**Borneo Engineering & Advanced Multidisciplinary International Journal (BEAM)**
Volume 4, Issue 1, May 2025, Pages 39-44

# Deep Learning-Based Defect Detection from Ultrasonic Testing Imaging of Stainless Steel Plate

**Full Paper**

**Article history**
Received
*17 January 2025*
Received in revised form
*12 March 2025*
Accepted
*17 March 2025*
Published online
*1 May 2025*

Saipul Azmi Mohd Hashim[1]*, Muhammad Nasir Marzuki[2], Mohd Firdaus Abdullah[1], Muhammad Zhafran Mohaiyiddin[1]

[1]Unit Teknologi Pembuatan, Kolej Komuniti Kepala Batas, 13200, Kepala Batas, Pulau Pinang, Malaysia
[2]Unit Pengurusan, Kolej Komuniti Kepala Batas, 13200, Kepala Batas, Pulau Pinang, Malaysia

*Corresponding author: saipulazmi@kkkba.edu.my
Please provide an **official organisation email** of the corresponding author

## Abstract

Ultrasonic Testing is widely applied to inspect product and building structure with unseen defect in manufacturing industries. Furthermore, the advancement of the test changes from waveform signal into image which clarifies the unseen defects. The clarity of the unseen defects causes it is widely accepted in the industries. Though, ultrasonic images screening by operators the to detect the defect are prone to misjudgments. Therefore, this paper aims to automate the test using Deep Learning based approach using SqueezeNet model. Besides, the automated system is tested stainless steel plate with artificial defects from ultrasonic test image. The image is designed in two classes – 100 images with defect and 100 images without defect. Then the total number of 200 images is labelled and classes into 70% of learning data, 20% for testing data, and 10% for validating data for the following system modelling stage. The detection rate stands at 81.67% based on testing and validating data. Moreover, for the all evaluation measures resulted above 80% – Precision, Recall, and Accuracy. Statistically, the best performance of the model rated by F-score at 81.36%. In addition, by margin of error value, the used data and by chance to get the similar result at above 73% with 99% confidence level. These findings suggest the used model is fit to detect defects based on the ultrasonic test imaging image at best performance.

*Keywords: - Ultrasonic test, SqueezeNet model, defect detection*



## 1. Introduction

Early defect detection inspection is a crucial element in preventing any disaster due to defective structure. Besides, delaying detecting the defective may expand the defect size, hence resulted significant consequences. Moreover, internal defect is manually non-traceable, this assistance is needed to detect the defect. Due to the cost for destructive testing is not economical for a periodic inspection, hence Non Destructive Test (NDT) is the better alternative to lower the cost. The NDT such as Ultrasonic Test (UT) is well-known in detecting internal defect for multitude material, i.e. metals, plastics, composites, and ceramics (Ye et al., 2018).

UT uses an electric pulse went through a test material is manipulated to generate ultrasonic signal. Once the defect is detected, this resulted wave energy generates ultrasonic signal to the surface structure. These resonate wave energy used for indicating the location of the defect. The result is a waveform visualised via monitor which observed by an expert. This test advantages include it is highly perceptive to the most material defect, and ability in defect location and size determination (Taheri & Hassen, 2019).

Alternatively, UT can be presented by a B-scan image. The image is developed by a series of lines of the ultrasonic wave signal. Besides, the image is useful to visualize the subsurface visual in two dimensions. Thus, converting the waveform signal into ultrasonic imaginary image is expected to gain wider advantage in the industry. This is proven by the fact that B-scan image is widely used in medical applications. In manufacturing industries, the image usage is to detect invisible defect in manufactured products, and industrial building and machine structures. While ultrasound applied frequencies greater than 20 kHz in generating waveform signal, 2 Mhz and above frequencies are applied for ultrasound image. This frequency at shorter wavelength lets small internal structure resolution (Rathbun et al., 2023) than can visually see on the monitor. In addition, the generated image location determined by the time required for a pulse signal travels out from a pulse laser scan unit, it reflects from a structure, and return to a transducer. Then the signal is digitalised by an amplifier and a digital oscilloscope. This information helps the industries to locate the defect location effectively.

Deep Learning (DL) is the interest in this study due to it is successfully used in detecting defect structure in digital image. Bhole & Kumar (2020) suggests the use of DL in automating image classification surpassed other approaches in research and application. There are also cases, which DL surpasses human capability (Manakitsa et al., 2024). Besides, DL is the current research trend that is more focused image feature extraction and automating image classification (Archana & Jeevaraj, 2024). In addition, DL is reserved as the basic approaches in artificial intelligence studies due to its excellent performance (Elkorany & Elsharkawy, 2021). Moreover, DL works sensibly satisfactory in employing image usage due to it is inspired by the normal techniques of living thing conception (Obaid et al., 2022). Thus, these scenarios rationalized this approach, and it is comprehensible for used in this study.

A model development and training in DL approach can be conducted in either Training from Scratch, and Adjusting a Pre-Trained model (Bhole & Kumar, 2020). This study interested in applying the second approach, also known as Transfer Learning (TL). By TL, much work model development is simplified since the model is designed by previous researcher and proven effective. Because of the model is proven effective, thus it is predictable to be well-operational and applicable to other data. In order the model functions well with new data, Bhole & Kumar (2020) suggests the work required are model selection, modifying, fitting, and retraining. Due to it is a well-developed and pre-trained model, thus small amount of data and computational time are required for be re-training to a new data compare to the use of a new developed model. The instance of TL models includes Xception, Darknet-19, ResNet-50, and SqueezeNet (Beale et al., 2018). For this study, the interest is applied SqueezeNet to classify ultrasound test image either with or without defect.

Due to the UT data from the industry are often protected, thus limited data can be obtained, and lead to small amount of study to automated ultrasonic test is conducted (Posilović et al., 2022a). Hence, due to limited number of data is obtained, resulted supervised learning approach is the option in this study. Besides, many studies conducted in automated defect detection based on ultrasound image, still the researches are adopted supervised learning approach (Posilović, et al., 2022b). By this approach, automate the test is believed will improve the result's speed and reliability, and fortunately may reduce the test cost (Shipway et al., 2021).

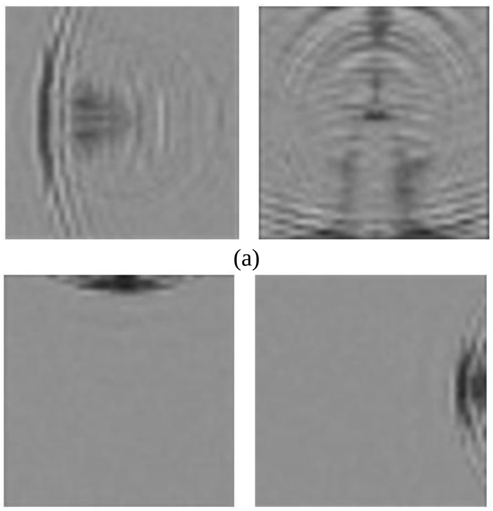
Through literature study, there are studies in this research line. For example, Ye et al. (2018) used two models -- USseqNet and USresNet model for ultrasonic imaging inspection in their work. Medak et al. (2022) self-developed model named by DefectNet tested on UI image inspected six steel blocks. In addition, Posilović et al. (2022a) applied Ganomaly, PaDIM, and DifferNet models; and Posilović et al. (2022b) applied Copy/Paste, DetectionGAN, and Modified SPADE GAN models both onto UT image of six blocks containing defects in the internal structure. McKnight (2024) used self-developed DL model for evaluating aerospace composite using UT image. Cheng et al. (2023) used CNN-LSTM model to inspect UT image of geometric shape of composite material with internal damage. Currently, Zhu et al. (2024) used UI image to inspect wind turbine blade damage using UCD-YOLO model. Hence, this finding suggests that there is no researcher applied SqueezeNet model in the research line.

It has become increasingly challenging to find a sample of the test image. This area of study has been gaining attention recently, and several DL models available for selection. As the result, this research exploring significant new opportunities. This paper aims to apply a DL-based approach to automatically detect defects in UT images. This paper begins by the spells out the methodology in section 2. Section 3, comprises the data result and analysis, and then wrap up section 4 for conclusion.

## 2. Methodology

### *2.1 Data Acquisition*

Due to the UT image is difficult to obtain, this study applied data from a secondary source from Ye et al. (2018). The source has 100 samples of data for each defected and with no defect images. Therefore, the total number of data is 200 samples and saved in two folders – defect image and no defect image as in Fig. 1. In addition, the data required special format for each model, hence for SqueezeNet framework each image is set to 224 x 224 bit size of data and in RGB image format. Besides, to randomised image data, the sequence of the image is rearranged using shuffle command. This is due to this research adopting unsupervised learning approach. Then, the data is divided into three sub-sets as follow: 70% (140) for training, 20% (40) for testing, and 10% (20) for validating. In addition, these three classes of image selection are set at random by using the shuffle command, so that the used system will learn the classes at a more even rate. The image from the source, is designed to form artificial defects are prepared in a batch of stainless steel plates with various types of flaws with 3mm thickness. There are four types of defect are fabricated to the plate – penetrated hole, 1.5mm depth of hole, penetrated slit, and 1.5mm depth of slit. The hole and slit defects prepared on both front and back sides of the steel plates. According to Ye, Ito, & Toyama (2018), following is the system's features to obtain the image: 1Mhz for Probe frequency, 90° for Beam angle, 500Hz pulse repetition frequency, and five Incident angles set at 0°, 22.5°, 45°, 67.5° and 90°. The captured images will not visualised the defect form, but it represents the cumulative of ultrasound wave that represents defect detection.



(a)

(b)

Fig. 1. Sample of RGB format images for the ultrasonic testing imaging (a) ultrasound image with defect and (b) ultrasound image without defect (Ye et al., 2018)

### *2.2 Deep Learning*

Elkorany & Elsharkawy (2021) defined that the SqueezeNet model designed as shown in Fig. 2 and Table 1. The model design starts with a Conv1 and Conv10 filters the input that produces the model influences detecting feature activation. Then, the layer developed a filter once the filter is multiplied by a set of weights which resulted input filter in two-dimensional. This resulted, a filter which suitable to filtered image inputs (Li et al., 2021). Afterward, the filtered input is designed with the interest information gathered from the whole image, to detect a specific feature in it, and not targeting on a point on the entire image area. Besides, this layer operates in two convolutional layers (Conv1 and Conv10). Conv1 squeezes the data into 1 x 1 convolutional filter and Conv10 expends the data using 1 x 1 and 3 x 3 convolutional filters (Ashhar et al., 2020). The squeeze data operation is to generate the input array multiple times at different points on the entire image area. This mechanism caused the model an effective model (Wang, 2023). Furthermore, while the parameter minimization is initiated in Conv1, the model still needs supplementary eight fire modules, i.e. fire2–fire9. These eight fire modules reduce the inputs resulted parameter number minimization (Obaid et al., 2022). These modules designed in sequence progressively minimizing input parameters, resulted only effective parameters obtained. Moreover, the three Max pool layers and an Average pool layer, are operated to chain the output generation (Guo et al., 2017). Lastly, Soft Max layer activates a function and place on the outer layer. This used to predict the score for each output. Based on the output convolutional result, is gathered for following result and analysis section.

Table 1. The design of SqueezeNet model

| Layer (Symbol) | Number of Layer |
|---|---|
| convolution layers (Conv1 and Conv10) | 2 |
| fire module layers (fire2-fire9) | 8 |
| max pooling layers (Max pool) | 3 |
| global average pooling layer (Average pool) | 1 |
| softmax layers (Soft Max) | 1 |

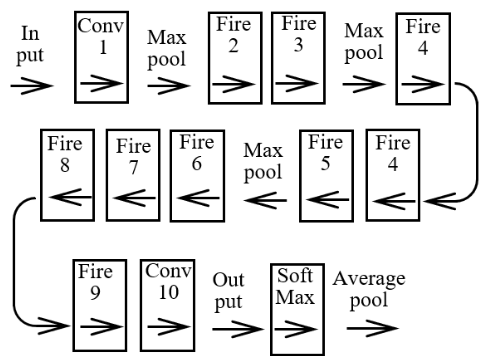



Fig. 2. SqueezeNet model

#### *2.2.1 System Modelling*

In this study, Matlab 2023 software is used to design the model. Through the software, Designer stage is simplified by the use existing model template. Thus, only selection and setting of input and output of the model at Designer stage is required. Then, in Data stage the 140 images of training data are Augmented set on following setting random rotation: Random Rotation (-90° minimum, 145° maximum), Random Rescaling (0.9 minimum, 1.3 maximum), Random horizontal pixels translation (-30 minimum, 45 maximum), and random vertical pixels translation (30 minimum, 45 maximum). By augmentation setting the trained model is expected can be be robust enough to detect the correct image class in following stage. The 20 validating image data is selected a validating process. In Training stage, the model training settings need to be set include Learning Rate, Max Epochs, and Validation Frequency. These ready data is applied to the deep learning model the learning rate is set at 0.005 and maximum epochs are set at 5. The number of image sets is priory done in the previous stage and input in this stage.

The training process produced charts showing both accuracy and loss as shown in Figure 3. The repetitive work is done at this stage, the attempt is to get close to the ideal graph, i.e. increasing in accuracy and decreasing in loss can be obtained. If the data is not following as the attempt, hence the training data, model architecture, and training options likely to be adjusted. In this study, the training is options are reselected so that's close to ideal result is obtained.

## 3. Result and Discussion

### *3.1 System Framework Training Accuracy and Loss*

Beale et al. (2020) suggests that an epoch is an entire full training cycle of learning data, for learning process. This process is set to a maximum epoch of 5 to avoid time wastage by using the software defaulted epoch. By the setting, the result of training shown in Figure 3. Full accuracy and loss for the result is measured through the training stage, hence it is resulted beginning after the iteration is a 0 value. Accordingly, the iteration is ends at epoch 5 as initially setup. In addition, the rate of correct classification represents by accuracy, and the rate of incorrect classification represents by loss. Fig. 3 shows the accuracy curve is going uphill, the loss curve showing downhill. If the poor accuracy is obtained, thus the re-setting is needed. By prediction command, testing data and validating data can be used as the performance indicator. The predicted result is shown in following subsection.

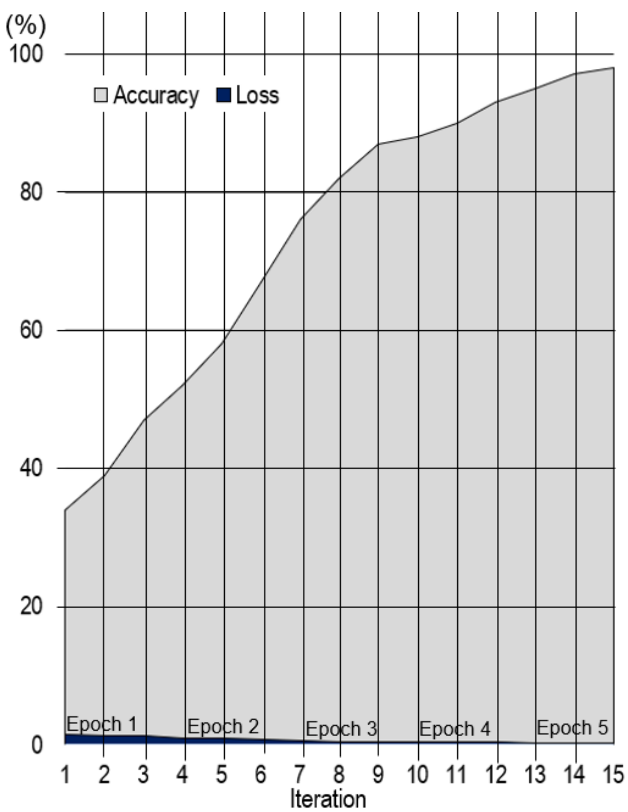



Fig. 3. Accuracy and loss chart

### *3.2 System Performance Evaluation*

The Correctly Classified Common charts shown three Confusion Charts in Fig. 4 and Fig. 5 for Histogram charts, indicate the categories that are being confused by the proposed system as other category in this case image with defect, and image with no defect. The chart shows the summary of how many detections are correct and incorrect per category. Besides, testing and validating data is fully used as the performance indicator in this study. The chart shown in Fig. 4 indicates the true category in vertical and the predicted category in horizontal. In addition, Correct and Incorrect Rate Histogram chart shown in Fig. 5. Based on validating data, the correct rate for image with defect is 90% and image with no defect is 80% – the average classification rate is 85%. The average classification rate for testing data is 80%. Then, the combined data, i.e. validating and testing data, the average classification rate found is 81.67%.

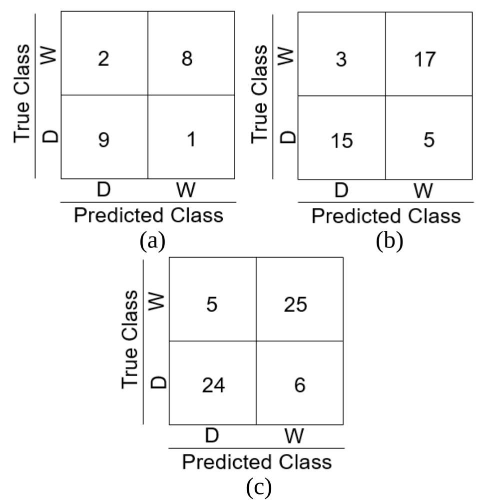
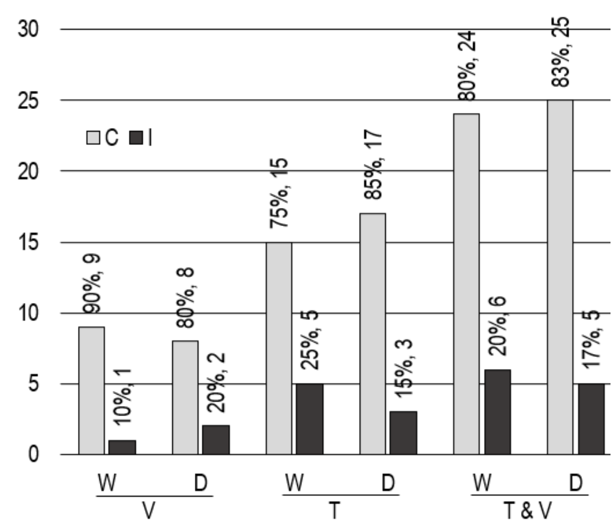



Fig. 4. Confusion chart for (a) validating data, (b) testing data and (c) testing and validating data




Note: C = Correct, I = Incorrect, W = Without defect, D = With defect, T = Testing Data, V = Validating Data

Fig. 5. Histogram chart

Based on technical performance evaluation, the Classification Rate (CR) shown in equation (1) is used to measure the model performance (Wen et al., 2020) in images classifying. For overall performance used Testing and Validating data, the number of Correctly Classified (CC), i.e. 49, and divided by the total number of samples (n), i.e. 60, the value of CR is 81.67%. This value indicates good CC value for the model classifying the images.

$$CR = CC/n \; x \; 100\% \qquad (1)$$

### 3.3 Evaluation Measurement

In addition to prior analysis, researchers evaluate image data by evaluation measurement to access defect detection (Hashim et al., 2025). This data category is image correctly classified (P) and image incorrectly classified (N). Broaden categorized of the data that represents parameters in this study includes: the images with defect correctly classified number known as True Positive (TP); True Negative (TN) represents the images with no defect correctly classified; images with no defect incorrectly classified represented by False Positive (FP); and images with defect incorrectly classified represents as False Negative (FN). Based on these parameters, following are the measurement performance evaluation parameters used to measure the proposed model performance – Precision (Pr), Recall (Re), and Accuracy (ψ).

Precision value computes the system's accuracy in classifying an image as positive or normal image. Recall is the name measures the system's competence to classify positive images. In addition, the value of accuracy is stated as the ratio of the correct classified images for the total number of images. These values are displayed classification performance in percentage. Based on the validating data result, the value of P is 17, N is 3, TP is 9, TN is 8, FP is 1, and FN is 2. Thus, the value of Pr is 0.9 (90%), Re is 0.818 (81.8%), and Ψ is 0.85 (85%). These values are calculated based on equation (2), (3), and (4). Besides, the performance is calculated by combining validating and testing data, resulted P is 49, N is 11, TP is 24, TN is 5, FP is 24, and FN is 25. Thus, the value of Pr is 0.8276 (82.76%), Re is 0.8 (80%), and Ψ is 0.8167 (81.67%).

$$\mathrm{Pr} = \frac{TP}{TP + FP} \quad (2) \qquad (2)$$

$$\mathrm{Re} = \frac{TP}{TP + FN} \quad (3) \qquad (3)$$

$$\Psi = \frac{TP + TN}{P + N} \quad (4) \qquad (4)$$

### 3.4 Statistics Tests

This study measures performance indicators using two sample subsets: testing and validating data. Therefore, two option are available for performing the statistical tests. The fact is, the minimum required sample size (n) for the statistical test is equal or above 30. Due to number of validating data is below 30, for statistical test both finding from testing and validating data is combined. Hence, the total number of data is 60.

F-score (γ) as in equation (5) is a statistical parameter indicates the best performance achieved by a system. The highest value is 1.0, that representing ideal value for $Pr$ and $Re$. The excellent contributions of the DL with SqueezeNet model used in this study, by this analysis the F-score value is 81.36%. In addition, a statistic test is conducted with the data in this study to know the result margin of error and confidence interval. By the test, the idea is to know in following prediction what is the percentage to be re-obtained the result CR at 81.67%. By 95% desire confidence (z* = 1.96), the margin of error is plus and minus 6.52%, and the confidence interval is between 88.19% to 75.15%. In addition, by 99% desire confidence (z* = 2.58), the margin of error is plus and minus 8.58%, and the confidence interval is between 90.25% to 73.58%. These findings told that for the model used, the confidence level is 95% to re-obtained the value (Rumsey, 2019) for CR is above 75.15% and the confidence level is 99% to re-obtained the value of CR is above 73.58%.

$$\gamma = 2\frac{Pr \times Re}{Pr + Re} \quad (5) \qquad (5)$$

## 4. Conclusion

Two classes of UT images are used for classification: images with defect and images without defects, utilizing SqueezeNet model in DL. The model prediction resulted outstanding performance at 80% and above –precision, recall, and accuracy. In addition, for the second conclusion, statistically confirmed the system high level performance at F-score is 81.36%. Therefore, this paper indicates the used model is an effective to do classification tasks by replicating UT operation. The third conclusion suggests that the result of validating data agrees with the result by testing data by average at 83.34% of accuracy. This indicates the collected data is consistently at the whole level. Besides, the statistical test also concludes that the results through this study is at high statistical confidence – non-bias nor the result is not by chance, it can 73% be achieved repetitively at 99% confidence level.

For future study, the used model is suggested to embed in a more general system that should be applied with realistic amount of data. Hence, unsupervised learning approach can be adopted. Moreover, for future study the system can be extended as a dynamic image, e.g. video, since in manufacturing industry the inspection process involved the whole part or structure with defect. Furthermore, the suggestion is to apply other NDT tests in manufacturing facilities such as Radiographic Testing and Manual Visual Inspection to widen the model applicability and efficiency, and exploring the model's adoption for other manufacturing process system.

**Acknowledgement:** This research was not funded by any grant.

**Author Contributions:** The research study was carried out successfully with contributions from all authors.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

Archana, R., & Jeevaraj, P. E. (2024). Deep learning models for digital image processing: a review. *Artificial Intelligence Review*,

Ashhar, S. M., Mokri, S. S., Abd Rahni, A. A., Huddin, A. B., Zulkarnain, N., Azmi, N. A., & Mahaletchumy, T. (2021). Comparison of deep learning convolutional neural network (CNN) architectures for CT lung cancer classification. *International Journal of Advanced Technology and Engineering Exploration*, *8*(74), 126.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2020). Deep learning toolbox. *R2020a User's Guide; The MathWorks, Inc.: Natick, MA, USA*.

Bhole, V., & Kumar, A. (2020). Analysis of convolutional neural network using pre-trained squeezenet model for classification of thermal fruit images. In *ICT for competitive strategies* (pp. 759-768). CRC Press.

Cheng, X., Ma, G., Wu, Z., Zu, H., & Hu, X. (2023). Automatic defect depth estimation for ultrasonic testing in carbon fiber reinforced composites using deep learning. *Ndt & E International*, *135*, 102804.

Elkorany, A. S., & Elsharkawy, Z. F. (2021). COVIDetection-Net: A tailored COVID-19 detection from chest radiography images using deep learning. *Optik*, *231*, 166405.

Guo, K., Sui, L., Qiu, J., Yu, J., Wang, J., Yao, S., Han, S., & Wang, H. (2017). Angel-eye: A complete design flow for mapping CNN onto embedded FPGA. *IEEE transactions on computer-aided design of integrated circuits and systems*, *37*(1), 35-47.

Hashim, S. A. M., bin Abd Malek, N. M. S., Kanagaraj, M., & bin Yasak, M. F. (2025). Automatic Weld Bead Discontinuities Detection based on Dye Penetrant Test by Deep Learning. *Malaysian Journal of Information and Communication Technology (MyJICT)*, 1-9.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.

Manakitsa, N., Maraslidis, G. S., Moysis, L., & Fragulis, G. F. (2024). A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. *Technologies*, *12*(2), 15.

McKnight, S. (2024). *Deep learning for ultrasonic non-destructive evaluation of aerospace composites*. Doctoral dissertation, University of Strathclyde. STAX.

Medak, D., Posilović, L., Subašić, M., Budimir, M., & Lončarić, S. (2022). DefectDet: A deep learning architecture for detection of defects with extreme aspect ratios in ultrasonic images. *Neurocomputing*, *473*, 107-115.

Obaid, T., Abu-Naser, S. S., Abumandil, M. S., Mahmoud, A. Y., & Ali, A. A. A. (2022, November). Age and Gender Classification from Retinal Fundus Using Deep Learning. In *The International Conference of Advanced Computing and Informatics* (pp. 171-180). Cham: Springer International Publishing.

Posilović, L., Medak, D., Milković, F., Subašić, M., Budimir, M., & Lončarić, S. (2022b). Deep learning-based anomaly detection from ultrasonic images. *Ultrasonics*, *124*, 106737.

Posilović, L., Medak, D., Subašić, M., Budimir, M., & Lončarić, S. (2022a). Generating ultrasonic images indistinguishable from real images using Generative Adversarial Networks. *Ultrasonics*, *119*, 106610.

Rathbun, K. M., Patel, A. N., Jackowski, J. R., Parrish, M. T., Hatfield, R. M., & Powell, T. E. (2023). Incorporating ultrasound training into undergraduate medical education in a faculty-limited setting. *BMC Medical Education*, *23*(1), 263.

Rumsey, D. J. (2019). *Statistics essentials for dummies*. John Wiley & Sons.

Shipway, N. J. (2019). *Automated defect detection for fluorescent penetrant inspection using machine learning*. Doctoral dissertation, Imperial College London.

Taheri, H., & Hassen, A. A. (2019). Nondestructive ultrasonic inspection of composite materials: A comparative advantage of phased array ultrasonic. *Applied Sciences*, *9*(8), 1628.

Wang, J., Khan, M. A., Wang, S., & Zhang, Y. (2023). SNSVM: SqueezeNet-guided SVM for breast cancer diagnosis. *Computers, materials & continua*, *76*(2), 2201.

Wen, L., Li, X., & Gao, L. (2020). A new reinforcement learning based learning rate scheduler for convolutional neural network in fault classification. *IEEE Transactions on Industrial Electronics*, *68*(12), 12890-12900.

Ye, J., Ito, S., & Toyama, N. (2018). Computerized ultrasonic imaging inspection: From shallow to deep learning. *Sensors*, *18*(11), 3820.

Zhu, X., Guo, Z., Zhou, Q., Zhu, C., Liu, T., & Wang, B. (2024). Damage identification of wind turbine blades based on deep learning and ultrasonic testing. *Nondestructive Testing and Evaluation*, 1-26.