UDK 666.9.019:004.896
Original scientific article/Izvirni znanstveni članek

ISSN 1580-2949
Mater. Tehnol.

DEFECT CHARACTERIZATION OF METALLIC MATERIALS USING SELF-PARAMETERIZED DENSITY-BASED CLUSTERING AND COMPUTATIONAL INTELLIGENCE TECHNIQUES

KARAKTERIZACIJA NAPAK V KOVINSKIH MATERIALIH Z UPORABO SAMOPARAMETRIZIRANE GOSTOTE GROZDOV IN TEHNIK RAČUNALNIŠKE INTELIGENCE

J. V. Johnsonselva*, J. Raja Sekar

Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India

Prejem rokopisa – received: 2023-09-30; sprejem za objavo – accepted for publication: 2024-09-30

doi:10.17222/mit.2023.982

Rapid growth of manufacturing industries is propelled by transformative technologies such as machine intelligence, autonomous computing and non-destructive testing (NDT). During the manufacturing of wrought products, there is no guarantee that the final product is 100-% flawless. Thus, all final products are subjected to quality checking to identify and eliminate defective products. In industries, most of internal defects are identified using NDT techniques, which fail to precisely characterize the defects. In this paper a novel algorithm, called Self-Parameterized Density-Based Clustering (SPDBC), is proposed for defect characterization. The proposed clustering method uses spatial parameters to identify the size and position of defects by filtering out the noise and other data that correspond to the non-defect area. Using these filtered data, computational intelligence techniques are employed to predict the defect type. SPDBC achieved Jaccard indices of 97.02 % and 98.78 % for identifying the defect size and position, respectively. Gradient boosting regression trees (GBRT) achieved a maximum accuracy of 97.44 % in predicting the defect type. As a result, the proposed approach can assist NDT experts in various sectors to differentiate between problem severities faster and replace defective parts before any major breakdown occurs.

Keywords: material defect characterization, ultrasonic non-destructive testing, density-based clustering, artificial intelligence

Hitro rast industrij različnih izdelkov poganjajo nove prodirajoče tehnologije, kot so strojna inteligenca, avtonomno računalništvo in neporušno testiranje materialov (NDT; angl.: Non-Destructive Testing). Med proizvodnjo surovih izdelkov je težko zagotoviti, da bodo le-ti 100 %-no brez napak. Zato se končne proizvode oz. izdelke obvezno kontrolira z namenom, da se odstrani slabe, preden se jih dobavi naročnikom ali potencialnim kupcem. V industriji se večina notranjih napak ugotavlja z NDT tehnikami, toda le-te pogosto niso dovolj natančne, da bi z njimi lahko identificirali vse napake v določenem materialu. V tem članku avtorji opisujejo razvoj novega algoritma imenovanega samo-parametrizirana gostota na osnovi skupljanja (SPDBC; angl.: Self Parameterized Density Based Clustering) in ga predlagajo kot metodo za karakterizacijo napak v materialih. Predlagana metoda uporablja prostorske parametre za identifikacijo velikosti in položaja napak in nato izvrši filtriranje »hrupa« in ostalih podatkov, ki predstavljajo področja brez napak. S tehnikami računalniške inteligence nato uporabijo filtrirane podatke za napoved vrste napake. Z metodo SPDBC so avtorji dosegli Jaccardov indeks 97,02 % pri identifikaciji velikosti in 98,78 % za identifikacijo položaja napak. Gradientno ojačana regresijska drevesa (GBRT; angl.: Gradient Boosting Regression Trees) so dala maksimalno natančnost 97,44 % za napoved vrste napak. Posledično avtorji predlagajo, da bi lahko ta pristop pomagal ekspertom s področja NDT tehnik pri hitrejšem razlikovanju nevarnosti prisotnih napak in hitro zamenjavo z novimi rezervnimi deli, še preden pride do resnejših poškodb in/ali zlomov končnega izdelka.

Ključne besede: karakterizacija napak v materialih, ultrazvočne neporušne tehnike testiranja materialov, gostota na osnovi združevanja v grozde, umetna inteligenca

1 INTRODUCTION

In general, defects are defined as any kind of unwanted irregularities present inside the material structure. In an industrial environment, material defects pose a serious threat of causing equipment malfunctions. Some defects of industrial materials originate from the earlier stages of the manufacturing process such as casting and moulding. During quality control (QC),[1] most of defective materials are identified and removed but some may pass through because of their small size. There is a high possibility that defects which slipped through QC may become larger during further processing of wrought products. Thus, an industrial material may contain both internal and external defects. While an external defect is easy to observe with the naked eye, internal defects are hard to identify. Industrial equipment failures are common due to the growth of internal defects, which occur due to uneven temperature and stress in the material structure. Since internal defects are uncertain, it is challenging to accurately detect their presence without breaking the material structure. To identify internal defects precisely, non-destructive testing (NDT) is adopted in industries.[2,3] Some industries employ magnetic-based NDT to identify defects but it cannot be used for non-magnetic alloys and multi-layered composite materials. To overcome the limitations of magnetic-based NDT, ultrasonic-based NDT with the pulse-echo tech-

*Corresponding author's e-mail:
johnsoninboxx@gmail.com (J. V. Johnsonselva)

nique is widely used in industrial environments. Ultrasonic equipment operates very well in a harsh industrial environment.[4] Ultrasonic-based NDT uses through-transmission and pulse-echo techniques to analyze materials. The through-transmission technique is used on composite materials,[5] concrete,[6] aluminum alloys,[7] carbon fiber,[8] etc. The pulse-echo technique is used on metals and its alloys. The proposed work is concentrated only with metals, and so the pulse-echo technique is employed.

In the industrial domain, there are numerous challenges in performing a NDT analysis.[9] As per the literature, the ultrasonic NDT techniques are widely used for observing the integrity of metals. The researchers have developed a method to observe discontinuity in adhesive bonds using the ultrasonic pulse-echo technique.[10] This technique succeeds in observing the discontinuity even in multi-layered metal composites. A research proposal included an automatic approach for the inspection of composite materials using the ultrasonic pulse-echo technique.[11] However, the success rate of an NDT analysis depends upon the method and ability to collect the data.[12]

The most challenging and complex problem with ultrasonic NDT is that it needs a level-II expert to diagnose defects precisely. Defect characteristics that are collected with the ultrasonic equipment need to be analyzed manually by a human expert. Some methodologies for defect characterization using computational intelligence techniques are presented in the literature.[13,14] To deal with incomplete data, the system uses the Improved Mean Imputation Clustering Algorithm[15] and Kernel-Based Fuzzy C-Means Algorithm[16] with a considerably positive outcome. Research experts have developed a defect classification method using Probabilistic Neural Networks (PNNs).[17,18] The Convolution Neural Network (CNN) based ultrasonic testing is used to detect and classify railhead surfaces and subsurface defects.[19] A method has been developed using a neural network-based solution and radiographic images to categorize defects, but it is limited only to the identification of cracks.[20] The existing methods presented in the literature are limited only to the classification between defect and non-defect signals. The computation of defect characteristics such as defect size, position and type are not addressed. There are currently no studies in the literature covering automatic computation of defect characteristics such as defect size, position and type using computational intelligence techniques. Hence, an effective and automatic approach is required to aid the NDT experts in defect characterization.

In this work, a novel Self-Parameterized Density Based Clustering (SPDBC) algorithm is proposed to identify the defect size and position in wrought products. Several classifiers are employed in this work to categorize the defect type from the clustered data. This will assist the NDT experts to identify the severity of defects more effectively.[21] It is extremely beneficial for the industries that produce heavy-duty vehicles to locate defects in wrought products before they are subjected to the manufacturing process.

2 METHODOLOGY

The proposed design methodology contains loosely coupled hardware and software modules, which can detect, transmit and process the material defect information.

The existing ultrasonic hardware design was incorporated in this system which is commonly used in industrial applications. Still, this hardware model needs a level-II expert for precise evaluation of defects. To enable the user to operate the ultrasonic machinery without any expertise, a software model with high accuracy and precision in finding defects is required. Thus, a software model is proposed, which includes two stages, covering defect size, position calculation and defect type prediction as depicted in **Figure 1**. To implement the first stage, a novel algorithm called Self-Parameterized Density Based Clustering (SPDBC) is proposed to compute the size and position of a defect. The second stage deals with the defect type prediction using computational intelligence techniques.

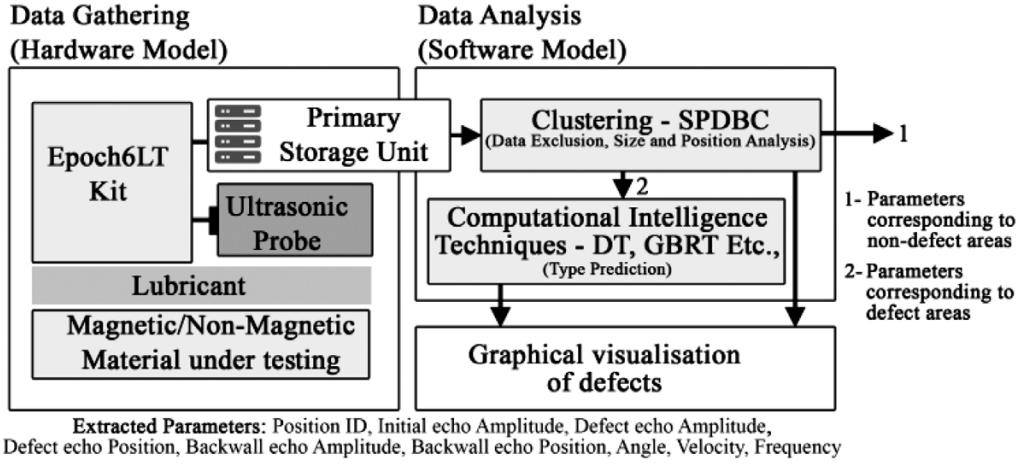

Figure 1: Workflow of the proposed methodology

2.1 Hardware model set-up

The hardware model is responsible for gathering raw data using ultrasonic NDT and storing it in the primary storage. The ultrasonic NDT in this research work is performed with the Epoch6LT ultrasonic kit, using the pulse-echo technique to detect the abnormalities in the internal material structure. The ultrasonic pulse-echo technique can be utilized using any ultrasonic probe with an oscillating frequency of 2–20 MHz. In ultrasonic NDT, the penetration ability of the ultrasonic wave is unaffected by changes in the material's magnetic properties, making it suitable for both magnetic and non-magnetic materials. Due to a unique behavior of ultrasonic waves in different materials, internal defects in a wide range of materials can be studied. Using the NDT data observed with the ultrasonic pulse-echo technique, structural details of defects can be clearly observed. Although the ultrasonic pulse-echo technique is suitable for all types of industrial materials, it needs a trained expert to diagnose defect characteristics in detail. Industrial research should be focused on honing ultrasonic NDT using knowledge mining algorithms. Additionally, defects need to be categorized to assess the extent of damage accurately. This classification helps prevent an unnecessary replacement of an entire part when the defect is minor. Hence, an effective method is proposed combining computational intelligence techniques and ultrasonic NDT to analyze and characterize internal defects found in industrial materials in order to support NDT experts in the industry.

To gather NDT data using the ultrasonic pulse-echo technique, the ultrasonic probe is placed on the material's surface with a layer of lubricant between them and a high-power ultrasonic beam of 4 MHz is passed inside the material as seen in **Figure 2a**. Due to the changes in the aquatic impedance between the material and atmospheric air, ultrasonic waves bounce back to the probe at the end of the material, forming a pulse-echo signal. The same reflection takes place in the presence of impurities, air molecules and voids inside the material structure. These pulse-echo signals may contain data that correspond to defects, non-defects, and outliers. Outliers often appear in data due to random noise, data loss, data misunderstanding, etc. In NDT, these outliers have to be handled effectively to increase the result accuracy as it degrades the performance of a prediction. A standard pulse-echo signal contains initial echo (IE), defect echo (DE) and back-wall echo (BE) as illustrated in **Figure 2b**. If there is no defect present inside the material, and nothing interferes with the beam path, there is no defect echo in the signal. The presence of a defect can be described with three different cases as follows below.

Case 1: Bigger back wall and smaller defect echo

Defects like cracks and discontinuities are thin defective structures present inside materials. Under ultra-sonic NDT, such defects only block a very small portion of the ultrasonic beam path. As a result, the defect echo of the pulse-echo signal is much smaller than the back-wall echo.

Case 2: Same size of the defect and back-wall echo

It is rare that both back wall and defect echo have the same size. This pattern can be seen in the verification of drill holes in an industrial environment and in the presence of an internal structure similar to a blister, porosity and shrinkage.[22] This kind of defects is classified as highly critical and need to be taken care of as soon as possible.

Case 3: Absence of the back-wall echo

This case is only possible if a defect is bigger than the beam path. Materials with bigger defects pose a serious threat to the material structure and the operating environment. The scenario of having no back wall during ultrasonic NDT is considered as most critical case in an industrial environment.

The data observed using the Epoch6LT ultrasonic kit is stored in its primary storage unit to be retrieved as '*.xml' and '*.csv' files for further analysis. The hardware model outputs several parameters, and among those, nine parameters are chosen for further processing. They include position ID, initial echo amplitude, defect echo amplitude, defect echo position, back-wall echo amplitude, back-wall echo position, angle, velocity, and frequency, as described in **Table 1**. These nine parame-

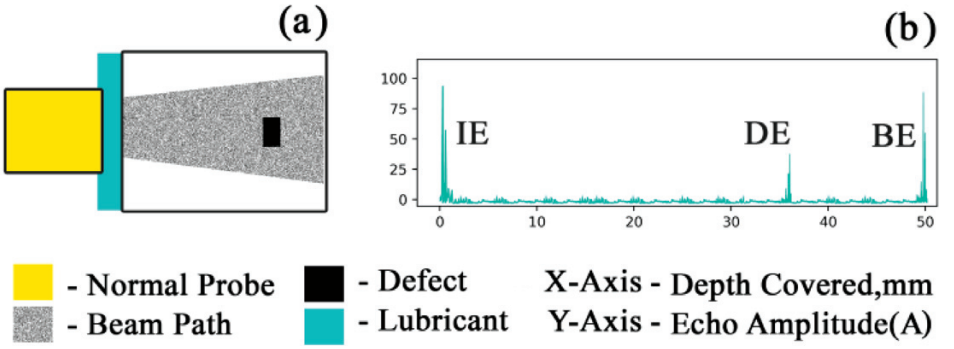

Figure 2: a) Ultrasonic pulse-echo principle, b) pulse-echo structure

ters of the pulse-echo signal significantly represent the characteristics of defects observed using ultrasonic NDT.

Table 1: Description of the ultrasonic NDT dataset

Attribute	Type	Value
Position ID (PID)	Integer	[1–8]
Initial echo amplitude (IEA)	Integer	[0–110] A
Initial echo position (IEP)	Integer	[0–2] mm
Defect echo amplitude (DEA)	Integer	[10–100] A
Defect echo position (DEP)	Float	[0–200] mm
Back-wall echo amplitude (BEA)	Integer	[10–100] A
Back-wall echo position (BEP)	Float	[1–200] mm
Angle	Integer	[0°, 45°, 60°]
Velocity	Integer	[1200–6300] m/s
Frequency	Float	[5–12] Hz

Table 2: Description of the materials utilized

Material ID	Sample type	Defect type	Length, *mm*	Width, *mm*	Height, *mm*
S_001	Industrial	*Ali*	123	64	15
S_002	Industrial	*Bli*	74	82	15
S_003	Industrial	*La*	74	195	15

All the defective materials utilized in this work were collected from broken industrial equipment and for a better understanding of the samples, their details are described in **Table 2**.

2.2 Software design

The software model retrieves the parameters stored in the primary storage and processes them using machine learning techniques. Initially, the SPDBC algorithm uses the defect echo position and back-wall echo position to cluster all defects individually and to exclude non-defect areas. From the clustered defect data, the size and position of defects are calculated using the SPDBC algorithm. Meanwhile, the algorithm filters out the parameters that correspond to the non-defect areas. Following that, computational intelligence techniques are used to accurately predict the defect type.

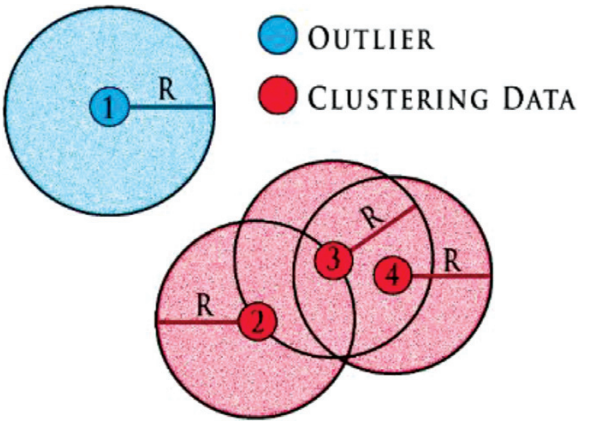

Figure 3: Illustration of SPDBC

2.2.1 Proposed SPDBC algorithm for defect size and position identification

The density-based clustering method works based on the spatial metrics between data. This method is well known for its ability to cluster spatial data and filter out noise/outliers. It identifies a cluster as a region of high point density, separated from regions of low point density. The idea here is to segregate groups of data with similar traits and assign them to the same cluster. By using this method for a defect analysis, multiple defects within a material can be identified precisely, with exact positions and sizes. The accuracy of density-based clustering can also be increased by adjusting the spatial parameters. However, the limitation of density-based clustering lies in the fact that it uses more computational time as it iterates through all possible parametric values and requires human assistance.

To overcome this limitation and achieve accurate results effectively, the SPDBC method is proposed. It is a clustering method derived from density-based clustering like Density Based Dynamically Self-Parameterized Clustering for Material Inspection (DBDSPCMI).[23] Unlike other density-based clustering methods, SPDBC uses a classification and clustering approach. This approach uses a minimum threshold (T_{min}) value of 10 and a maximum threshold (T_{max}) value of 100 to identify the data as belonging to defects and non-defects. The values of T_{min} and T_{max} are determined to be the possible amplitudes of the data gathered from defective portions of materials as mentioned earlier in **Table 1**. The data from the defective portion of a material is considered valid (V), while the data from the non-defective portion is considered invalid (I), using Equation (1). This approach helps us to improve the speed and accuracy of SPDBC by identifying defect and non-defect data during clustering.

$$f(x)=\begin{cases} V, DEA > T_{min} \text{ and } T_{max} < 100 \\ I, \text{ Otherwise} \end{cases} \quad (1)$$

During SPDBC, effective spatial parameters such as reachability (R) and density factor (D) can be computed automatically. R is the distance metric used to identify a possible connected neighbor (q) around the data point (p) where D is the minimum density of q required by p. Data points p and q are said to be connected neighbors only if they are within the q distance and have a minimum of D neighbors within R, as illustrated in **Figure 3**. These spatial parameters are computed using Equations (2) and (3).

$$R=\left(\frac{\text{Material Length in mm}}{\text{Total No. of Samples Observed}}\right) \quad (2)$$

$$D=\frac{1}{R} \quad (3)$$

The defect size can also be roughly determined using Equation (4). If a defect is identified to be big, then each cluster is considered as a defect boundary. In case of a

big defect, defect boundaries are reconstructed as a single defect to observe its size and position.

$$f(x) = \begin{cases} \text{Smaller, } DEA \le BEA \\ \text{Bigger, Otherwise} \end{cases} \quad (4)$$

We define x as the data, where $x(i,j)$ are the data points. For each data point $x(i,j)$ the number of connected neighbors within R should be D_n, This can be expressed as

$$f(x) = \begin{cases} 1, D_{\min} < D_n < D_{\max} \\ 0, \text{ Otherwise} \end{cases} \quad (5)$$

In Equation (5), the number 1 stands for a defect whereas number 0 stands for a non-defect/outlier. Also, ΔD ($\Delta D = D_{max} - D_{min}$) depends on the number of observations per millimeter and varies accordingly.

2.2.2 Computational intelligence techniques for a defect-type prediction

It is observed that the amplitude of the defect echo and back-wall echo varies depending on the defect structure even though the initial echo amplitude is kept as a constant. Using the parametric changes in the pulse-echo structure that varies based on the defect's properties, the type of a defect that resides inside the material can be recognized precisely. In machine learning and data science, the computational intelligence techniques are primarily used for data classification and prediction. To identify the optimal existing model for predicting defects, a set of algorithms such as K-Nearest Neighbor (KNN),[24] Support Vector Machine (SVM),[25] Decision Tree (DT),[26] Naive Bayes Classifier (NBC),[27] Random Forest Classification (RFC),[28] Adaptive Boost (AdaBoost),[29] Stochastic Gradient Descent (SGD),[30] Artificial Neural Network (ANN)[31] and Gradient Boosting Regression Trees (GBRT),[32] are used. The goal is to determine, which algorithm can forecast defects more accurately during ultrasonic NDT by comparing their applicability based on their features.

2.2.3 Performance evaluation metrics

The performance of the proposed SPDBC algorithm is statistically evaluated in terms of the Jaccard index (*JI*) and the classifiers are comparatively evaluated using the accuracy metric. The Jaccard index, often known as the Jaccard similarity coefficient, is a metric used to measure the similarity between two sets of data as given in Equation (6). Accuracy is the ratio of correctly classified samples to the total number of samples as given in Equation (7).

$$JI = \frac{TP}{TP + FP + FN} \quad (6)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

Abbreviations *TP*, *FP*, *TN*, and *FN* stand for true positive, false positive, true negative, and false negative, respectively. These values are obtained from the confusion matrix.

3 RESULTS AND DISCUSSIONS

Here, the structural integrity of defect samples used and the rate, at which the ultrasonic pulse-echo observations are made are discussed in detail. After that, defect characterization using the proposed SPDBC algorithm and computational techniques are observed.

3.1 Details of defect samples employed in the work

The sample materials used in the analysis are made of low-carbon iron and include defect such as alligatoring (*Ali*), blister (*Bli*) and lamination (*La*). For a better understanding of defect samples used in the proposed work, some of their cross-sections are shown in **Figure 4**. In an ideal environment, these defect structures are not necessarily considered as a threat, but under industrial circumstances, stress and temperature changes cause them to weaken the material structure. Thus, mate-

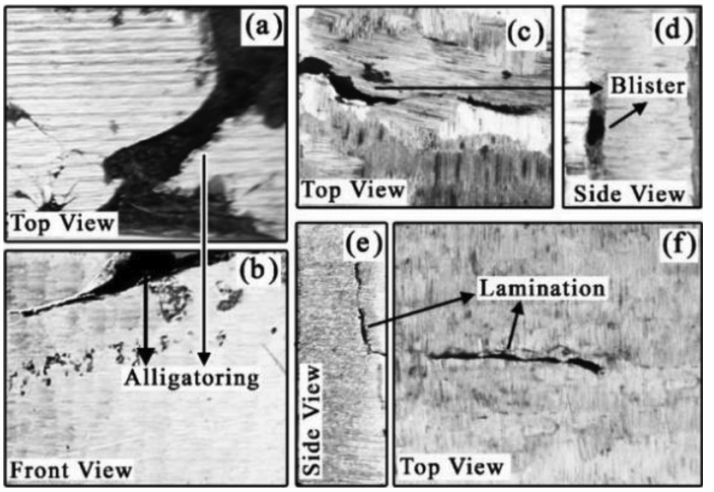

Figure 4: Cross-sections of (a–b) S_001, (c–d) S_002, (e–f) S_003

rials with internal defects need to be eliminated from wrought products during manufacturing to prevent failure.

Ali is a defective structure, caused by accidental splitting of the material during hot rolling. Although defect *Ali* in **Figure 4a** and **4b** seems to appear at the edge of the material structure, ultrasonic inspection reveals the presence of microcracks around it. Thus, it is not safe to just trim off the defective edges as they already caused microcracks on the material. These microcracks cause structural failure under stress even after the welding of the surface. During casting, gases like nitrogen are absorbed by the material during heating and released during cooling. If they have no space to move out of the material, the released gases are trapped inside, forming porosity.

When a material with porosity is processed as a wrought product, trapped gases expand along with the material and form a *Bli* defect inside the final product, as shown in **Figure 4c** and **4d**. During material processing, rapid heating and cooling cause the material to expand and contract. This kind of expansion and contraction of the molecular structure causes shrinkage or cavity inside the material. When it is processed as a wrought product using milling, the cavity will be elongated in the material structure, forming *La* as shown in **Figure 4e** and **4f**.

3.2 Experimental analysis of data gathering

To visualize a defect's structure with ultrasonic NDT, the number of observations per millimeter (n/mm) must be sufficient to define the defect structure accurately. Also, the distance for each observation must be as consistent as possible. To select the effective n/mm, a series of inspections with different n values such as n {0.1, 0.2, 0.5, 1.0 and 2.0} is carried out on S_000, as seen in **Figure 5**. As the sample is fabricated in a laboratory, its dimensions are well known, as illustrated in **Figure 5a**.

The visual similarity between the observed defect structure and the actual known structure can be well differentiated by plotting the *DEP* against the material length. This also shows how effective the ultrasonic NDT can be if properly used on materials. From the analysis, it is observed that only 0.1's/mm, as in **Figure 5b**, can recognize a defect. Increases in the pulse-echo signal such as 0.2's/mm and 0.5's/mm make it difficult to identify the geometrical structure of a defect as these observations do not cover all the defect structure clearly. **Figure 5c** and **5d** implies that the increase in n's/mm increases the visualization of defect accuracy. However, it also confirms that the increase in n's/mm increases the similarity between the actual and observed defect. As we increase n's/mm, a higher accuracy in visualizing defects is achieved using 1.0's/mm and 2.0's/mm, as shown in **Figure 5e** and **5f**. To identify the effective n's/mm, the similarity measure for all series of observations is computed using the Jaccard coefficient or Jaccard index, Minkowski distance, Manhattan distance and Cosine distance, as shown in **Figure 5g**. By multiplying these similarity measures by one hundred, the Jaccard index of the size and position identification can be computed. All similarity measures imply that the convergence has been achieved at 1.0's/mm. Comparably, the outcome after convergence gives better results than before convergence. To achieve accurate results of defect identification, anything higher than 1.0's/mm is acceptable. By inspecting all available samples at n's/mm, a labeled dataset with a total of 74800 observations is created for further analysis of defect prediction.

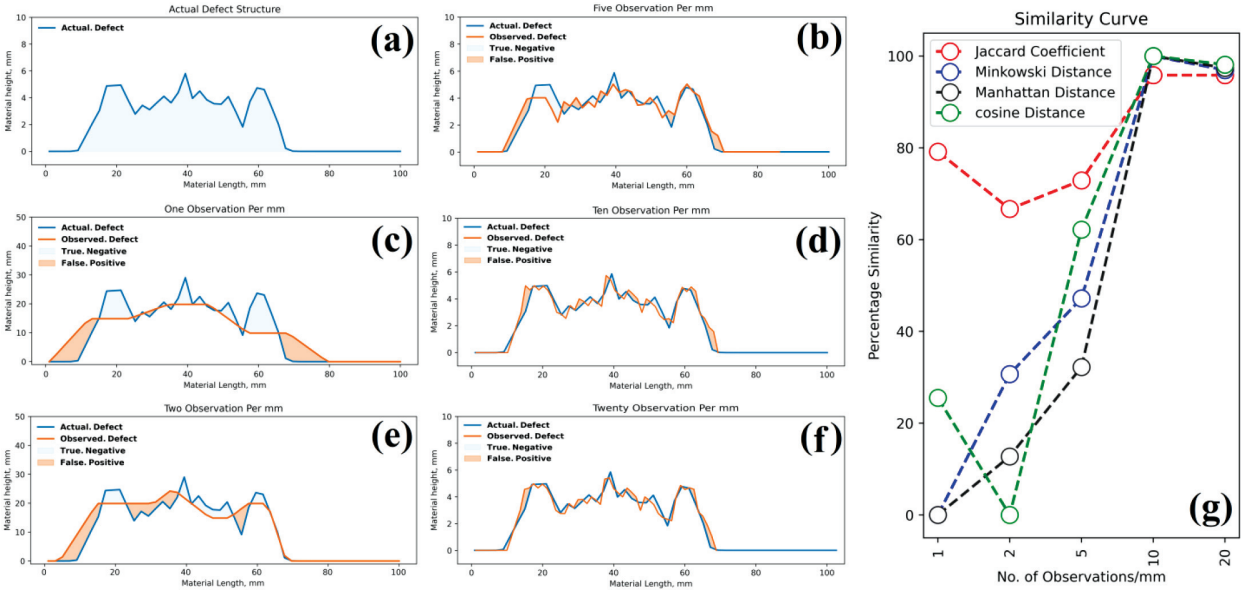

Figure 5: a) Illustration of a defect in S_001; (b, c, d, e, f) series of different numbers of observations/mm; (g) similarity evaluation curves for different observations/mm

3.3 Defects size & position computation using the proposed SPDBC algorithm

To identify a defect inside the S_001 material sample, 350 pulse-echo observations are made at a rate of 2.8 s/mm on the material surface and the data is filtered using Equation (1). In the clustering process, only the data identified as *V* are clustered using parametric values $R = 0.35$ and $D = 2.8$ obtained from Equations (2) and (3). A plot of *DEP* against *PID* is created. On relationship graph between the defect and back-wall amplitude in **Figure 6a** shows that the defect echo is much smaller than the back-wall echo in all instances, and the pattern represents the presence of smaller defects such as microcracks or discontinuity as per Equation (4). Thus, each cluster in the spatial representation is considered as an individual defect. From the spatial representation of defects observed in **Figure 6d**, the size and position of defects can be analyzed precisely. The result shows that material S_001 has 9 defective structures (A1-A9).

The S_002 sample is subjected to the same pulse-echo technique using a 4 MHz frequency and 180 observations made at a rate of 2.4 s/mm. The observed data is filtered using Equation (1) and processed using SPDBC. In the clustering process, the data identified as *V* are clustered using parametric values $R = 0.41$ and $D = 2.4$ obtained with Equations (2) and (3). The amplitude relationship graph in **Figure 6b** shows that *DEA* and *BEA* intersect and that there is no back wall at some sites representing a bigger defect as per Equation (4). Thus, the clusters present in the spatial representation in **Figure 6e** are considered as the boundaries of the defect.

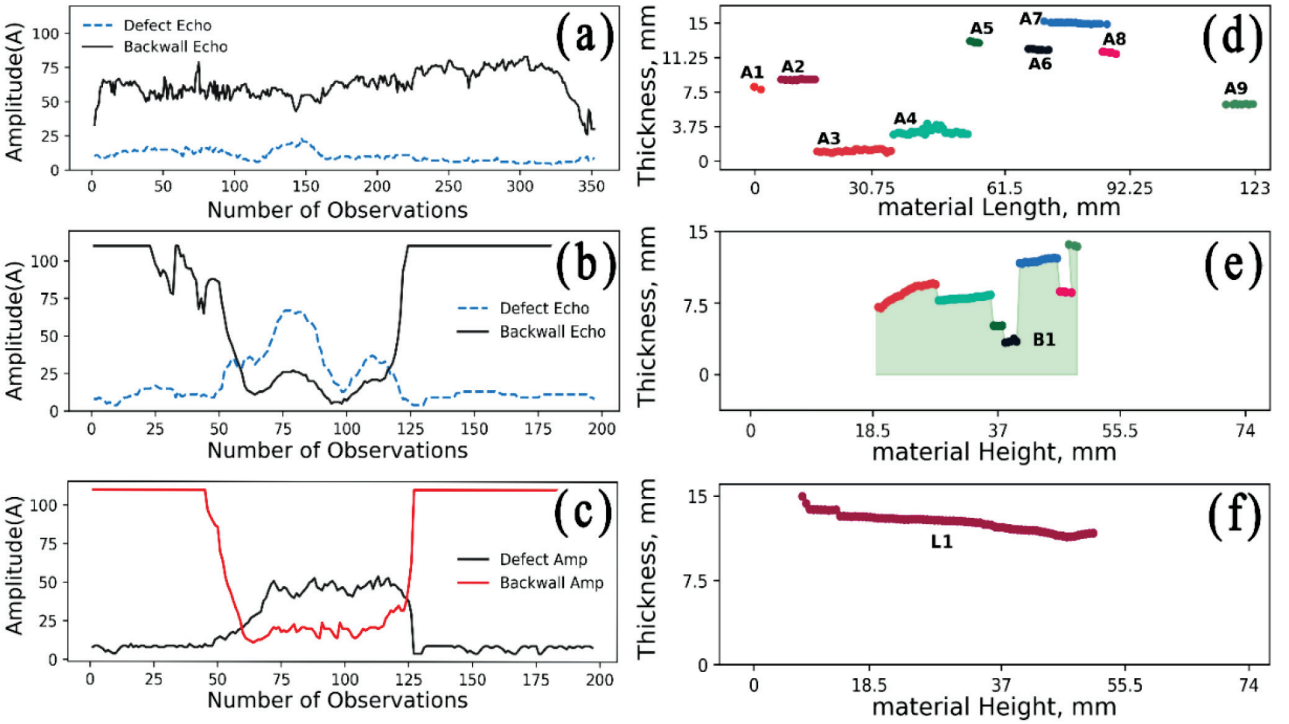

Figure 6: (a-c) Amplitude relationship graphs, (d-f) SPDBC results

Table 3: Defects observed on S_001, S_002 and S_003

Material ID	*R*	*D*	Amplitude (A)		Defect ID	Position, mm		Size, mm	
			DEA	BEA		X-axis	Y-axis	X-axis	Y-axis
S_001	0.35	2.8	4–24	26–83	A1	0–2.6	7.25–8.5	2.6	2
					A2	6–17	8.25–9.25	11	1.4
					A3	15.5–36.6	0.5–1.75	21.1	1
					A4	35–56.5	2.45–4.5	21.5	2.2
					A5	55–59.5	12.45–13.5	4.5	1.9
					A6	66.25–73.5	11.6–12.6	7.25	1.5
					A7	70.5–89	14.45–15.55	18.5	1
					A8	85.5–91	11.25–12.25	5.5	1.7
					A9	118–123	5.75–6.7	5	1.5
S_002	0.41	2.4	4–67	5–110	B1	18.5–48	3–14.1	29.5	11.1
S_003	0.37	2.7	4–67	11–110	L1	7.5–52	11–14.5	44.5	3.5

Defect boundaries are reconstructed to visualize defect geometry from the point of contact of the ultrasonic probe. The result shows that the S_002 material includes only one defective structure (B1). On the S_003 sample, 198 observations are made using the pulse-echo technique at a rate of 2.7 s/mm and the data is filtered with Equation (1).

During clustering, the data identified as *V* are clustered using parametric values $R = 0.37$ and $D = 2.7$ obtained with Equations (2) and (3). From the amplitude relationship graph in Figure 6(c), it is observed that *DEA* and *BEA* intersect each other, but there is no sign of the absence of the back wall. As per Equation (4), it is indeed a big defect but not big enough to block the beam path. The result shows that the S_003 material includes only one defective structure (L1), as illustrated in **Figure 6f**. For a better understanding, a detailed description of defects observed during SPDBC is included in **Table 3**. The accuracy of defects identified using SPDBC is 98.78 % (Jaccard index: 0.9878) in terms of the defect position and 97.02 % (Jaccard index: 0.9702) in terms of the defect size, obtained with the help of reference defects fabricated in our laboratory.

3.3.1 Comparison of the proposed algorithm with existing techniques

The ultrasonic data gathered in this work was studied using also different existing clustering algorithms such as K-means,[33] hierarchical algorithm,[34] affinity propagation clustering (APC),[35] DBSCAN[36] and DBDSPCMI.[23] Individual industrial materials might include several defects, like in the case of S_001. Clustering methods like K-means and hierarchical algorithm need to be initialized with a number of clusters. As a result, they fail to classify the data about materials with multiple defects as it would lead to misclassification. On the other hand, APC gives better results in finding defects by clustering data based on affinity otherwise termed as similarity. The only issue with using APC is that it cannot remove the noise from the data. As ultrasonic data includes noise as well, density-based clustering methods, such as DBSCAN and DBDSPCMI, are used.

Table 4: Comparative analysis using the Jaccard index (in %)

Algorithm	Defect size (Jaccard index)			Defect position (Jaccard index)		
	Single defect	Multiple defects	Average	Single defect	Multiple defects	Average
K-means[33]	95.91	13.09	54.50	94.53	11.51	53.02
Hierarchical[34]	96.02	17.51	56.80	94.07	11.74	52.91
APC[35]	93.06	92.51	92.79	93.43	92.79	93.11
DBSCAN[36]	95.87	95.00	95.44	95.12	94.65	94.89
DBDSPCMI[23]	97.23	96.70	96.97	95.68	94.80	95.24
SPDBC	**99.23**	**98.32**	**98.78**	**97.68**	**96.36**	**97.02**

From these results, it is observed that these algorithms successfully classify data from the materials with multiple defects, using spatial metrics and removing the noise as well. Using density-based clustering methods, the defect position and size can be observed precisely. The Jaccard index, used as an accuracy measure, shows that the existing methods are inferior to the proposed method in identifying a defect structure accurately, as described in Table 4. It is observed that the highest accuracy, in terms of the Jaccard index, among the existing methods is 96.97 % for defect size calculation and 95.24 % for defect position computation. However, the proposed SPDBC algorithm achieves Jaccard indices of 98.78 % and 97.02 % for defect size and position computations.

3.4 Defects type prediction using computational intelligence techniques

After filtering out the parameters of non-defect areas using the proposed SPDBC algorithm, the remaining parameters only partly correspond to defects *Ali*, *Bli* and *La*. The nine parameters that correspond to them are only used as input data for the classifiers. An analysis is performed using two-fold, three-fold, five-fold, ten-fold and twenty-fold cross-validation ratios and a common training and testing ratio of 7:3. The input data contains 74,800 row vectors and nine column vectors for defect characterization. A row vector indicates the number of observations and a column vector indicates the nine parameters. The input data include five different classes such as *Ali*, *Bli*, *La,* empty and outlier. The empty class refers to the data gathered from the non-defect portion of the material, and the outlier is random data, considered as an error of data collection. Testing results for the classifier are derived from the confusion matrix and performance metrics such as accuracy, area under curve (AUC), F1 score and precision, calculated as shown in **Tables 5** to **9**.

Table 5: Testing results (two-fold, 7:3 ratio)

Model	Accuracy	AUC	F1	Precision
KNN	90.22 %	0.961	0.891	0.888
SVM	86.22 %	0.966	0.850	0.845
DT	95.11 %	0.959	0.938	0.926
NBC	81.77 %	0.940	0.835	0.861
RFC	92.44 %	0.962	0.919	0.904
AdaBoost	93.33 %	0.951	0.930	0.929
SGD	80.88 %	0.839	0.771	0.805
ANN	80.44 %	0.954	0.749	0.799
GBRT	**94.66 %**	**0.972**	**0.940**	**0.934**

Table 5 displays testing results for the two-fold cross-validation ratio, where DT predicts a defect with a high accuracy of 95.11 %. In comparison, the results obtained with GBRT are higher than those obtained with DT in terms of AUC, F1 score and precision value.

Table 6: Testing results (three-fold, 7:3 ratio)

Model	Accuracy	AUC	F1	Precision
KNN	90.22 %	0.963	0.891	0.886
SVM	85.78 %	0.964	0.947	0.935
DT	96.00 %	0.958	0.947	0.935
NBC	88.00 %	0.963	0.879	0.878
RFC	96.00 %	0.976	0.954	0.943
AdaBoost	93.78 %	0.953	0.941	0.949
SGD	85.33 %	0.892	0.863	0.864
ANN	86.22 %	0.955	0.849	0.845
GBRT	**96.44 %**	**0.997**	**0.961**	**0.958**

Table 6 displays testing results for the three-fold cross-validation ratio. Here, the performance results obtained with GBRT are higher than those of other algorithms in terms of accuracy, AUC, F1 and precision value. It is evident that an increase in the folding value from two to three increases the accuracy of the entire algorithm considerably, except for KNN, which shows no significant changes.

Table 7: Testing results (five-fold, 7:3 ratio)

Model	Accuracy	AUC	F1	Precision
KNN	92.00 %	0.972	0.908	0.898
SVM	85.84 %	0.960	0.850	0.845
DT	96.42 %	0.964	0.951	0.939
NBC	88.89 %	0.961	0.884	0.879
RFC	96.00 %	0.976	0.956	0.943
AdaBoost	96.00 %	0.967	0.951	0.943
SGD	86.22 %	0.888	0.861	0.861
ANN	86.67 %	0.955	0.854	0.851
GBRT	96.44 %	0.992	0.961	0.958

Table 7 displays testing results for the five-fold cross-validation ratio. Here, we see no great changes in the GBRT parameters, but AdaBoost seems to perform better than with three-fold cross-validation. Still, GBRT achieves the highest accuracy in all aspects, while KNN, SVM, NBC, AdaBoost, SGD and ANN also show increased accuracy.

Table 8: Testing results (ten-fold, 7:3 ratio)

Model	Accuracy	AUC	F1	Precision
KNN	91.77 %	0.977	0.904	0.895
SVM	86.77 %	0.969	0.850	0.845
DT	93.82 %	0.966	0.951	0.939
NBC	83.82 %	0.963	0.870	0.870
RFC	93.53 %	0.992	0.966	0.973
AdaBoost	94.12 %	0.969	0.957	0.954
SGD	86.47 %	0.891	0.861	0.861
ANN	87.94 %	0.957	0.850	0.844
GBRT	95.29 %	0.989	0.974	0.974

Tables 8 and **9** display testing results for the ten-fold and twenty-fold cross-validation ratios. For the ten-fold cross-validation, most algorithms first start to reach convergence, after which accuracy starts to decrease. At this stage, SGD and ANN still show an increase in accuracy, but their performance is too low compared to GBRT. The final testing result for the twenty-fold cross-validation shows that GBRT achieves the highest accuracy and precision along with the highest AUC and F1 score.

Table 9: Testing results (twenty-fold, 7:3 ratio)

Model	Accuracy	AUC	F1	Precision
KNN	92.00 %	0.981	0.908	0.899
SVM	87.11 %	0.973	0.859	0.853
DT	95.56 %	0.959	0.945	0.934
NBC	88.89 %	0.961	0.889	0.892
RFC	97.33 %	0.988	0.960	0.948
AdaBoost	96.89 %	0.975	0.962	0.961
SGD	86.67 %	0.886	0.857	0.857
ANN	88.89 %	0.967	0.876	0.870
GBRT	97.78 %	0.993	0.974	0.974

Thus, the proposed work uses an ultrasonic-based NDT technique, which is extremely suitable for all types of metals as well as their alloy compounds, achieving high accuracy of defect characterization. The prime significance of the proposed SPDBC algorithm is that it excludes non-defect parameters from the whole set of parameters covering both defect and non-defect areas. Since the proposed SPDBC algorithm automatically selects defect parameters, the time needed for the entire defect analysis decreases. The main advantage of the proposed model is that it helps an analyst detect a material defect faster and more effectively, without acquiring the knowledge of NDT.

4 CONCLUSIONS

In this work, a hybrid approach combining hardware and software models was developed to detect the internal defective structures of metals and metal alloys, using ultrasonic NDT. The proposed work aims to determine the size and position of a defect using the proposed SPDBC algorithm, followed by computational intelligence techniques to predict the defect type.

The proposed SPDBC algorithm accurately identified a defect with Jaccard indices of 98.78 % for defect position computation and 97.02 % for defect size computation.

By removing non-defect data prior to clustering, the SPDBC algorithm clusters defects 10.89 % faster than the existing DBSCAN and DBDSPCMI algorithms.

The system achieved the highest accuracy of 96.44 % in predicting the defect type using the GBRT computational intelligence technique.

Therefore, the proposed approach will assist NDT professionals in industries in identifying and differentiating the severity of faults faster so that they can replace defected parts before a significant breakdown may occur. This model is extremely useful for locating defects in wrought products before they are used in the industries that produce heavy-duty vehicles.

5 REFERENCES

[1] A. Sifa, A. S. Baskoro, S. Sugeng, B. Badruzzaman, T. Endramawan, Identification of the Thickness of Nugget on Worksheet Spot Welding Using Non Destructive Test (NDT) – Effect of Pressure, IOP Conf. Ser.: Mater. Sci. Eng., 306 (**2018**), 012009, doi:10.1088/1757-899X/306/1/012009

[2] S. Sambath, P. Nagaraj, N. Selvakumar, S. Arunachalam, T. Page, Automatic detection of defects in ultrasonic testing using artificial neural network, Int. J. Microstruct. Mater. Prop., 5 (**2010**) 6, 561–574, doi:10.1504/IJMMP.2010.038155

[3] D. Selvathi, I. H. Nithilla, N. Akshaya, Image Processing Techniques for Defect Detection in Metals using Thermal Images, Proc. of the 3rd Inter. Conf. ICOEI, 7 (**2019**) 2, 939–944, doi:10.1109/ICOEI.2019.8862616

[4] A. El Kouche, H. S. Hassanein, Ultrasonic non-destructive testing (NDT) using wireless sensor networks, Procedia Comput. Sci., 10 (**2012**) 2, 136–143, doi:10.1016/j.procs.2012.06.021

[5] M. Jolly, Review of Non-Destructive Testing (NDT) Techniques and their Applicability to Thick Walled Composites, Procedia CIRP, 38 (**2015**) 1, 129–136, doi:10.1016/j.procir.2015.07.043

[6] Y. Jia, L. Tang, P. Ming, Y. Xie, Ultrasound-excited thermography for detecting microcracks in concrete materials, NDT E Int., 101 (**2019**) 2, 62–71, doi:10.1016/j.ndteint.2018.10.006

[7] A Savin, Influence of the defect characteristic in aluminum alloy 7075 in non-destructive electromagnetic testing, IOP Conf. Ser.: Mater. Sci. Eng., 126 (**2022**) 2, 012029, doi:10.1088/1757-899X/1262/1/012029

[8] E. Jasiūnienė, L. Mažeika, V. Samaitis, V. Cicėnas, D. Mattsson, Ultrasonic non-destructive testing of complex titanium/carbon fibre composite joints, Ultrasonics, 95 (**2019**) 1, 13–21, doi:10.1016/j.ultras.2019.02.009.

[9] H. Towsyfyan, A. Biguri, R. Boardman, T. Blumensath, Successes and challenges in non-destructive testing of aircraft composite structures, Chinese J. Aeronaut., 33 (**2020**) 3, 771–791, doi:10.1016/j.cja.2019.09.017

[10] R. Leiderman, A. M. B. Braga, Scattering of guided waves by defective adhesive bonds in multilayer anisotropic plates, Wave Motion, 74 (**2017**) 3, 93–104, doi:10.1016/j.wavemoti.2017.05.007

[11] T. D'Orazio, M. Leo, A. Distante, C. Guaragnella, V. Pianese, G. Cavaccini, Automatic ultrasonic inspection for internal defect detection in composite materials, NDT E Int., 41 (**2008**) 2, 145–154, doi:10.1016/j.ndteint.2007.08.001

[12] M. E. Buchanan, Methods of data collection, AORN J., 33 (**1981**) 1, 137–149, doi:10.1016/S0001-2092(07)69400-9

[13] J. Peters, Computational Intelligence: Principles, Techniques and Applications, Comput. J., 50 (**2007**) 6, 758–765, doi:10.1093/comjnl/bxm073

[14] L. Sepulvene, Performance Evaluation of Machine Learning Techniques for Fault Diagnosis in Vehicle Fleet Tracking Modules, Comput. J., 65 (**2021**) 8, 2073–2086, doi:10.1093/comjnl/bxab047

[15] H. Shi, P. Wang, X. Yang, H. Yu, An Improved Mean Imputation Clustering Algorithm for Incomplete Data, Neural Process. Lett., 54 (**2020**) 5, 203–212, doi:10.1007/s11063-020-10298-5

[16] D. Q. Zhang, S. C. Chen, Clustering Incomplete Data Using Kernel-Based Fuzzy C-means Algorithm, Neural Process. Lett., 18 (**2003**) 3, 155–162, doi:10.1023/B:NEPL.0000011135.19145.1b

[17] T. M. Meksen, B. Boudraa, M. Boudraa, Defects clustering using kohonen networks during ultrasonic inspection, IAENG Int. J. Comput. Sci., 36 (**2009**) 3, 1–4

[18] S. Arivazhagan, J. Jasline Tracia, N. Selvakumar, Mater. Res. Express, 6 (**2019**) 9, 096539, doi:10.1088/2053-1591/ab2d83

[19] I. Ghafoor, P. W. Tse, N. Munir, A. J. C. Trappey, Non-contact detection of railhead defects and their classification by using convolutional neural network, Optik (Stuttg.), 253 (**2022**) 2, 168607, doi:10.1016/J.IJLEO.2022.168607

[20] V. A. Golodov, A. A. Maltseva, Approach to weld segmentation and defect classification in radiographic images of pipe welds, NDT E Int., 127 (**2022**) 3, 102597, doi:10.1016/J.NDTEINT.2021.102597.

[21] Y. Ahmad Al-Maharma, P. Sandeep Patil, B. Markert, Effects of porosity on the mechanical properties of additively manufactured components: a critical review, Mater. Res. Express, 7 (**2020**) 12, 122001, doi:10.1088/2053-1591/abcc5d

[22] S. Sambath, P. Nagaraj, N. Selvakumar, Automatic Defect Classification in Ultrasonic NDT Using Artificial Intelligence, J. Nondestruct. Eval., 30 (**2011**) 1, 20–28, doi:10.1007/s10921-010-0086-0

[23] P. Radha, N. Selvakumar, J. Raja Sekar, J. V. Johnsonselva, Density-Based Dynamically Self-Parameterized Clustering for Material Inspection, Comput. J., 66 (**2021**) 2, 416–428, doi:10.1093/comjnl/bxab169

[24] N. S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat., 46 (**1992**) 3, 175–185, doi:10.1080/00031305.1992.10475879

[25] H. F. Chen, In silico log p prediction for a large data set with support vector machines, radial basis neural networks and multiple linear regression, Chem. Biol. Drug Des., 74 (**2009**) 2, 142–147, doi:10.1111/j.1747-0285.2009.00840.x

[26] J. R. Quinlan, Induction of decision trees, Mach. Learn., 1 (**1986**) 1, 81–106, doi:10.1007/bf00116251

[27] Z. Muda, W. Mohamed, M. D. Nasir, I. U. Nur, K-Means Clustering and Naive Bayes Classification for Intrusion Detection, Journal of IT in Asia, 4 (**2016**) 2, 13–25, doi:10.33736/jita.45.2014.

[28] T. K. Ho, Random decision forests, Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, 5 (**1995**) 1, 278–282, doi:10.1109/ICDAR.1995.598994

[29] Y. Freund, R. E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, J. Comput. Syst. Sci., 55 (**1997**) 1, 119–139, doi:10.1006/jcss.1997.1504.

[30] Y. Tian, Y. Zhang, H. Zhang, Recent Advances in Stochastic Gradient Descent in Deep Learning, Mathematics., 11 (**2023**) 3, 682, doi:10.3390/math11030682

[31] E. Grossi, M. Buscema, Introduction to artificial neural networks, Eur. J. Gastroenterol. Hepatol., 19 (**2007**) 12, 1046–1054, doi:10.1097/MEG.0b013e3282f198a0

[32] D. D Nguyen, T. H. Nguyen, GBRT-based model for predicting the axial load capacity of the CFS-SOHS columns, Asian J. Civ. Eng., 24 (**2003**) 4, 3679–3688, doi:10.1007/s42107-023-00743-w

[33] S. Na, L. Xumin, G. Yong, Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm, ISIIT&SI'10, 5 (**2010**) 3, 63–67, doi:10.1109/IITSI.2010.74

[34] X. Ran, Y. Xi, Y. Lu, Comprehensive survey on hierarchical clustering algorithms and the recent developments, Artif. Intell. Rev., 56 (**2003**) 7, 8219–8264, doi:10.1007/s10462-022-10366-3

[35] B. J. Frey, D. Dueck, Clustering by passing messages between data points, Science, 315 (**2007**) 5814, 972–976, doi:10.1126/science.1136800

[36] J. R. Sekar, N. Selvakumar, P. Radha, J. V. Johnsonselva, Supervised and unsupervised learning for characterizing the industrial material defects, International Journal of Business Intelligence and Data Mining, 21 (**2022**) 3, 233–246, doi:10.1504/IJBIDM.2022.10039148