 사용하였다.

SSE와 유사하게 클러스터 품질을 평가하는 척도인 DBI(Davies-Bouldin Index)를 사용하였으며, DBI는 클러스터 내에서의 Distribution과 비교하여 다른 클러스터 간의 분리 정도의 비율로 계산되는 값이다^[11]. 즉, DBI는 두 개의 클러스터에 대해 각 클러스터의 중심 간 거리를 각 클러스터의 크기의 합으로 나눈 값으로 식 (2)와 같이 나타낼 수 있으며, 값이 가장 작을 때 최적의 클러스터 개수로 판단한다^[12].

$$DB = \frac{1}{n} \sum_{i=1}^n \max \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right), i \neq j \quad (2)$$

여기서, n은 클러스터의 개수, c_x 는 클러스터 x의 중심점(centroid), σ_x 는 클러스터 x내의 모든 데이터로부터 중심점 c_x 까지의 평균거리 그리고 $d(c_i, c_j)$ 는 클러스터 i와 클러스터 j 사이의 중심점의 거리를 표시한다. 클러스터 개수에 대한 DBI 값을 Fig. 8과 같이 나타내었으며, 클러스터 개수가 4개일 때,

DBI 값이 가장 작게 나타났다.

따라서, 평가 척도 SSE와 DBI를 고려하였을 때, 적절한 클러스터 개수는 4개로 선정하였으며, 데이터셋을 4개로 클러스터링하면 Fig. 9와 같이 나타낼 수 있다.

한편, DBSCAN 알고리즘은 클러스터 개수를 사전에 지정하지 않고, 클러스터가 이웃 포인트를 찾을 수

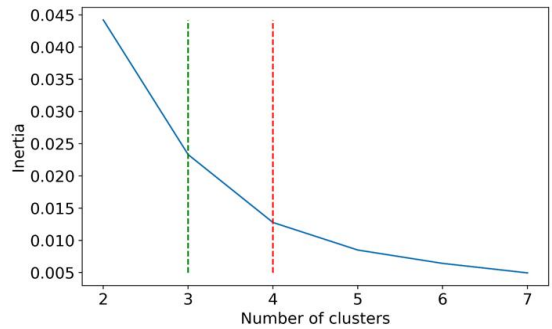


Fig. 7 SSE for number of clusters

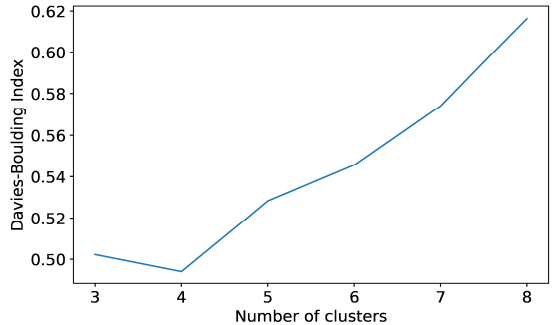


Fig. 8 DBI for number of clusters

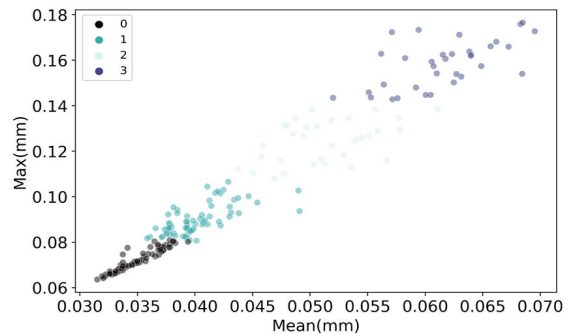


Fig. 9 k-Means results for corrosion height

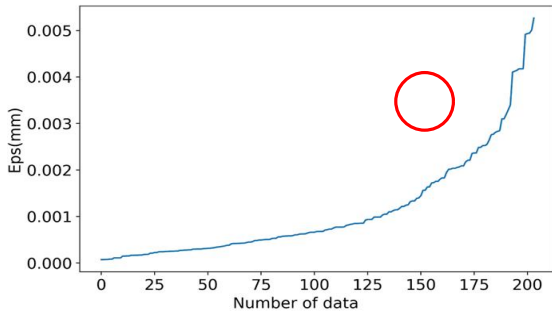


Fig. 10 4-distance graph

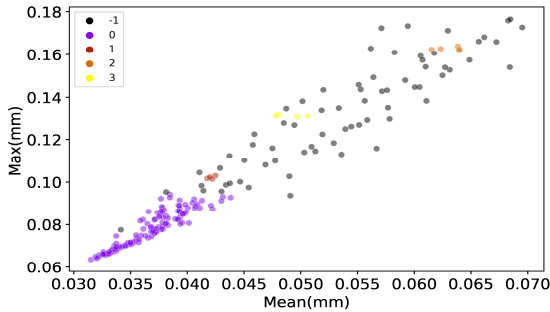


Fig. 11 DBSCAN results for corrosion height

있는 반경 또는 최대거리를 의미하는 Eps와 클러스터로 인정하기 위한 최소 데이터 개수인 MinPts를 선정해야 한다. MinPts는 DBSCAN 원문에서 제시하는 Heuristic method를 참고하였으며, 2차원 데이터는 MinPts를 4개로 하는 것을 권장하고 있다^[12]. Eps는 k-dist graph에서 엘보우 방법을 사용하였으며, Fig. 10과 같이 꺾이는 부분인 0.002를 Eps로 선정하였다^[13]. 도출된 MinPts와 Eps를 DBSCAN 알고리즘에 적용하면 Fig. 11과 같이 분류되며, 4개의 군집과 노이즈로 구분되는 것을 확인할 수 있다.

k-Means와 DBSCAN 알고리즘 모두 4가지 상태로 분류되었으나, DBSCAN 분류 결과는 노이즈가 많이 발생하므로 k-Means의 분류 결과가 적절함을 알 수 있다.

4.3 상태 기준 범위

k-Means 알고리즘으로 분류된 4개의 군집을 A 상태, B 상태, C 상태, D 상태로 설정하였다. 각 상태의 평균 부식 높이와 최대 부식 높이의 분포를

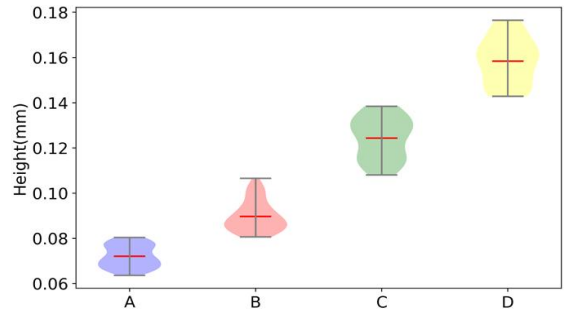


Fig. 12 Maximum distribution of corrosion height for states

Table 3 Maximum corrosion height by states

	A	B	C	D
Mean	0.0717	0.0893	0.124	0.158
Range	0.0636 ~0.0804	0.0807 ~0.1065	0.1081 ~0.1384	0.1428 ~0.1765

획득하였으나, 평균 부식 높이는 상태마다 중복되는 범위가 발생하여 제외하였으며, 최대 부식 높이에 대해 각 상태의 최대, 최소 및 평균값을 표시한 분포를 Fig. 12에 나타내었으며, 각 상태의 최대 부식 높이의 평균값과 범위를 Table 3에 나타내었다.

위의 그림과 표로부터 최대 부식 높이가 A 상태 배관은 0.0636~0.0804mm, B 상태 배관은 0.0807~0.1065mm, C 상태 배관은 0.1081~0.1384mm, D 상태 배관은 0.1428~0.1765mm로 분류할 수 있으며, 이러한 결과로부터 레이저 변위 센서를 이용하여, 배관의 상태 분류가 가능함을 알 수 있다.

5. 결론

배관의 부식 정도에 따른 상태를 분류하기 위해 레이저 변위 센서를 이용한 측정장비를 제작하였으며, 표면 정보를 통해 배관 부식 높이를 획득하였다. 비지도 학습을 통해 4가지 상태로 분류하였으며, 각 상태의 최대 부식 높이의 기준을 도출하였다. 얻어진 결과를 요약하면 다음과 같다.

- 1) 배관 부식에 대한 상태분류는 비지도 학습의 클러스터링 방법인 k-Means와 DBSCAN 알고리즘 중 k-Means 알고리즘이 적절하다.
- 2) 배관 부식 상태에 따라 4가지 상태로 구분이 가능하

며, 각 상태에 대한 최대 부식 높이의 범위를 선정하였다.

- 3) 차후에는 배관의 일부표면만을 측정하는 것이 아닌 모든 면을 검사할 수 있는 장비를 설계하고, 배관의 상태를 분류할 수 있도록 할 예정이다.

후 기

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업의 연구결과로 수행되었음(IITP-2022-2020-0-01612).

REFERENCES

1. Cheon, K. M., Shin, B. C., Shin, G. H., Go, J. I., Lee, J. H. and Hur, J. W., "Classification of the Rusting State of Pipe Using a Laser Displacement Sensor", Journal of the Korean Society of Manufacturing Process Engineers, Vol. 21, No. 5, pp. 46-52, 2022.
2. Yeo, H. J., "Development of a Robot System for Repairing a Underground Pipet", Journal of the Korea Academia-Industrial cooperation Society, Vol. 13, No. 3, pp. 1270-1274, 2012.
3. Park, J. W., Kim, W., Jung, B. C., Moon, S. J. and Yang, S. J., "SVM-based Damage Detection Technique Using Smart Valves in a Pipeline System", Transactions of the Korean Society of Mechanical Engineers, Vol. 45, No. 11, pp. 1019-1028, 2021.
4. Yun, K. W., Kim, D. L., Jung, H. C., Hong, D. P. and Kim, K. S., "Application Defects Detection in the Small-Bore Pipe Using Infrared Thermography Technique", Journal of the Korean Society for Nondestructive Testing, Vol. 33, No. 1, pp. 34-39, 2013.
5. Kim, H. M. and Park, G. S., "A Study on Determining the Shape of Small Axial Cracks by using Magnetic Flux Leakage in NDT System for Underground Pipe", Journal of the Korean Magnetism Society, Vol. 23, No. 1, pp. 18-25, 2013.
6. Park, H. S., Choi, M. Y., Park, J. H., Lee, S. S., Huh, Y. H., Lee, B. Y. and Kim, J. S., "Study on the Micro Crack Detection in Joints by Using Ultrasound Infrared Thermography", Journal of the Korean Society for Nondestructive Testing, Vol. 32, No. 2, pp. 162-169, 2012.
7. Technical Standard for Pipe Thickness Calculation and Inspection, Korea Occupational Safety and Health Agency, pp. 6-7, 1998.
8. Laser Scanners RF627 Series User's Manual, RIFTEK, pp. 4-5, 2019.
9. Yang, G. C., "An Efficient Conceptual Clustering Scheme", Journal of the Korea Entertainment Industry Association, Vol. 14, No. 4, pp. 349-354, 2020.
10. Data Mining (Third Edition), The Morgan Kaufmann Series in Data Management Systems, pp. 443-495, 2012.
11. Kim, J. W. and Shin, G. S., "Comparative Analysis for Clustering Based Optimal Vehicle Routes Planning", The Korea Journal of BigData, Vol. 5, No. 1, pp. 155-180, 2020.
12. Davies, D. L. and Bouldin, D. W., "A Cluster Separation Measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-1, Issue. 2, pp. 224-227, 1979.
13. Rahmah, N. and Sitanggang, I. S., "Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra", IOP Conference Series: Earth and Environmental Science, Vol. 31, 2015.