EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Accurate ultimate tensile strength classification in friction stir welding of symmetric AA5052 weld seams using ensemble deep learning model

Chakat Chueadee[1)], Rungwasun Kraiklang*[1)], Surasak Matitopanum[1)] and Sarayut Gonwirat[2)]

[1)]Department of Industrial Engineering, Faculty of Engineering and Technology, Rajamangala University of Technology Isan, Nakhon Ratchasima 30000, Thailand
[2)]Department of Computer Engineering and Automation, Faculty of Engineering and Industrial Technology, Kalasin University, Kalasin 46000, Thailand

Received 9 October 2023
Revised 28 December 2023
Accepted 16 January 2024

Abstract

This research introduces a comprehensive classification and analysis system tailored for accurately determining the ultimate tensile strength (UTS) of weld seams. Traditional UTS assessment methods typically depend on destructive testing, which tends to be both lengthy and expensive, limiting their continuous application in quality control of welding procedures. This study leverages deep learning techniques, segmenting the dataset into subsets for training and validation in both multi-class and binary classification scenarios. The model devised in this study incorporates cutting-edge methodologies including geometric enhancement, U-Net based image segmentation, an image augmentation of diverse deep learning models, and decision fusion tactics. A significant aspect of this research was the success of Experiment 33, which skillfully combined various methodologies, resulting in outstanding performance. This experiment demonstrated exceptional accuracy in multiclass classification, alongside impressive outcomes in binary classification, achieving a high accuracy rate of 97.4% and an F1 score of 96.5%. This level of accuracy is indicative of the average performance across all models that incorporated the He-UWA for decision fusion strategy. It encompasses the efficacy of all models using He-UWA, with or without image segmentation. These findings underscore the effectiveness of our proposed model in accurately classifying UTS in friction stir welding. This represents a crucial advancement in assessing the quality of welding processes and provides a solid foundation for future investigations in this area.

Keywords: Deep learning, Weld seam strength, UTS Classification, Image segmentation, Decision fusion

1. Introduction

Friction Stir Welding (FSW) has garnered significant attention as a solid-state welding technique renowned for producing high-quality welds in diverse metallic alloys [1, 2]. In particular, FSW has emerged as a promising method for joining aluminum alloys, offering distinct advantages such as improved mechanical properties, reduced distortion, and enhanced corrosion resistance [1, 3]. Among these alloys, AA5052, a symmetric alloy with wide-ranging applications in automotive, aerospace, and marine industries, has attracted considerable interest due to its favorable combination of strength, formability, and weldability [4]. The ultimate tensile strength (UTS) of a weld seam plays a critical role in evaluating the structural integrity and performance of the joined components. Accurate and efficient determination of UTS is paramount to ensure reliability and safety [5]. However, conventional UTS evaluation methods often rely on destructive testing, which is time-consuming, costly, and impractical for real-time quality control during welding processes. In recent years, deep learning techniques have demonstrated great potential across various fields by effectively extracting complex patterns and making accurate predictions from extensive datasets. Leveraging the power of deep learning, researchers have successfully developed models for a wide range of classification tasks, including applications in materials science. Despite these advancements, the classification of UTS for weld seams in AA5052 alloy using deep learning techniques remains largely unexplored.

Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have revolutionized image classification tasks [6-8], showcasing their remarkable capabilities in various domains. However, accurately classifying the ultimate tensile strength (UTS) of weld seams in the context of friction stir welding (FSW) presents unique challenges, including architectural limitations and the intrinsic complexity and variability of weld seam images. To overcome these challenges and significantly enhance the accuracy of UTS classification in FSW weld seams, this research introduces an innovative approach that harnesses the power of ensemble deep learning [9-11]. The proposed ensemble model combines multiple CNN architectures, including EfficientNetV2B3, EfficientNetV2-Small, ShuffleNetV2, SqueezeNetV2, and MobileNetV3-Small, to leverage diverse representations and improve the overall classification performance. By incorporating a variety of CNN architectures, the model can capture different aspects and features of weld seam images, leading to more robust and accurate UTS classification. Furthermore, advanced techniques are integrated into the proposed model to further enhance its capabilities. The U-Net architecture is employed for image segmentation, enabling precise localization of weld seam regions and extraction of relevant features. Geographic augmentation is applied for image augmentation [12-

*Corresponding author.
Email address: rungwasun@rmuti.ac.th
doi: 10.14456/easr.2024.21

14], introducing variations in lighting conditions, angles, and other factors to improve the model's ability to generalize. Additionally, unweight average ensemble (UWA) are utilized as a decision fusion strategy, effectively combining predictions from multiple CNNs to arrive at a final UTS classification [15].

The application of deep learning techniques in weld quality assessment has been the subject of prior research. A comprehensive review by Smith et al. [16] highlights the potential of deep learning approaches and emphasizes the significance of ensemble learning in achieving improved classification accuracy. Furthermore, Chiaranai et al. [9] conducted a survey focusing on image segmentation techniques in weld seam analysis, shedding light on the challenges and opportunities in accurately delineating weld seam regions. Despite the considerable amount of existing research, there is still a noticeable research gap concerning the development of an accurate ultimate tensile strength (UTS) classification model for weld seams.

This study aims to develop a comprehensive ensemble deep learning model to address the limitations of individual models and improve the accuracy of ultimate tensile strength (UTS) classification for friction stir welding (FSW) weld seams. The effectiveness of the proposed model is evaluated by comparing it with the conventional testing approach. The results highlight the significant potential of the image-based method in providing reliable and consistent evaluations of weld seam strength. This positions it as a promising alternative to traditional UTS testing techniques. Furthermore, the non-destructive nature of the proposed model allows for repeated assessments without compromising the integrity of the weld seam, enabling its application in quality control and real-time monitoring during welding processes.

To achieve the objective of accurate UTS classification in FSW weld seams, this study addresses the following research questions:

1. What is the effect of geographic augmentation techniques on the performance of the ensemble deep learning model for UTS classification?

2. How does image segmentation using the U-Net architecture enhance the accuracy of weld seam classification for UTS assessment?

3. How does the combination of EfficientNetV2B3, EfficientNetV2-Small, ShuffleNetV2, SqueezeNetV2, and MobileNetV3-Small architectures in a heterogeneous ensemble deep learning model the accuracy of UTS classification for FSW weld seams?

4. To what extent can the application of unweight average as a decision fusion strategy enhance the classification accuracy of the ensemble deep learning model for UTS assessment in FSW weld seams?

The remaining sections of the study are organized as follows. Section 2 provides an extensive discussion on related research concerning sentiment feature engineering and classification algorithms. In Section 3, two distinct feature set strategies are described, accompanied by a comprehensive framework for ensemble classification. The presentation and analysis of the experimental results are presented in Section 4. Finally, in Section 5, we present the concluding remarks and propose potential research directions for future investigations.

2. Literature review

Friction stir welding (FSW) is a widely employed solid-state joining process extensively utilized in various industries for the purpose of joining materials, particularly aluminum and its alloys. FSW offers numerous advantages compared to conventional welding techniques, including improved mechanical properties, reduced distortion, and enhanced weld quality. Multiple factors, such as rotational speed, significantly influence the quality and properties of FSW welds [17]. As an example, Lombard et al. [18] conducted a study that showcased the influence of rotational speed on defect formation within the weld of aluminum alloy 5083-H321. Their research revealed a robust correlation between friction feed and tensile strength, emphasizing the importance of rotational speed as a controlling factor. Similarly, Khalafe et al. [19] identified several key parameters crucial to the weldability of aluminum joints using FSW, including welding speed, rotational speed, plunge depth, spindle torque, shoulder design, base material, pin profile, and tool type. The careful selection of these parameters is of utmost importance as they directly influence the quality and performance of the joint. Furthermore, the physical parameters of FSW joints directly correspond to the microstructural transformations that occur during the welding process [20]. Therefore, in order to optimize the FSW process, a comprehensive understanding of the impact of welding parameters on joint strength is essential. The aim of this study was to enhance the overall joint strength and mechanical performance of FSW.

Image classification is a crucial task in computer vision, aiming to automatically assign images to specific classes or categories. With the advent of deep learning, particularly convolutional neural networks (CNNs), significant advancements have been made in this domain. Deep learning models have revolutionized image classification by effectively addressing the challenges associated with manual feature extraction in traditional machine learning methods. In previous research studies, various techniques in deep learning and welding quality assessment have been introduced to address the limitations and challenges associated with ultimate tensile strength (UTS) classification for friction stir welding (FSW) welds. Wang et al. [21] conducted a review on ensemble deep learning techniques, highlighting their potential in improving classification model performance. The incorporation of ensemble deep learning in the proposed research aligns with the findings of Wang et al., as it aims to leverage the strengths of multiple convolutional neural network (CNN) architectures to enhance the accuracy of UTS classification in FSW weld seams. Yao et al. [22] introduced differentiable neural architecture search, enabling the automatic discovery of optimal neural network architectures. Although not directly applied in the current research, the work of Yao et al. showcases advancements in neural architecture search and the potential for further optimizing deep learning models for weld quality assessment. He et al. [23] proposed deep residual learning, which introduced residual connections to address training challenges in deep neural networks. The insights from this work have had a significant impact on the development of deep learning models, including the CNN architectures employed in the proposed ensemble model. Furthermore, Ioffe and Szegedy [24] introduced batch normalization, a technique that enhances the training of deep neural networks by normalizing intermediate activations. The integration of batch normalization, as demonstrated in the proposed ensemble model, aligns with the insights of Ioffe and Szegedy [24], aiming to expedite training and improve the overall performance of the UTS classification model. Several state-of-the-art CNN architectures have been developed for image classification tasks. Notable examples include EfficientNetV3 [25], EfficientNetV2-Small [26], ShuffleNetV2 [27, 28], Squeeze Net [29], and MobileNetv3 [30]. These models have demonstrated superior performance in terms of accuracy and have been widely adopted in various applications. These previous studies have provided valuable insights and methodologies that inform and contribute to the development and enhancement of the proposed ensemble deep learning model for accurate UTS classification in FSW weld seams.

The ensemble technique, recognized for its ability to amalgamate the outputs of multiple base classification models to yield a unified result, has emerged as a remarkably effective strategy for classification in diverse domains [31]. Notably, researchers have successfully employed ensemble techniques to improve classification accuracy in the context of topical text classification tasks [32, 33]. However, within the specific realm of sentiment categorization, there exists a noticeable scarcity of comparable publications, and a comprehensive evaluation of ensemble methods remains an unexplored area of research. Addressing this research gap, Khoder and Dornaika [34] conducted an extensive investigation wherein they developed multiple classifiers trained on diverse feature sets. The primary focus of their study centered around the development of these component classifiers, laying the foundation for a meticulous selection and combination process [34]. Furthermore, Xu et al. [35] have devised an algorithm that enables the automatic detection of flaws in radiographic images. Deep learning algorithms, in contrast to traditional detection methods, possess notable advantages such as robust generalization capabilities and automatic feature extraction. Consequently, deep learning algorithms have found successful application in the detection of welding defects [35], the output of a convolutional neural network (CNN) demonstrates superior generalization performance. The ensemble technique offers several advantages, including improved classification accuracy, enhanced model robustness, and increased resistance to noise and outliers. Moreover, ensemble methods enable the exploitation of complementary strengths across individual classifiers, leading to improved overall performance. In the context of deep learning, ensemble models are particularly advantageous in addressing complex classification problems and achieving state-of-the-art results.

3. Materials and methods

This section elucidates the research methodology employed to develop the deep learning model aimed at discriminating the ultimate tensile strength (UTS) of friction stir welded (FSW) seams. The construction process of the proposed model is illustrated in Figure 1, outlining the sequential steps undertaken in its development.

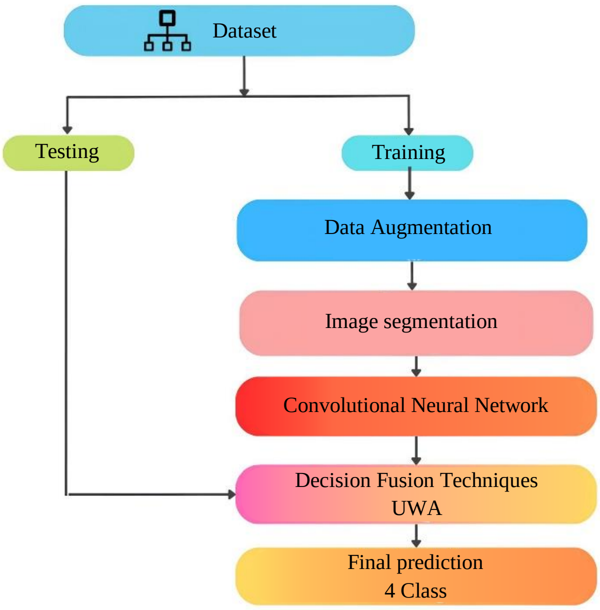


Figure 1 Framework of the method

The research methodology commences with the production and compilation of the dataset. This dataset is subsequently divided into two separate groups, specifically referred to as the training dataset and the testing dataset. The purpose of the training dataset is to facilitate the development of a comprehensive model, which encompasses four pivotal steps: (1) data augmentation, (2) image segmentation, (3) construction of a convolutional neural network (CNN) model, and (4) the implementation of decision fusion techniques. Following this, the proposed model is subjected to a thorough evaluation process, utilizing the designated test dataset. In the subsequent sections, each step is expounded upon in meticulous detail.

3.1 Weld seam dataset preparation

Dataset Collection: Compile a comprehensive dataset comprising images of weld seams from symmetric AA5052 samples used in Friction Stir Welding (FSW) which was obtained through laboratory Friction Stir Welding (FSW) experiments aluminum plates with a thickness of 6 mm were employed, and their chemical and mechanical properties are detailed in Table 1. The tool utilized in the experiment was machined from DC53 cold working tool steel and the tool was a cylindrically threaded pin profile featuring a diameter of 6 mm, a pitch distance of 1 mm, a pin height of 5.5 mm, and a shoulder diameter of 18 mm.

Table 1 Chemical and mechanical properties of the as received AA5052

Chemical composition w/%	Si	Fe	Cu	Mn	Mg	Cr	Zn
	0.08	0.23	0.002	0.005	2.24	0.20	0.007
Mechanical properties	Tensile strength (MPa)		Elastic modulus (GPa)			Hardness (HB)	
	208.99		70			47	

The experimental design used in this study was based on Full factorial's method. In order to evaluate this approach, several key Friction Stir Processing (FSP) parameters, including rotational speed, weld speed, plunge depth, and the tilt angle, were considered. Following Full factorial's experimental design principles, four parameters divided into three levels (3^4) were selected for the investigation. The process parameters and their corresponding levels are verified and presented in Table 2.

Table 2 Process parameters and levels

Parameters	**Code**	**Levels**		
		1	**2**	**3**
rotational speed (RPM)	A	800	1000	1200
weld speed (mm/min)	B	15	45	70
plunge depth (mm)	C	0.3	0.4	0.5
tilt angle (degrees)	D	0	0.15	0.3

In this study, a large dataset comprising 769 images from 81 unique specimens was used. Top-view images of these specimens were processed using various random techniques to simulate real-world scenarios and detect welding defects. Scaling the images, a vital pre-processing step in computer vision, greatly influences the effectiveness of the training models, with smaller images typically yielding better performance. The original digital camera images, with a resolution of 1275x890 pixels, were downscaled to 225x225 pixels. These specimens underwent thorough testing to evaluate their Ultimate Tensile Strength (UTS). Figure 2 in the study depicts the methodical procedure used for dataset preparation. The friction stir welding process was conducted in strict accordance with the experimental guidelines. Post-welding, tensile test specimens were precisely crafted using a water jet cutter, following the standards set in ASTM E8M-04, leading to the final assessment of their ultimate tensile strength. To ensure an unbiased and comparison, the training set was further divided, with 80% of the data allocated for training purposes, while the remaining 20% was set aside for validation. Deep learning models were applied to the dataset to accomplish both multi classification task. For the deep learning classification modeling, the dataset's output was categorized into four distinct groups based on the UTS values in relation to the reference value established by Mishra and Ma [17]. The dataset's specific details are provided in Table 3, and an outline of the classification of welding quality into separate categories can be found in Table 4.

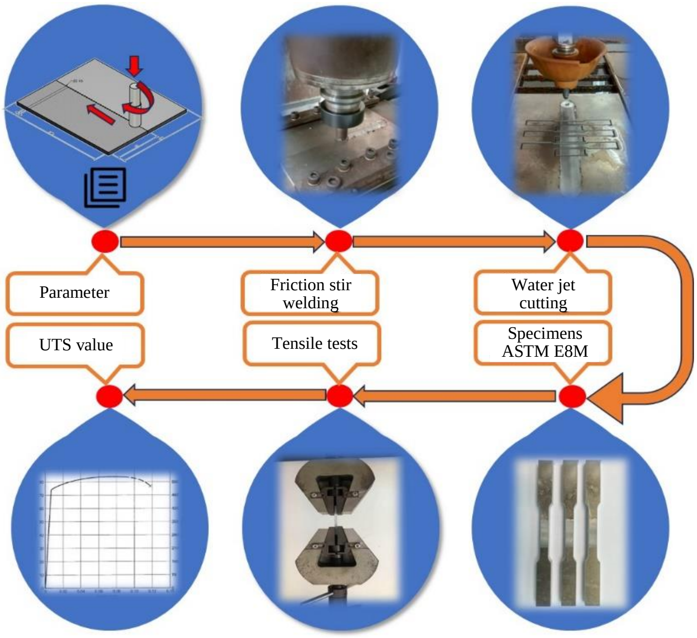


Figure 2 Experimental methodology approach

Table 3 Characteristics of our proposed dataset

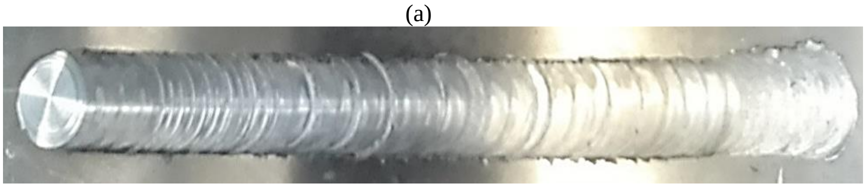
UTS Range	**4 Classes**			
	≤ 60%	**60–70 %**	**70-80%**	**>80%**
	≤ 125 MPa	**125 –146 MPa**	**146 -167 MPa**	**>170 MPa**
Train set	154	172	176	150
Test set	23	35	32	27
Total	177	207	208	177

Table 4 Classification of the welding quality of AA5053 mechanical properties.

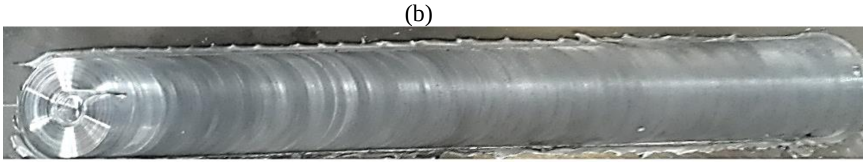
Class	**Quality**	**% of base material**	**UTS (MPa)**
1	Good	80% ≤	167 ≤ UTS
2	Fair	70% < UTS ≤ 80%	146 < UTS ≤ 167
3	Poor	60% < UTS ≤ 70%	125 < UTS ≤ 146
4	Unacceptable	≤ 60%	≤ 125

Before initiating data training and testing, the initial stage of pre-processing takes place. In this study, top-view alignment images obtained from specimens were utilized to simulate real-world scenarios, employing four distinct random image techniques [36]. This approach aimed to detect welding defects accurately. Furthermore, the concluding step of this preprocessing phase involves encoding labels in a one hot manner, as certain machine learning algorithms cannot directly handle data labeling. The assigned labels for the data include categories such as "Good weld seams," "Fair," "Poor," or "Unacceptable." Figure 3 serves as an illustration of the dataset's labeling. It is crucial to ensure that all input and output variables, including those used within the algorithm, possess numerical values. As a result, the labeled data was converted into numerical labels to facilitate comprehension and interpretation by the algorithm.

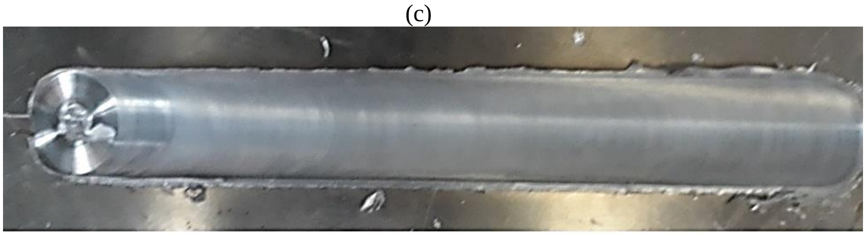
(a)



(b)



(c)



(d)

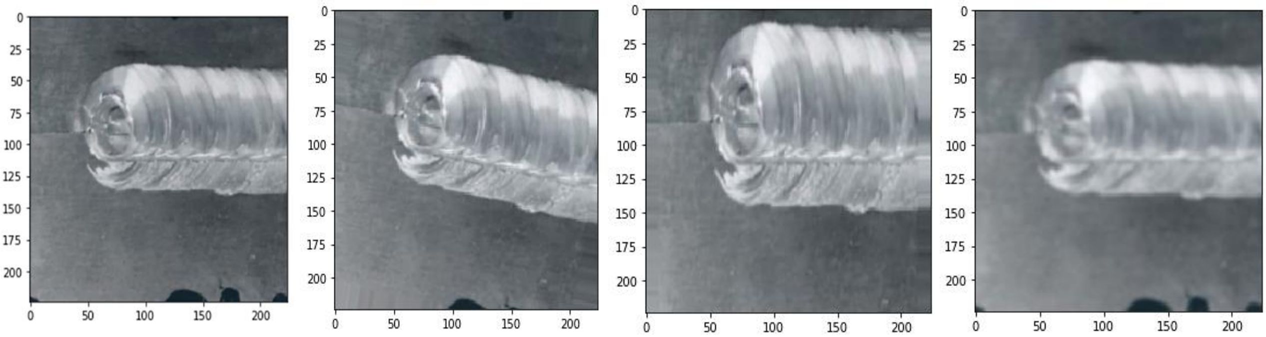
Figure 3 Example of (a) Unacceptable, (b) Poor, (c) Fair and (d) Good weld seams

3.2 Development of the ensemble deep learning model

Ensemble Deep Learning Model Development: Design and implement an ensemble deep learning model by utilizing state-of-the-art architectures and, unweight Average as a decision fusion strategy. The research findings presented here contribute significantly to the advancement of UTS assessment accuracy in FSW weld seams, thereby facilitating enhanced quality control and performance evaluation in welding processes. Each aspect of this research will be elaborated upon in the following sections.

3.2.1 Data augmentation

Enriching Geospatial Data: To enhance the training dataset, sophisticated geospatial augmentation methods, including rotation, translation, scaling, and flipping, are employed. These techniques introduce deliberate variations into the image data, augmenting the model's ability to effectively handle diverse spatial orientations and deformations observed in weld seam images. The outcomes of the image augmentation technique, exemplified in Figure 4, present instances of the original weld seam image (depicted in Figure 4a), juxtaposed with images subjected to geometric transformations such as rotation, zoom, and blur (represented as Figure 4b, 4c, and 4d, respectively).



(a) original (b) rotation (c) zoom (d) blur

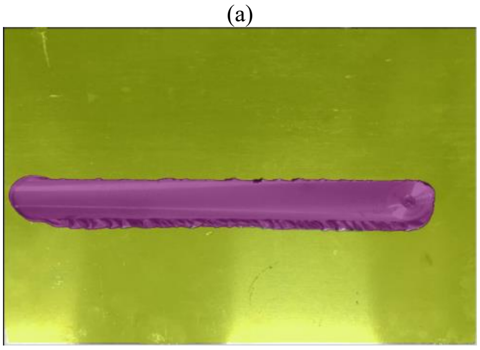
Figure 4 Image augmentation example

3.2.2 Image segmentation

Image segmentation plays a crucial role in various computer vision applications as it enables the precise delineation and extraction of specific objects or regions of interest from an image. In the context of weld seam analysis, accurate segmentation is essential for further analysis, defect detection, and quality assessment. The U-Net architecture is specifically designed for biomedical image segmentation but has been successfully adapted and applied to various domains, including industrial and engineering applications. It consists of an encoder-decoder network structure, where the encoder captures the contextual information and the decoder generates the segmentation map with detailed localization. To train the U-Net model, a carefully curated dataset of segmented images is utilized. The dataset contains examples of weld seam regions, along with corresponding ground truth masks that precisely outline these regions. During the training process, the model learns to map the input image to the corresponding segmented output, effectively acquiring the knowledge required to accurately segment weld seams.

Leveraging the U-Net architecture and training it with a segmented dataset enables precise weld seam segmentation using deep learning techniques. This advancement contributes to improved quality control, defect detection, and overall performance evaluation in welding processes by enhancing our understanding and analysis of weld seams. Figure 5 depicts an example of the segmentation technique that was employed in this study.

(a)



(b)

Figure 5 Image segmentation (a) before segmentation and (b) after segmentation

3.2.3 CNN Architectures

Comparison and Evaluation of CNN Architectures: In this high-quality research study, we conducted a comprehensive analysis of popular CNN architectures, including EfficientNetV2B3, EfficientNetV2-Small, ShuffleNetV2, SqueezeNetV2, and MobileNetV3-Small. These specific network models were selected due to their well-established reputation for achieving superior accuracy while maintaining comparable prediction times when compared to other networks. This architecture serves as the baseline model for single structure classification, as depicted in Figure 6.

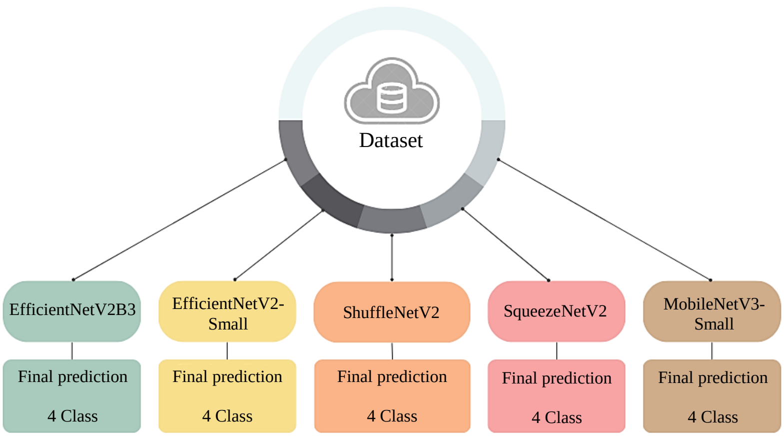


Figure 6 Frameworks of the single structures

A comprehensive explanation of the CNN architecture employed in the aforementioned network models for classifying Ultimate Tensile Strength (UTS) from welds has been presented the following

EfficientNetV2B2, an advanced convolutional neural network architecture that has gained attention in the field of deep learning. It highlights the architecture's improved performance and efficiency compared to previous versions and represents a significant advancement in deep learning, providing a compact and efficient CNN architecture with competitive performance. Its adoption can lead to improved accuracy, faster inference, and better resource utilization in computer vision systems.

EfficientNetV2-Small is a variant of the EfficientNetV2 convolutional neural network (CNN) architecture specifically designed to be compact and efficient while maintaining competitive performance. It is part of the Efficient Net family, which is known for achieving state-of-the-art results in various computer vision tasks. EfficientNetV2-Small is a compact and efficient variant of the EfficientNetV2 CNN architecture. It offers competitive performance in image classification tasks while requiring fewer resources. Its application can enable efficient and accurate computer vision solutions, particularly in resource-constrained environments.

ShuffleNetV2 is a lightweight CNN architecture that offers a balance between model size, accuracy, and computational efficiency. Its channel shuffling technique and optimized building blocks enable efficient information exchange and reduce computational complexity. With its deployment-friendly characteristics, ShuffleNetV2 is a valuable tool for resource-constrained environments and mobile computer vision applications.

SqueezeNetV2 offers a lightweight CNN architecture that balances model size and accuracy. Its efficient fire module design, deep supervision, and feature map concatenation techniques contribute to its superior performance. By leveraging SqueezeNetV2, researchers and practitioners can benefit from an efficient and effective model for image classification tasks in resource-constrained scenarios.

MobileNetV3-Small offers a compact and efficient CNN architecture suitable for resource-constrained environments. Its combination of depth-wise separable convolutions, linear bottlenecks, and squeeze-and-excitation blocks results in a powerful model capable of accurate image classification while maintaining computational efficiency.

3.2.4 Decision fusion strategy

The decision of the final prediction is reported by the individual an ensemble deep learning model (EDL). The decision fusion strategy is used to combine the different solutions from the different methods into the single final prediction of the ensemble model. In this study, Unweight Average Ensemble (UWA) decision fusion strategies are used, which are explained below.

Unweighted Average Ensemble (UWA): The Decision Fusion Strategy (DFS) plays a pivotal role in amalgamating outputs from both homogeneous (Ho) and heterogeneous (He) models into a unified solution, representing the proposed model's solution. In this study, the outputs of the base learners within the ensemble undergo an unweighted averaging process to determine the model's fusion decision. Our methodology entails the meticulous implementation of a carefully selected set of five baseline models. Each individual baseline model plays a crucial role in constructing a homogeneous ensemble comprising multiple models. Notably, we present the MobileNetV3-Small model as a representative example within this framework, as visually depicted in Figure 7. Furthermore, Figure 8 presents a visual representation of the diverse ensemble structures utilized in our study.

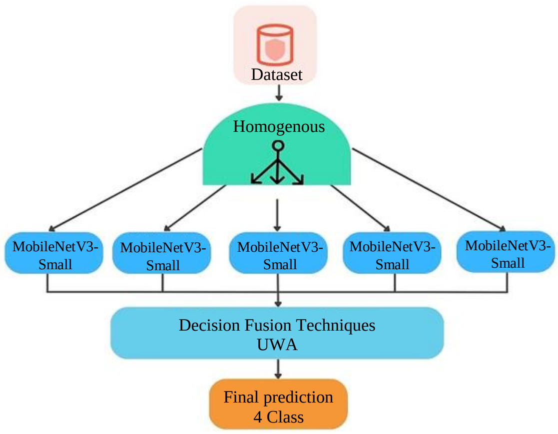


Figure 7 Frameworks of the homogenous ensemble structures

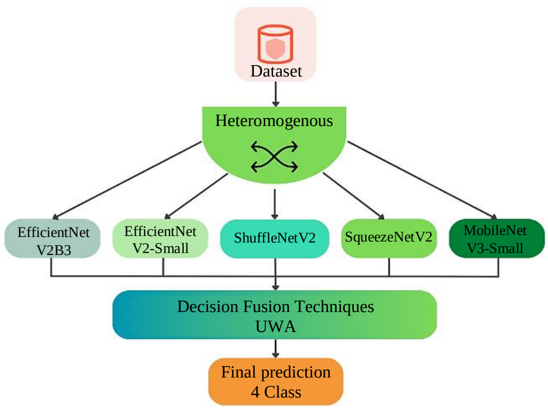


Figure 8 Frameworks of the heterogenous ensemble structures

3.2.5 Comparative analysis of CNN architectures

Conduct a comprehensive comparative analysis of the EfficientNetV2B3, EfficientNetV2-Small, ShuffleNetV2, SqueezeNetV2, and MobileNetV3-Small. And the proposed model using Unweight Average Ensemble (UWA). The decision fusion strategy from many homogenous (Ho-UWA) and heterogenous models (He-UWA). Evaluate the performance of each architecture in terms of UTS classification accuracy, model complexity, and computational efficiency. The experiment comprised a total of 33 models, as shown in Table 5.

Table 5 Details of the experiment

Experiment	**Augmentation**		**Segmentation**		**CNN architecture**					**Decision fusion strategy UWA**		
	Without	**With**	**Without**	**With**	**Efficient Net V2B3**	**Efficient Net V2-Small**	**Shuffle Net V2**	**Squeeze Net V2**	**Mobile Net V3-Small**	**Single**	**Homo**	**Hetero**
1	1		1		1					1		
2	1		1			1				1		
3	1		1				1			1		
4	1		1					1		1		
5	1		1						1	1		
6		1	1		1					1		
7		1	1			1				1		
8		1	1				1			1		
9		1	1					1		1		
10		1	1						1	1		
11		1		1	1					1		
12		1		1		1				1		
13		1		1			1			1		
14		1		1				1		1		
15		1		1					1	1		
16	1		1		1						1	
17	1		1			1					1	
18	1		1				1				1	
19	1		1					1			1	
20	1		1						1		1	
21		1	1		1						1	
22		1	1			1					1	
23		1	1				1				1	
24		1	1					1			1	
25		1	1						1		1	
26		1		1	1						1	
27		1		1		1					1	
28		1		1			1				1	
29		1		1				1			1	
30		1		1					1		1	
31	1		1		1	1	1	1	1			1
32		1	1		1	1	1	1	1			1
33		1		1	1	1	1	1	1			1

4. Result

To develop a robust classification and analysis tool, we implemented the proposed model using the Python programming language, leveraging the Keras and TensorFlow frameworks. The model's performance evaluation was conducted in a Google Colab environment, equipped with 16 GB of RAM and an NVIDIA Tesla V100. To replicate the proposed model, a computer setup with two Intel Xeon CPUs operating at 2.30 GHz, 52 GB of RAM, and a Tesla K80 GPU with 16 GB of GPU RAM was utilized. This configuration was thoughtfully chosen to ensure ample computational resources for efficient model operation. The results of our extensive computational analysis are presented in this section, organized into three distinct subsections. Firstly, we present the identification of the optimal multiclass classification model from a pool of 33 models. This model is specifically designed to accurately classify the Ultimate Tensile Strength (UTS) of weld seams. Secondly, we showcase the superior binary classification models selected from the set of 33 proposed models. Finally, a comprehensive comparative analysis is conducted to assess the performance and efficacy of the models derived from (1) and (2) when compared to state-of-the-art methods.

4.1 Assessment of multiclass classification performance

The proposed model encompasses four essential components: image augmentation, image segmentation, ensemble deep learning, and decision fusion strategies. This section aims to investigate several research inquiries, including: (1) exploring the potential benefits of integrating geometric augmentation as a form of image augmentation within the model; (2) analyzing the advantages of incorporating image segmentation using the U-Net architecture; (3) assessing the effectiveness of employing homogeneous and heterogeneous ensemble deep learning models; and (4) determining the optimal decision fusion strategy, including options such as unweighted average (UWA). The objective is to gain comprehensive insights into these aspects and analyze their respective contributions to the overall performance of the model. Two distinct performance metrics were utilized to assess weld seam UTS classification, encompassing the

F1-score (a measure representing the harmonic mean of precision and recall) and accuracy. The equations for computing accuracy and F1-score are outlined as Equations (1) and (2), correspondingly.

$$accuracy = \frac{TP_j + TN_j}{TP_j + TN_j + FP_j + FN_j} \tag{1}$$

Here, the variables TP, FP, FN, and TN denote True Positive, False Positive, False Negative, and True Negative values, respectively, assigned to the classification of class j.

$$\mathrm{F1} = 2 \times \frac{\left(\frac{TP_j}{TP_j+FN_j} \times \frac{TP_j}{TP_j+FP_j}\right)}{\left(\frac{TP_j}{TP_j+FN_j} + \frac{TP_j}{TP_j+FP_j}\right)} \tag{2}$$

Table 6 KPIs of the proposed methods classifying the UTS of the weld seam.

	Accuracy	F1		Accuracy	F1
1	0.759	0.744	18	0.718	0.712
2	0.782	0.777	19	0.742	0.733
3	0.687	0.680	20	0.744	0.737
4	0.699	0.681	21	0.820	0.804
5	0.715	0.701	22	0.847	0.841
6	0.796	0.779	23	0.780	0.773
7	0.817	0.812	24	0.773	0.756
8	0.736	0.720	25	0.784	0.770
9	0.727	0.713	26	0.879	0.862
10	0.755	0.743	27	0.897	0.875
11	0.841	0.828	28	0.828	0.822
12	0.863	0.851	29	0.823	0.812
13	0.794	0.778	30	0.848	0.841
14	0.776	0.763	31	0.820	0.815
15	0.819	0.803	32	0.882	0.869
16	0.799	0.787	33	0.916	0.898
17	0.811	0.800			

Table 7 Evaluation metrics for each entity include average accuracy and F1-score

KPI	Augmentation		Segmentation		CNN model					Decision Fusion Strategies		
	Without	With	Without	With	Efficient Net V2B3	Efficient Net V2-Small	Shuffle Net V2	Squeeze Net V2	Mobile Net V3-Small	Single	Ho-UWA	He-UWA
Accuracy	0.753	0.823	0.776	0.846	0.842	0.817	0.800	0.800	0.799	0.773	0.808	0.873
F1-score	0.741	0.812	0.764	0.836	0.832	0.809	0.791	0.789	0.789	0.763	0.796	0.861

Table 6 Experiment 33, showcasing the most accurate multiclass classification model. Developed via a four stage method data augmentation, image segmentation, heterogeneous ensemble integration, and He-UWA decision fusion it rigorously assesses each component's impact on the model's effectiveness. Based on the data presented in Table 7, it becomes apparent that the model incorporating augmentation techniques outperforms the model without such implementation, exhibiting a notable increase in accuracy rate by 9.30 %. This finding is consistent with the outcomes derived from the F1-score metrics, where the augmented model consistently surpasses the non-augmented version by 9.58 %. Furthermore, the integration of image augmentation results in a remarkable enhancement, as evidenced by a 6.05 % increase in accuracy, and a 6.28 % increase in F1-score, in comparison to its counterpart that lacks image segmentation. In terms of enhancing solution quality, image segmentation demonstrates a 12.35 % improvement compared to a model without image augmentation, while image segmentation alone leads to a 9.02 % enhancement. The heterogeneous ensemble structure achieves a higher accuracy of 9.67 % compared to other approaches. Consequently, the proposed He-UWA solution surpasses on base CNN Architectures by 6.85 % in terms of accuracy.

4.2 Evaluation of binary classification performance

To assess the effectiveness of the binary classification model, a comprehensive set of 33 experiments was undertaken, as referenced in Table 5. The results derived from these experiments have been meticulously documented and are presented concisely in Table 8.

Among the 33 conducted experiments, Experiment 33 exhibits the highest accuracy in the binary classification. This particular model follows a meticulously designed four-stage methodology, encompassing essential steps such as (1) data augmentation, (2) image segmentation, (3) integration of a heterogeneous ensemble structure, and (4) utilization of He-UWA as the decision fusion strategy. The comprehensive evaluation of the proposed model components and their individual contributions is effectively summarized in Tables 8, offering valuable insights into the performance and efficacy of the model.

To provide a concrete example, the accuracy values showcased in the "no segmentation" column of Table 8 correspond to the average accuracy derived from 33 carefully documented experiments outlined in Table 9. In contrast, the average accuracy values presented in the "segmentation" column, meticulously summarized in Tables 7 and 9, specifically pertain to experiments that incorporate image segmentation. Notably, the final models listed was Experiment 33 in Table 8 consist of augmented and segmented models, such model achieved a remarkable accuracy rate of 97.4% and an F1 score of 96.5%, indicative of their synergistic qualities. Furthermore, the accuracy indicated in the "He-UWA" row of Table 9 reflects the average accuracy across all models that implemented He-UWA for decision fusion. This average encompasses the performances of all models using He-UWA, irrespective of image segmentation implementation.

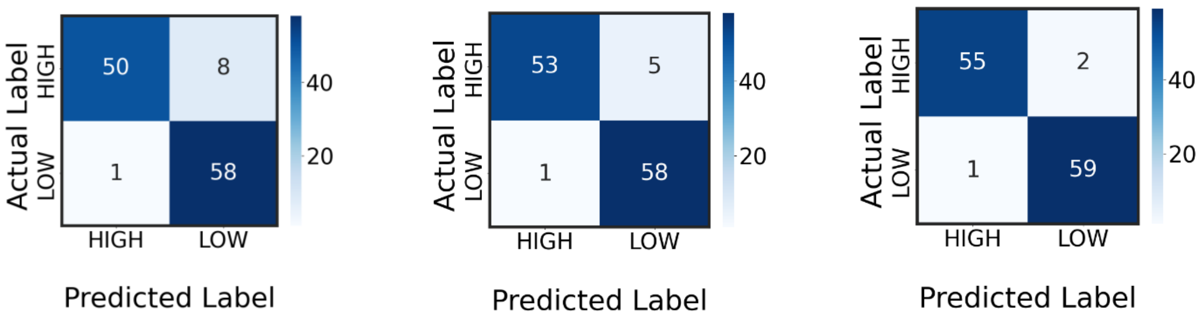
Table 8 KPIs of the proposed binary classification model

	Accuracy	F1		Accuracy	F1
1	0.791	0.783	18	0.772	0.766
2	0.809	0.804	19	0.759	0.743
3	0.727	0.714	20	0.783	0.769
4	0.728	0.720	21	0.877	0.863
5	0.760	0.747	22	0.887	0.870
6	0.842	0.835	23	0.793	0.775
7	0.859	0.847	24	0.794	0.777
8	0.757	0.751	25	0.837	0.827
9	0.763	0.757	26	0.938	0.933
10	0.799	0.787	27	0.949	0.928
11	0.899	0.881	28	0.860	0.848
12	0.923	0.911	29	0.843	0.834
13	0.820	0.814	30	0.880	0.860
14	0.817	0.810	31	0.872	0.860
15	0.852	0.834	32	0.918	0.908
16	0.821	0.804	33	0.974	0.965
17	0.856	0.837			

Table 9 The effects of the components contributing to the proposed model

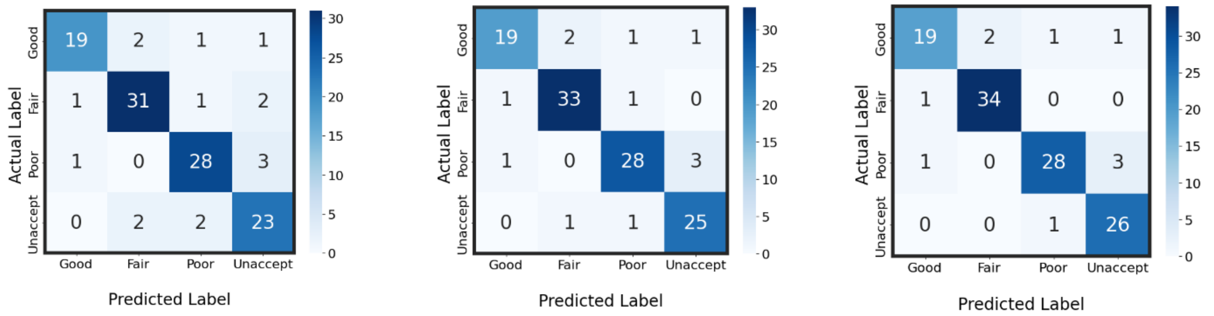
KPI	Augmentation		Segmentation		CNN model					Decision Fusion Strategies		
	Without	With	Without	With	Efficient Net V2B3	Efficient Net V2-Small	Shuffle Net V2	Squeeze Net V2	Mobile Net V3-Small	Single	Ho-UWA	He-UWA
Accuracy	0.783	0.844	0.801	0.871	0.854	0.824	0.811	0.837	0.815	0.800	0.844	0.921
F1-score	0.773	0.834	0.790	0.861	0.844	0.818	0.801	0.823	0.798	0.789	0.822	0.911

Incorporating image segmentation into the model significantly improved its accuracy, elevating it to 0.871 with an F1 score of 0.861. In contrast, models lacking image segmentation showed an average accuracy of 0.801 and an F1 score of 0.790, indicating a correlation between increased accuracy and F1 scores. This improvement is credited to the system's efficiency in distinguishing foreground and background in images by comparing pixel intensity values against a set threshold during segmentation. The system's use of sophisticated thresholding techniques allows it to adjust to various lighting conditions, such as uneven light distribution and shadows. Consequently, the processed images exhibit enhanced clarity and minimized noise, leading to greater accuracy than those without modifications. This observation aligns with findings from related literature. Additionally, the integration of image augmentation techniques contributes to a notable 7.79% improvement in solution quality compared to models without such augmentation. The He-UWA, a heterogeneous ensemble model, demonstrates superior performance, achieving an accuracy that is 11.77% higher than other methods. Moreover, when compared to the homogeneous ensemble model (Ho-UWA), the He-UWA shows a remarkable advantage of 9.12% in accuracy. These findings suggest that the He-UWA model consistently delivers reliable results, whether in multiclass or binary classification tasks. This consistency implies that the model is well-suited for categorizing the quality of weld seams, even as the ultimate tensile strength (UTS) classification potentially extends beyond five levels in the future. The model's applicability in weld seam quality assessment remains robust. The confusion matrices illustrating this categorization are presented in Figures 9 and 10.



(a) (b) (c)

Figure 9 Confusion matrices of 2 classes (a), single model, (b) Ho-UWA, and (c) He-UWA.



(a) (b) (c)

Figure 10 Confusion matrices of 4 classes, (a) single model, (b) Ho-UWA, and (c) He-UWA.

Analyzing the confusion matrices depicted in Figure 10, with particular attention to a single model as an example, it becomes apparent that the model effectively categorized 23 images (85.18%) depicting Unacceptable-quality weld seam visuals. In contrast, the proposed ensemble model exhibited an enhanced performance, accurately recognizing 26 images (96.29%) from the test dataset. It's worth highlighting that the target and output class predictions, encompassing all classes and datasets, consistently fall within a comparable range. This consistency affirms the credibility of the reported accuracy as a faithful representation of the proposed model's proficiency across diverse weld seam quality scenarios.

4.3 Comparison between the proposed methodology and state-of-the-art methods

Within this section, we aim to assess the effectiveness of the proposed method in comparison to other state-of-the-art methodologies. The proposed approach is derived from the optimal model suggested in sections 4.1 and 4.2. In-depth analysis and simulation outcomes of both the proposed method and the compared methodologies are comprehensively presented in Table 10, offering detailed insights into their respective performance.

Table 10 Details and simulation results of the proposed and compared methods

Models	Size of model (MB)	Training time (min)	Testing time (second/image)	Accuracy	F1-score
Efficient Net V2B3	49.35	246	0.030	84.80	83.80
Efficient Net V2-Small	77.58	387	0.047	82.05	81.35
Shuffle Net V2	10.06	143	0.006	80.55	79.60
Squeeze NetV2	4.73	64	0.003	81.85	80.60
Mobile NetV3-Small	3.59	47	0.002	80.70	79.35
Ho- UWA	246.75	1230	0.148	82.61	80.09
He-UWA	145.31	887	0.088	89.70	88.60

Table 10 provides a clear comparison, highlighting the He-UWA model's superior performance in accuracy against models like Efficient Net V2B3, Efficient Net V2-Small, Shuffle Net V2, Squeeze NetV2, and Mobile NetV3-Small. This model exhibits an increase in accuracy within the range of 5.46% to 10.20%. Looking at F1 score metrics, the He-UWA model continues to outshine these counterparts with an average improvement of approximately 9.56%. Additionally, this model is particularly efficient, demonstrating 8.58 % higher accuracy than its counterpart, the Ho-UWA model, and proving to be a more effective option in terms of resource utilization. Despite a smaller size, the He-UWA model manages to reduce its size by 41.11% while maintaining a high level of computational speed in processing. There's a notable reduction in the training time for this model by 27.89%, and the testing time is significantly lowered by 40.54%. This comparative analysis suggests that our introduced model stands out as a more efficient alternative when assessed against the models considered in this study.

5. Discussion

An extensive comparison between our suggested models and existing leading-edge methods consistently underscores the proficiency of the ensemble deep learning model in precisely classifying the ultimate tensile strength (UTS) of weld seams in AA5052 alloy. The research findings consistently reinforce the potential of the ensemble deep learning model for UTS classification, demonstrating consistent and reliable outcomes across different aspects of the research [31-33]. This consistent demonstration of the model's effectiveness and potential further strengthens its position as a valuable tool for quality control and real-time monitoring during welding processes.

The non-destructive nature of the proposed model allows for repeated assessments without compromising the integrity of the weld seam, positioning it as a promising alternative to traditional UTS testing techniques. The study evaluates the effect of geographic augmentation techniques on the performance of the ensemble deep learning model for UTS classification [9], highlighting the potential for consistent and reliable evaluations of weld seam strength. This consistent emphasis on the non-destructive nature of the model and its potential for reliable evaluations underscores its significance in the field of industrial engineering and technology.

The model we have developed presents a thorough and efficient method for precisely classifying the Ultimate Tensile Strength (UTS) of welding joints [31]. By incorporating techniques such as image segmentation, augmentation, and a diverse ensemble framework, the model's accuracy and overall performance have seen substantial improvements, aligning with the guidelines set by prior research [12-14]. This study not only advances the understanding of weld seam quality classification but also offers essential perspectives for upcoming investigations in this domain.

The four-phase model we introduced, which integrates image segmentation, image enhancement, a varied ensemble framework, and a decision fusion method, attained top-tier accuracy in the multiclass classification task, consistent with the principles established by earlier studies [15]. The effects of these elements were comprehensively summarized and displayed in tabular form, offering a clear depiction of their impact on the overall effectiveness of the suggested approach.

Based on our findings, we recommended five foundational CNN models, namely EfficientNetV2B3, EfficientNetV2-Small, ShuffleNetV2, SqueezeNetV2, and MobileNetV3-Small for UTS classification in weld seams. These models were evaluated using the FSW image dataset, and their training and validation accuracy were analyzed [9, 16]. Impressively, our developed model outperformed these standards, demonstrating superior effectiveness. The outcomes of this study reveal that our proposed model delivers stable and reliable performance, whether applied in multiclass or binary classification scenarios.

6. Conclusions and outlook

This study successfully developed a comprehensive classification and analysis tool for accurately classifying the ultimate tensile strength (UTS) of weld seams. The proposed model demonstrated superior performance in both multiclass and binary classification tasks, surpassing other state-of-the-art methods. The incorporation of image segmentation and augmentation techniques, along with a heterogeneous ensemble structure and the decision fusion approach, contributed significantly to the model's accuracy and solution quality. A notable highlight of the study was Experiment 33, which effectively merged these methodologies, delivering a stellar performance. It achieved top-level accuracy in multiclass classification and notable results in binary classification, with an accuracy of 97.4% and an F1 score of 96.5%. This level of accuracy is indicative of the average performance across all models that incorporated the He-UWA for decision fusion strategy. It encompasses the efficacy of all models using He-UWA,

The efficacy of all models that adopted the He-UWA as their decision fusion approach was observed to be superior. These models exhibited a significant increase in accuracy compared to others like Efficient Net V2B3, Efficient Net V2-Small, Shuffle Net V2, Squeeze NetV2, and Mobile NetV3-Small, with improvements in accuracy ranging from 5.46% to 10.20%. Additionally, in terms of F1 score metrics, the He-UWA model consistently outperformed similar models, achieving an average accuracy enhancement of approximately 9.56%.

The He-UWA model distinguishes itself through its enhanced efficiency, achieving an 8.58% improvement in accuracy over the Ho-UWA model, thereby demonstrating greater effectiveness in resource use. Remarkably, the He-UWA model has achieved a 41.11% reduction in its size while still retaining a robust computational processing speed. Additionally, the model has shown a significant decrease in training time, reduced by 27.89%, and a considerable reduction in testing time, lowered by 40.54%.

The suggested base CNN models, such as EfficientNetV2B3, EfficientNetV2-Small, ShuffleNetV2, SqueezeNetV2, and MobileNetV3-Small, proved to be effective for UTS classification in weld seams. These models exhibited strong performance, laying a foundation for further exploration and advancement in this area.

6.1 Outlook

Although the proposed model achieved remarkable accuracy and reliability in the UTS classification of weld seams, there are several avenues for further research and improvement. Future studies could focus on the following aspects:

Expansion of UTS Classification: The proposed model primarily focused on the classification of UTS into four levels. To enhance its practical applicability, future research could explore the extension of UTS classification to include more than four levels. This would provide a more comprehensive analysis of weld seam quality.

Integration of Additional Features: While the proposed model achieved excellent performance using image-based features, the inclusion of additional features, such as textural or structural characteristics, may further enhance its accuracy. Investigating the fusion of multiple modalities could lead to a more comprehensive and robust classification system.

Optimization of Model Architecture: Although the proposed model exhibited high accuracy, there is potential for optimizing the model architecture to further improve its efficiency and speed. Exploring alternative CNN architectures or employing advanced optimization techniques could lead to more streamlined and efficient classification models.

Real-Time Implementation: The application of the proposed model in real-time scenarios, such as welding processes, would be a valuable direction for future research. Developing an embedded system or deploying the model on edge devices could enable on-site UTS classification, facilitating immediate feedback and quality control in welding operations.

7. References

[1] Ahmed MMZ, El-Sayed Seleman MM, Fydrych D, Çam G. Friction stir welding of aluminum in the aerospace industry: the current progress and state-of-the-art review. Materials. 2023;16(8):2971.

[2] Shinde G, Gajghate S, Dabeer PS, Seemikeri CY. Low cost friction stir welding: a review. Mater Today Proc. 2017;4(8):8901-10.

[3] Yang Y, Zhou L. Improving corrosion resistance of friction stir welding joint of 7075 aluminum alloy by micro-arc oxidation. J Mater Sci Technol. 2014;30(12):1251-4.

[4] Sato YS, Sugiura Y, Shoji Y, Park SHC, Kokawa H, Ikeda K. Post-weld formability of friction stir welded Al alloy 5052. Mater Sci Eng A. 2004;369(1-2):138-43.

[5] Matitopanum S, Pitakaso R, Sethanan K, Srichok T, Chokanat P. Prediction of the Ultimate Tensile Strength (UTS) of Asymmetric Friction Stir Welding Using Ensemble Machine Learning Methods. Processes. 2023;11(2):391.

[6] Taye MM. Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. Computation. 2023;11(3):52.

[7] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8:1-74.

[8] Tripathy S, Singh R. Convolutional neural network: an overview and application in image classification. In: Poonia RC, Singh V, Singh Jat D, Diván MJ, Khan MS, editors. Proceedings of Third International Conference on Sustainable Computing. Advances in Intelligent Systems and Computing, vol 1404. Singapore: Springer; 2022. p. 145-53.

[9] Chiaranai S, Pitakaso R, Sethanan K, Kosacka-Olejnik M, Srichok T, Chokanat P. Ensemble deep learning ultimate tensile strength classification model for weld seam of asymmetric friction stir welding. Processes. 2023;11(2):434.
[10] Yang Y, Lv H, Chen N. A survey on ensemble learning under the era of deep learning. Artif Intell Rev. 2023;56(6):5545-89.
[11] Zhou X, He J, Yang C. An ensemble learning method based on deep neural network and group decision making. Knowl.-Based Syst. 2022;239:107801.
[12] Alomar K, Aysel HI, Cai X. Data augmentation in classification and segmentation: a survey and new strategies. J Imaging. 2023;9(2):46.
[13] Negassi M, Wagner D, Reiterer A. Smart (sampling) augment: optimal and efficient data augmentation for semantic segmentation. Algorithms. 2022;15(5):165.
[14] Lee J, Lim K, Cho J. Improved monitoring of wildlife invasion through data augmentation by extract–append of a segmented entity. Sensors. 2022;22(19):7383.
[15] Chakraborty S, Saha AK, Ezugwu AE, Agushaka JO, Zitar RA, Abualigah L. Differential evolution and its applications in image processing problems: A comprehensive review. Arch Computat Methods Eng. 2023;30(2):985-1040.
[16] Smith AG, Petersen J, Selvan R, Rasmussen CR. Segmentation of roots in soil with U-Net. Plant Methods. 2020;16(1):1-15.
[17] Mishra RS, Ma ZY. Friction stir welding and processing. Mater Sci Eng R Rep. 2005;50(1-2):1-78. 2005.
[18] Lombard H, Hattingh DG, Steuwer A, James MN. Optimising FSW process parameters to minimise defects and maximise fatigue life in 5083-H321 aluminium alloy. Eng Fract Mech. 2008;75(3-4):341-54.
[19] Khalafe WH, Sheng EL, Bin Isa MR, Omran AB, Shamsudin SB. The effect of friction stir welding parameters on the weldability of aluminum alloys with similar and dissimilar metals. Metals. 2022;12(12):2099.
[20] Liu G, Murr LE, Niou CS, McClure JC, Vega FR. Microstructural aspects of the friction-stir welding of 6061-T6 aluminum. Scripta materialia. 1997;37(3):355-61.
[21] Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. Energy Convers Manag. 2019;198:111799.
[22] Yao Q, Wang M, Chen Y, Dai W, Li YF, Tu WW, et al. Taking human out of learning applications: a survey on automated machine learning. arXiv preprint arXiv:181013306. 2018:1-20.
[23] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas, USA. USA: IEEE; 2016. p. 770-8.
[24] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning; 2015 Jul 6-11; Lille, France. p. 1-9.
[25] Aggarwal M, Khullar V, Goyal N, Singh A, Tolba A, Thompson EB, et al. Pre-Trained deep neural network-based features selection supported machine learning for rice leaf disease classification. Agriculture. 2023;13(5):936.
[26] Maddury S, Desai K. Using deep learning to determine amyloid deposition through pet and clinical data for Alzheimer's prognosis. medRxiv. 2022:1-12.
[27] Ma N, Zhang X, Zheng HT, Sun J. Shufflenet v2: practical guidelines for efficient cnn architecture design. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer Vision – ECCV 2018. Lecture Notes in Computer Science, vol 11218. Cham: Springer; 2018. p. 122-38.
[28] Li G, Wu Y, Hao X. Smaller and more efficient image classification neural network for mobile devices. Research Square. 2022:1-18.
[29] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv:160207360. 2016:1-13.
[30] Zhu J, Zhang C, Zhang C. *Papaver somniferum* and *Papaver Rhoeas* classification based on visible capsule images using a modified MobileNetV3-Small network with transfer learning. Entropy. 2023;25(3):447.
[31] Mohammed A, Kora R. A comprehensive review on ensemble deep learning: opportunities and challenges. J King Saud Univ - Comput Inf Sci. 2023;35(2):757-74.
[32] Jabbar MA. Breast cancer data classification using ensemble machine learning. Eng Appl Sci Res. 2021;48(1):65-72.
[33] Alsulami AA, AL-Ghamdi ASAM, Ragab M. Enhancement of E-learning student's performance based on ensemble techniques. Electronics. 2023;12(6):1508.
[34] Khoder A, Dornaika F. Ensemble learning via feature selection and multiple transformed subsets: application to image classification. Appl Soft Comput. 2021;113:108006.
[35] Xu H, Yan ZH, Ji BW, Huang PF, Cheng JP, Wu XD. Defect detection in welding radiographic images based on semantic segmentation methods. Measurement. 2022;188:110569.
[36] Zhang Y, You D, Gao X, Zhang N, Gao PP. Welding defects detection based on deep learning with multiple optical sensors during disk laser welding of thick plates. J Manuf Syst. 2019;51:87-94.