Deep Learning in Manufacturing: A Focus on Welding Defect Classification with CNNs

Tin Chang Ting[1], Hameedur Rahman[2], Tiong Hoo Lim[3], Chin Hong Wong[4,5], Chun Kit Ang[1], Mohamed Khan Afthab Ahamed Khan[1], Sew Sun Tiang[1*], Wei Hong Lim[1*]

[1]*Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia*
[2]*Faculty of Computing and Artificial Intelligence, Air University, Islamabad Capital Territory 44000, Pakistan*
[3]*Faculty of Engineering, Universiti Teknologi Brunei, Bandar Seri Begawan 1410, Brunei Darussalam*
[4]*Maynooth International Engineering College, Maynooth University, Maynooth, Co Kildare, Ireland*
[5]*Maynooth International Engineering College, Fuzhou University, Fujian, 350116, China*
Email: 1002058096@ucsiuniversity.edu.my, rhameedur@gmail.com, lim.tiong.hoo@utb.edu.bn, chinhong.wong@mu.ie, angck@ucsiuniversity.edu.my, mohamedkhan@ucsiuniversity.edu.my, tiangss@ucsiuniversity.edu.my, limwh@ucsiuniversity.edu.my

Abstract

Welding is integral to modern manufacturing, yet the complex process often leads to defects, impacting the quality of the final product. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have shown remarkable results in applications like defect recognition. This study evaluated AlexNet, ResNet-18, ResNet-50, ResNet-101, MobileNet-v2, ShuffleNet, and SqueezeNet for their effectiveness in identifying welding defects, using accuracy, precision, sensitivity, specificity, and F-score as metrics. The dataset covered defects like cracks, lack of penetration, porosity, and a no-defect class. Our analysis shows that most of these architectures deliver promising results in accuracy, sensitivity, specificity, precision, and F1-score, highlighting their potential in defect recognition.

Keywords: Convolutional neural network, Classification, Deep learning, Welding defects

1. Introduction

Welding is a crucial industrial process integral to various production sectors, including high-performance industries like aerospace, automotive, marine, and power generation. This complex process involves multiple parameters that directly impact the quality of the weld joint. Factors such as welding procedures, methods, environmental conditions, and the operator's skill level can lead to various defects during the welding of pipelines and pressure vessels [1]. Common defects include blowholes, cracks, incomplete fusion, incomplete penetration, slag inclusions, and undercutting, which significantly compromise the sealing and strength of the products. In addition to welding parameters, unforeseen events in the manufacturing process can also cause weld defects. Therefore, rigorous welding quality inspection and testing are essential during manufacturing, especially for products like pipelines and pressure vessels. This is critical to identify the root causes of welding defects and implement targeted corrective measures to ensure product quality and safety [1].

Considerable research has been devoted to addressing the challenges of weld defect detection and identification. These defects are typically detected using non-destructive testing (NDT) techniques, favored for their non-invasive interaction with specimens. Common NDT methods for weld defect detection fall into various categories, including visual or manual inspection, radiographic testing with ionizing radiation sources (such as gamma rays or X-rays), eddy current testing, ultrasonic testing, and dye penetrant testing [2], [3]. However, each of these techniques has inherent limitations. For example, eddy current testing is only applicable to metallic specimens, while X-ray radiographic testing poses potential health risks due to prolonged exposure to radiation. Additionally, the majority of current weld surface defect recognition relies on manual inspection by NDT experts analyzing these radiographic images. This manual process of interpreting and evaluating images can be complex, subjective, inconsistent, time-consuming, labor-intensive, and prone to errors, particularly when distinguishing between defects with similar features [4]. Consequently, developing an automated inspection solution that offers more efficient and accurate recognition of weld surface defects is essential to overcome the drawbacks of manual inspections.

The adoption of automated machine vision systems, integrating deep learning techniques, offers a promising solution for weld surface image classification challenges. Convolutional Neural Networks (CNNs), renowned for their ability to learn nonlinear relationships between inputs and outputs, mimic the human brain's learning process and have been effectively applied in various real-world scenarios. These applications include signal/image classification [5], [6], [7], cybersecurity [8], [9], medical diagnosis [10], [11], [12], fault detection [13], [14], [15] and prediction [16], [17]. Encouraged by deep learning's success, numerous network architectures have been developed, such as AlexNet [18], ResNet-18 [19], ResNet-50 [19], ResNet-101 [19], MobileNet-v2 [20], ShuffleNet [21], and SqueezeNet [22]. Transfer learning, a method for adapting these existing CNN architectures to new tasks, has gained popularity. It utilizes smaller datasets and reduces training time, making it an efficient approach for training CNNs in new application domains.

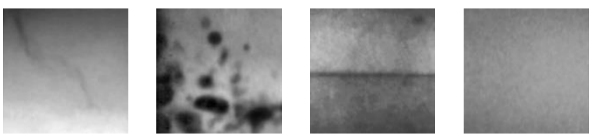
While previous studies [4], [23], [24], [25], [26], [27] have explored welding defect classification using deep learning, the optimal selection of CNN architectures remains relatively unexplored, leaving many architectures yet to be assessed. Addressing this gap, our paper presents an extensive study evaluating the performance of seven popular CNN architectures: AlexNet, ResNet-18, ResNet-50, ResNet-101, MobileNet-v2, ShuffleNet, and SqueezeNet, in classifying weld defects from digital radiographic images. Our aim is to objectively identify the most effective CNN architecture using various performance metrics. This study could lead to replacing manual inspection methods with a more accurate automated weld defect classification system, potentially reducing production costs and increasing throughput.

2. Methodology

2.1. *Data acquisition and preprocessing*

This study focuses on training and evaluating a deep learning model using the RIAWELC welding defect dataset [28]. Unlike other commonly used datasets such as GDXRay [29] and WDXI [30], RIAWELC offers a larger, open-source dataset, which is beneficial for training CNNs without the risk of overfitting.

The RIAWELC dataset comprises 24,407 radiographic images, each of size 224×224 pixels, categorized into four classes of welding defects: lack of penetration (LP), cracks (CR), porosity (PO), and no defect (ND). A representative image for each defect type is displayed in Fig. 1, and Table 1 details the distribution of images across these classes. The substantial size of RIAWELC dataset supports the development of automated methods for identifying and classifying welding defects, crucial for reliable inspection and quality control.



(a) (b) (c) (d)

Fig.1 Sample of radiographic images for each welding defect class from RIAWELC dataset: (a) LP, (b) CR, (c) PO and (d) ND.

Table 1. Data distribution of RIAWELC dataset for each welding defect class.

Defect Types	CR	PO	LP	ND
No. of Image	7,635	6,320	4,452	6,000

To facilitate deep learning model training, the RIAWELC dataset is divided into three subsets: 70% for training, 10% for validation, and 20% for testing. Additionally, all radiographic images from the dataset are resized to match the input requirements of the selected pretrained network, as outlined in Table 2, ensuring compatibility with the network architectures.

Table 2. Summary of seven pretrained networks

Pretrained Networks	Depth	Size (MB)	Parameter (Millions)	Input Size
AlexNet	8	227	61.0	224×224
ResNet-18	18	44	11.7	224×224
ResNet-50	50	96	25.6	224×224
ResNet-101	101	167	44.6	224×224
MobileNet-v2	53	13	3.5	224×224
ShuffleNet	50	5.4	1.4	224×224
SqueezeNet	18	5.2	1.24	227×227

2.2. *Transfer learning of pretrained networks*

Training CNNs from scratch for specific tasks presents challenges due to significant resource requirements, such as training time, infrastructure, and input datasets. Transfer learning offers a viable solution by transferring knowledge from one or more source domains to a different target domain. In this study, we utilize transfer learning to extract the learnable parameters from selected pretrained network architectures (namely, AlexNet, ResNet-18, ResNet-50, ResNet-101, MobileNet-v2, ShuffleNet, and SqueezeNet) for the welding defect radiographic images classification tasks.

Specifically, the last three layers of these pretrained networks are replaced with a new fully-connected layer, a SoftMax layer, and a classification output layer. Additionally, the original output layers are substituted with new output layers tailored to four classes: lack of penetration (LP), cracks (CR), porosity (PO), and no defect (ND). These modified networks are then trained using the RIAWELC dataset, as detailed in Table 1.

2.3. *Hyperparameter tuning*

Stochastic Gradient Descent (SGD) is a popular optimizer in deep learning training, valued for its ability to effectively balance accuracy and efficiency. In this study, SGD is employed to train selected pretrained network architectures with the RIAWELC dataset, aiming to minimize the cross-entropy loss function. However, the performance and convergence of SGD are influenced by various training hyperparameters, including momentum, initial learning rate, L2 regularization, maximum epoch, and batch size.

Specifically, the initial learning rate dictates the step sizes in the parameter update process. L2 regularization helps to prevent network overfitting by adding a penalty term to the loss function. The maximum epoch limits the number of iterations to avoid overfitting, and the batch size determines the number of samples used per iteration, impacting the stability, speed, and memory usage of SGD.

To facilitate a fair comparison in performance evaluations, the hyperparameters for all selected pretrained networks are standardized. This includes setting the momentum to 0.9, the initial learning rate to $\mathbf{2 \times 10^{-3}}$, L2 regularization to 0.5, the maximum epoch to 5, and the batch size to 32.

2.4. *Performance metrics*

The overall performance of all pretrained CNN networks in classifying welding defect radiographic images is evaluated using five key metrics: accuracy, sensitivity, specificity, precision and F1 score. These metrics are derived from the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes produced during the testing phase of the CNN architectures. The mathematical formulations for each performance metric employed in this study are as follows.

Accuracy, indicating the overall correctness of the CNN architecture's predictions, is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

Sensitivity, measuring the CNN architecture's ability to correctly identify positive results, is computed as:

$$Sensitivity = \frac{TP}{TP + FN} \qquad (2)$$

Specificity, gauging the CNN architecture's capacity to accurately identify negative results, is determined as:

$$Specificity = \frac{TN}{TN + FP} \qquad (3)$$

Precision, assessing the accuracy of positive predictions among all positive cases identified, is calculated as:

$$Precision = \frac{TP}{TP + FP} \qquad (4)$$

F1 Score, the harmonic mean of precision and sensitivity, is defined as:

$$F1\ Score = 2 \times \left(\frac{Sensitivity \times Precision}{Sensitivity + Precision}\right) \qquad (5)$$

The abovementioned five performance metrics provide a comprehensive assessment of the CNN architectures' capabilities in distinguishing different classes of welding defects.

3. Results and Discussions

The performance of all selected pretrained networks, including AlexNet, ResNet-18, ResNet-50, ResNet-101, MobileNet-v2, ShuffleNet, and SqueezeNet, in classifying radiographic images of welding defects is comprehensively evaluated. To facilitate a quantitative comparison, each CNN architecture's effectiveness in the welding defect classification task is assessed using the five performance metrics previously mentioned, namely accuracy (Acc.), sensitivity (Sens.), specificity (Spec.), precision (Prec.) and F1 score. The simulation results are detailed in Table 3, where the highest values attained by each CNN architecture for these metrics are highlighted in boldface.

Table 3. Quantitative performance evaluation of all selected CNN architectures for welding defect classification task

CNN architecture	Acc. (%)	Sens. (%)	Spec. (%)	Prec. (%)	F1 (%)
AlexNet	71.43	71.43	90.48	80.21	70.92
ResNet-18	78.57	78.57	92.86	80.56	78.92
ResNet-50	82.14	82.14	94.05	83.93	82.49
ResNet-101	**85.71**	**85.71**	**95.24**	**87.50**	**86.06**
MobileNet-v2	71.43	71.43	90.48	69.00	69.06
ShuffleNet	75.00	75.00	91.67	80.63	75.98
SqueezeNet	71.43	71.43	90.48	75.00	70.92

Table 3 reveals a significant observation across all CNN architectures: the identical values of accuracy and sensitivity when classifying the RIAWELC dataset. This uniformity implies that the models' ability to correctly predict positive instances (sensitivity) significantly impacts their overall accuracy. Among these, ResNet-101 (85.71%) and ResNet-50 (82.14%) stand out with the best and second-best performance in both accuracy and sensitivity, highlighting their superior detection capabilities for welding defects in radiographic images. Conversely, architectures with lower complexity, such as AlexNet, MobileNet-v2, and SqueezeNet, demonstrate less effectiveness, each recording 71.43% in both accuracy and sensitivity.

Moreover, Table 3 indicates that all selected CNN architectures consistently achieve higher specificity and precision compared to their accuracy and sensitivity. This suggests they are more effective in correctly identifying true negatives (specificity) and ensuring accurate positive predictions (precision). Specifically, ResNet-101 exhibits the highest specificity (95.24%), while AlexNet shows the lowest (90.48%), indicating a broad capability across these models to identify non-defective instances accurately. Precision, however, varies more across the models, with ResNet-101 leading at 87.50%, suggesting its higher accuracy in classifying weld defects. Conversely, MobileNet-v2, despite similar accuracy and sensitivity to other models, records the lowest precision (69.00%), implying a higher rate of false positives.

The F1 score, reflecting the harmonic mean of precision and sensitivity, offers a more balanced evaluation of model performance. ResNet-101 and ResNet-50 achieve the highest F1 scores (86.06% and 82.49%, respectively), confirming their robust overall performance in classifying the radiographic images of welding defects. In contrast, AlexNet, MobileNet-v2, and SqueezeNet exhibit lower F1 scores (70.92%, 69.06%, and 70.92%, respectively), indicating a compromise between sensitivity and precision in these models.

Analysis of the simulation results in Table 3 reveals that more complex CNN architectures, such as ResNet-101 and ResNet-50, consistently outperform simpler ones like AlexNet and SqueezeNet across all performance metrics. This superior performance of advanced models like ResNet-101 and ResNet-50 highlights the significance of choosing the right model for complex image classification tasks, such as welding defect classification. It suggests that deeper networks are more adept at extracting essential features necessary for accurately identifying welding defects. The enhanced capability of these complex CNN models to discern intricate patterns in data meets the expectations associated with deeper network architectures.

However, it is important to note that while these sophisticated CNN architectures offer improved classification accuracy, they also demand greater computational resources. In contrast, simpler CNN models, despite being less precise, are more computationally efficient and may be preferable in scenarios with limited resources. In practical applications, factors like available computational power, latency considerations, and the criticality of accurately detecting defects should guide the selection of CNN architectures. Balancing these considerations is key to effectively deploying CNNs in real-world tasks.

4. Conclusions

This paper introduces a deep learning-based machine vision inspection algorithm that combines pretrained CNN architectures and transfer learning for classifying four types of welding defects—lack of penetration, cracks, porosity, and no defect—using radiographic images from the RIAWELC dataset. The objective is to thoroughly analyze the efficacy of seven different pretrained CNN architectures: AlexNet, ResNet-18, ResNet-50, ResNet-101, MobileNet-v2, ShuffleNet, and SqueezeNet, in addressing welding defect classification tasks. This analysis employs various performance metrics, including accuracy, sensitivity, specificity, precision, and F1 score. Simulation studies reveal that these pretrained networks exhibit diverse performance levels in defect classification, with ResNet-101 emerging as the most effective, achieving 85.71% accuracy, 85.71% sensitivity, 95.24% specificity, 87.50% precision, and an F1 score of 86.06%. In contrast, MobileNet-v2 shows the least effectiveness, with the lowest scores across all metrics: 71.43% accuracy and sensitivity, 90.48% specificity, 69.00% precision, and an F1 score of 69.06%. These results offer valuable insights into the applicability of pretrained deep learning networks for welding defect classification, providing a significant benefit to manufacturing companies seeking to enhance their quality control processes. In this study, the hyperparameter settings for all pretrained networks are manually configured. It is anticipated that their classification performance could be further improved by employing advanced metaheuristic search algorithms to optimize these hyperparameter settings, thereby refining the training process of the pretrained networks.

Acknowledgements

This work was supported by UCSI University's Research Excellence & Innovation Grant (REIG) with project code of REIG-FETBE-2022/038 and Billion Prima Sdn. Bhd.'s Industry Research Grant with project code of IND-FETBE-2023/006.

References

1. Q. Feng, R. Li, B. Nie, S. Liu, L. Zhao, and H. Zhang, Literature review: Theory and application of in-line inspection technologies for oil and gas pipeline girth weld defection, Sensors 17(1), 2017, pp. 50.
2. L. Yin et al., A novel feature extraction method of eddy current testing for defect detection based on machine learning, NDT & E International 107, 2019, pp. 102108.
3. C. Pei, D. Yi, T. Liu, X. Kou and Z. Chen, Fully noncontact measurement of inner cracks in thick specimen with fiber-phased-array laser ultrasonic technique, NDT & E International 113, 2020, pp. 102273.
4. K. Ding, Z. Niu, J. Hui, X. Zhou, and F. T. S. Chan, A weld surface defect recognition method based on improved MobileNetV2 algorithm, Mathematics 10(19), 2022, pp. 3678.
5. B. Jdid, W. H. Lim, I. Dayoub, K. Hassan and M. R. B. M. Juhari, Robust automatic modulation recognition through joint contribution of hand-crafted and contextual features, IEEE Access 9, 2021, pp. 104530-104546.
6. K. M. Ang et al, Optimal design of convolutional neural network architectures using teaching-learning-based optimization in image classification, Symmetry 14(11), 2022, pp. 2323.
7. K. M. Ang et al., An innovative approach for automated convolutional neural network design for image classification, Mathematics 11(19), 2023, pp. 4115.
8. T. Berghout, M. Benbouzid and Y. Amirat, Towards resilient and secure smart grids against PMU adversarial attacks: A deep learning-based robust data engineering approach, Electronics 12(12), 2023, pp. 2554.
9. T. Berghout and M. Benbouzid, EL-NAHL: Exploring labels autoencoding in augmented hidden layers of feedforward neural networks for cybersecurity in smart grids, Reliability Engineering & System Safety 226, 2022, pp. 108680.
10. L. S. Chow, G. S. Tang, M. I. Solihin, N. M. Gowdh, N. Ramli and K. Rahmat, Quantitative and qualitative analysis of 18 deep convolutional neural network (CNN) models with transfer learning to diagnose COVID-19 on chest X-ray (CXR) images, SN Computer Science 4(2), 2023, pp. 141.
11. A. Qayyum et al., Hybrid 3D-ResNet Deep Learning Model for Automatic Segmentation of Thoracic Organs at Risk in CT Images, 2020 International Conference on

Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russia, 2020, pp. 1-5.
12. H. Rahman et al. A systematic literature review of 3D deep learning techniques in computed tomography reconstruction. Tomography 9(6), 2023, pp. 2158-2189.
13. M. Alrifaey, W. H. Lim and C. K. Ang, A novel deep learning framework based RNN-SAE for fault detection of electrical gas generator, IEEE Access, 9, 2021, pp. 21433-21442.
14. M. Alrifaey et al., Hybrid deep learning model for fault detection and classification of grid-connected photovoltaic system, IEEE Access 10, 2022, pp. 13852-13869.
15. T. Berghout et al. Federated learning for condition monitoring of industrial processes: A review on fault diagnosis method, challenges and prospect. Electronics 12(1), pp. 158.
16. T. Berghout, M. Benbouzid, Y. Amirat and G. Yao, Lithium-ion battery state of health prediction with a robust collaborative augmented hidden layer feedforward neural network approach. IEEE Transactions on Transportation Electrification 9(3), 2023, pp. 4492-4502.
17. A. A. Abdelhamid et al. Deep learning with dipper throated optimization algorithm for energy consumption forecasting in smart households. Energies 15(23), 2022, pp. 9125.
18. A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Communication of the ACM 60(6), 2017, pp. 84-90.
19. K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778.
20. M. Sandler et al, MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 4510-4520.
21. X. Zhang, X. Zhou, M. Lin and J. Sun, ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 6848-6856.
22. F.Iandola et al. SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and <IMB Model Size, Preprint, submitted November 4, 2016. https://arxiv.org/pdf/1602.07360
23. S. Kumaresan, K. S. Jai Aultrin, S. S. Kumar and M. Dev Anand, Deep learning-based weld defect classification using VGG16 transfer learning adaptive fine tuning. International Journal on Interactive Design and Manufacturing 17, 2023, pp. 2999-3010.
24. W. Hou, Y. Wei, J. Guo, Y. Jin and C. Zhu, Automatic detection of welding defects using deep neural network, Journal of Physics: Conference Series 933, 2017, pp. 012006.
25. W. Du, H. Shen, J. Fu, G. Zhang and Q. He, Approaches for improvement of the X-ray image defect detection of automobile casting aluminium parts based on deep learning, NDT & E International 107, 2019, pp. 102144.
26. S. Kumaresan, K. S. Jai Aultrin, S. S. Kumar and M. Dev Anand, Transfer learning with CNN for classification of weld defect, IEEE Access 9, 2021, pp. 95097-951078.
27. D. Say, S. Zidi, S. M. Qaisar and M. Krichen, Automated categorization of multiclass welding defects using the X-ray image augmentation and convolutional neural network, Sensors 23(14), 2023, pp. 6422.
28. S. Perri, F. Spagnolo, F. Frustaci and P. Corsonello, Welding defects classification through a convolutional neural network, Manufacturing Letters 35(42), 2022, pp. 29-32.
29. D. Mery et al, GDXray: The database of X-ray images for nondestructive testing, Journal of Nondestructive Evaluation 34(42), 2015, pp. 42.
30. W. Guo, H. Qu and L. Liang, WDXI: The Dataset of X-ray Image for Weld Defects, 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 2018, pp. 1051-1055.

Authors Introduction

Mr. Tin Chang Ting

He is currently pursuing the Bachelor of Mechatronics Engineering with Honours in Faculty of Engineering, Technology and Built Environment, UCSI University, Malaysia. His research interests are machine learning, deep learning, and optimization algorithm.

Dr. Hameedur Rahman

He is an Associate Professor in Faculty of Computing and AI at AIR University in Pakistan. He received his PhD in Computer Science from Universiti Kebangsaan Malaysia in 2018. His research interests are Viral/Augmented Rality, Image Processing, Data Mining, Artificial Intelligence, Natural Language Processing and CyberSecurity.

Dr. Tiong Hoo Lim

He is a Senior Assistant Professor and Director of Planning Development Office at Universiti Teknologi Brunei in Brunie Darussalam. He received his PhD in Computer Science from University of York, United Kingdom. His research interests are artificial intelligence, advanced technology, competitive and forecasting analysis.

Dr. Chin Hong Wong

He is a Lecturer in Maynooth International Engineering College at Fuzhou University in China. He received his PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2017. His research interests are Energy harvesting and control system.

Dr. Chun Kit Ang

He is the Dean and Associate Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Mechanical and Manufacturing Engineering from Universiti Putra Malaysia in 2014. His research interests are artificial intelligence, soft computing, robotics and mechatronics.

Dr. Mohamed Khan Afthab Ahmed Khan

He is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. His research interests are artificial intelligence, robotics, control, medical rehabilitation and power electroncs.

Dr. Sew Sun Tiang

She is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. She received her PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2014. Her research interests are optimization and antenna design.

Dr. Wei Hong Lim

He is an Associate Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Computational Intelligence from Universiti Sains Malaysia in 2014. His research interests are optimization and artificial intelligence.