Effective Feature Selection Technique for Deep Learning-Based Weld Defect Classification in Gamma Radiographic Images

Refaat M. Fikry, M. E. Hammad, Zeinab F. Elsharkawy, H. Kasban

***Abstract*—The traditional defect detection techniques have poor detection accuracy and are strongly influenced by the industrial imaging environment. Concentrating on this shortcoming, this study presented a hybrid Deep Learning (DL) approach for automatic weld defects classification in Gamma Radiography Images (GRIs). We focus on improving accuracy by fusion and selection of deep-learned features extracted from five different DL models (e.g. SqueezeNet, GoogleNet, ShuffleNet, DarkNet19, and MobileNet-V2). To extract robust features from the DL models, the Pearson Correlation Coefficient (PCC), F-score (FS), and ReliefF (RF) feature selection algorithms are evaluated. The RF algorithm achieved the best result. The selected features are used for classification tasks using a Multiclass K-Nearest Neighbors (MKNN) classifier. Eight main types of weld defects and the normal type are considered in the utilized RGIs dataset. Several experiments are performed using the traditional feature extraction methods, DL methods, feature selection algorithms, and the proposed one. Their results are compared to evaluate the performance of the presented system. Several RF feature subsets were tested, and the 450-feature subset with the best classification performance was found. The results confirmed that the suggested strategy performs better than all traditional and DL methods with an overall classification accuracy (CA) of 99.75%.**

***Keywords*—Deep learning, gamma radiography, NDT, weld defects, ReliefF, MKNN.**

Manuscript received [11 June 2025]; revised [6 July 2025]; accepted [6 July 2025]. Date of publication [21 July 2025].

(*Corresponding author:* Zeinab F. Elsharkawy*)*.

Refaat M. Fikry is with Engineering Department, Nuclear Research Centre, Egyptian Atomic Energy Authority (EAEA), Egypt (e-mail: eng_refaat@yahoo.com).

M. E. Hammad is with Engineering Department, Nuclear Research Centre, Egyptian Atomic Energy Authority (EAEA), Egypt (e-mail: m_hammad2020@yahoo.com).

Zeinab F. Elsharkawy is with Engineering Department, Nuclear Research Centre, Egyptian Atomic Energy Authority (EAEA), Egypt (e-mail: zeinab_elsharkawy@yahoo.com).

H. Kasban is with Engineering Department, Nuclear Research Centre, Egyptian Atomic Energy Authority (EAEA), Egypt (e-mail: hany_kasban@yahoo.com).

I. INTRODUCTION

The demand for different industries to have high-quality weld structures is increasing nowadays, and hence it is essential to inspect these structures using methods of Non-Destructive Testing (NDT). The objective of these tests is to ensure highly reliable structures that maintain their competitiveness in markets. There are numerous industries, including nuclear, naval, chemical, and aerospace ones, that recognize X-ray testing as necessary for the quality control of welded joints as an NDT technique.

The use of NDT techniques to evaluate welded components is critical to ensure consistent performance. Ultrasonic imaging, radiography, and magnetic resonance imaging are the most often used modalities of NDT. Several scholars have conducted a thorough analysis and a contrast between these approaches. Radioisotopes have different NDT applications in the industry; one of these applications is Gamma Radiography Testing (GRT), which determines the quality of a weld by measuring the radiographs of welded components. GRT differs from previous approaches in that it is faster and less expensive, and it may be performed comprehensively and non-invasively. GRT differs from previous approaches in that it is faster and less expensive, and it may be performed comprehensively and non-invasively. GRT uses gamma radioisotopes in radiographic films to detect faults in welding materials. The obtained radiographic films are then analyzed by a professional operator or automatically using image processing algorithms where images are first digitized and then enhanced using different processes. Finally, the weld is classified using methods of detection and recognition. Various methods and approaches have been used by various researchers to study defect identification using radiography images.

Blowholes and cracks are the most internal defects that occur inside the welded assemblies and frameworks, so gamma radiography was introduced like X-ray radiography for weld defect detection [1]. Welded constructions must also be non-destructively evaluated with gamma rays, X-rays, eddy current, liquid penetrant, or ultrasonic testing. Radiographic inspection is a widely used NDT tool for searching for internal defects in welded structures. This technique is based on the premise that γ rays can penetrate through opaque substances like metal, which are invisible to light. [2]. Typically,

radiographic films are fairly dark. They have a comparatively high density. Therefore, using a radiograph, a standard scanner cannot provide adequate illumination. Gamma rays also create photographic records through the radiant energy that is transmitted. Hence, more energy is consumed by the defective areas. Consequently, penetrated γ-rays on the receiving films display differences in intensity. Thus, the defects in the image appear darker [3]. An explanation for the image context is the small variety of intensities the image capture system can handle. The noise pixels' grey level values are also much greater than those of their near neighbors. However, there is a significant bias towards the darker side in the distribution of grey levels in the original radiography image. As a result, these flaws are rarely noticeable. Consequently, the backdrop in the image is depicted by these dark regions. However, this noise is a representation of the high-frequency components of the image. As a result, it offers a method for examining a weld's interior structure. Compared to other methods, gamma radiography has many benefits [4]. It can be done more efficiently and cheaply, thoroughly, and non-invasively. GRT monitoring examines weld flaws in radiographic film materials using gamma radioisotopes [5].

Weld x-ray radiography images have been studied utilizing a variety of approaches and procedures for flaws detection and categorization. Multi-layer Perceptron Neural Network (MLP) approaches for classifying weld faults from simulated data were presented by Lim et al. [6]. The authors utilize six types of faults and 25 shape descriptors to discriminate between each class. The overall accuracy of 97.96% is obtained using 17 selected efficient features. Valavanis et al. [3] used three different classifiers (e.g., Artificial Neural Network (ANN), KNN, and Support Vector Machine (SVM)) for weld flaws categorization from x-ray images. Forty-three features are extracted from each segmented image and selected represented geometrical and texture measurements. The six-defect type classification accuracy of 85.5% is obtained using ANN with 3-fold cross-validation testing. An automatic detection approach of weld defects from radiography is presented in [7]. The process is based on image edge detection with an MLP classifier. To increase the identification accuracy of weld flaws, Liao et al. [8] used an Ant Colony Optimization (ACO)-based method for feature selection. The authors used four classifiers (e.g., nearest mean, KNN, fuzzy KNN, and center-based nearest neighbor) for the identification task. Hassan et al. [9] introduced a detection system of weld defects on radiographic images. The system consists of four steps: noise reduction and defects localization with maximum and minimum intra-class variance, geometric features extraction, and classification using ANN. To identify weld faults in radiographic images, [10] employs an Adaptive-Network-based Fuzzy Inference System (ANFIS) and ANN. After image preprocessing, 12 geometrical features are extracted and fed to the classifiers. The obtained accuracies are 82.6% for the ANFIS and 78.9% for the ANN. Bhat et al. [11] proposed a classification technique of friction stir weld based on Discrete Wavelet Transform (DWT) and SVM. Three features are extracted from the DWT of sub-regions called variance, energy, and entropy and fed to SVM for classification. In [12], an automated detection method of welded joint defects from Radiographic Double Wall Double Image (DWDI) is presented. A feed-forward MLP classifier is used to classify DWDI as defective and non-defective with 88% detection accuracy. Vishal et al. [13] reviewed the Artificial Intelligence (AI) techniques for welding radiography and ultrasonic image classification. The multivariate generalized Gaussian distribution (MGGD)-based finite mixture model was suggested by Nacereddine et al. [14] for weld radiography testing using certain steps, such as image preprocessing, image segmentation, features extraction, and defects classification steps of weld radiography testing. Four types of defects are considered. This method achieved more than 96% accuracy. A Deep Convolutional Neural Network (DCNN) is presented in [15] as an intelligent diagnosis system of weld X-ray images. Deep features are extracted and fed to MLP and fuzzy K-NN classifiers. The highest accuracy of 97.2% is obtained. In [16], a method for detecting weld faults using triplet deep neural networks has been described. To simplify flaw detection, the original X-ray image is preprocessed into a relief image. The relief image function vector is then obtained by mapping using a deep network-based triplet. It is essential that vectors with similar defect characteristics are closer together and that those with different types of defect features are farther apart. To achieve the automatic classification and identification of weld flaws, the defect classification model (SVM) is created in the end.

To identify weld flaws and increase accurate detection in their dataset, a unique classification method based on DL pre-trained network AlexNet topologies was developed [17]. To detect the welding images, a fine-tuning technique is used and is compared to the several pre-trained VGG-16, GoogleNet, VGG-19, ResNet50, and ResNet101 DCNN models as well as deep convolution activation features (DCFA). In [18], an approach for completely automatic weld segmentation and defect identification was proposed. Weld segmentation is performed using the DCNN FgSegNet, and fault detection is performed using the Neural Network (NN) with three convolution layers. Also, in [19, 20] transfer learning is employed with pretrained DCNNs for weld defect classification from X-ray radiography images. Four classes are classified in [19], and the parameters of the pre-trained VGG16 are adjusted to enhance the presented model performance. In [20], 14 types of weld defects are classified.

Moreover, recent advances in DL have significantly enhanced automated weld defect detection in radiographic images, addressing the limitations of manual inspection. Studies comparing multiple CNNs architectures for classification tasks demonstrate that deeper networks like ResNet-101 achieve superior performance (85.71% accuracy) on the large-scale RIAWELC dataset (24,407 images), outperforming simpler models such as AlexNet and SqueezeNet [21]. For localized defect detection, Faster R-CNN models have been adapted to handle challenging cases like small, sticky porosity in limited datasets (453 images), achieving 44.5% average precision (AP) through optimized

anchor boxes and data augmentation, though they lag behind classification-focused approaches in accuracy [22]. The state-of-the-art Weld-CNN, a hybrid model combining sequential and parallel convolutional blocks, sets a new benchmark with 99.83% test accuracy on the RIAWELC dataset, highlighting the efficacy of specialized architectures over transfer learning methods [23]. Collectively, these studies underscore the trade-offs between classification accuracy, defect localization capability, and computational efficiency in industrial applications.

Feature extraction is done with the VGG16 and ResNet50 CNNs, and classification tasks are done with the SVM classifier. Using transfer learning, performance is improved while training time is reduced. The Convolutional Neural Networks (CNNs) and SVM methods are used in [24] to recognize welded joint defects in radiographic images. After morphological filtration, CNNs are utilized for feature extraction and primary classification. After that, the SVM is employed to precisely determine the limits of defects. Stephen et al. [25] presented an automatic approach for detecting weld defects in radiographic images. This method proposes a DL-based approach to identify 4 types of welding defects from 200 x-ray images. DA technique is applied to get better generalization performance. Results show that 95% validation accuracy was obtained.

The detection and categorization of weld defects from GRIs was the subject of fewer researchers. In [23], a cepstral approach is suggested for using GRIs to find problems. The cepstral features are extracted with this method using different transforms like Discrete Sine Transforms (DST), Discrete Cosine Transforms (DCT), and DWT of the GRI. The ANN is used to match features. To identify flaws, Zahran et al. [4] used the spectral properties that they retrieved from the segmented GRIs' Power Density Spectra (PDSs). In [26], Higher-Order Spectra (HOS) are proposed for weld defect detection from GRIs. The authors used bi-spectrum and tri-spectrum that were estimated by direct, indirect, and ARMA methods. Then, the detection process is implemented by ANN. The authors in [4, 26, 27] studied the effect of noise in their approaches. El-Tokhy et al. [1] presented multi-scale Wavelet-Packets Transform (WPT) features extraction from GRI for defects detection. Additionally, the features are taken from the WPT faults signal's DCT and DST. Finally, feature matching can be achieved using SVM.

In various industries, it is vital to automatically detect and classify the weld faults in GRIs. This study employs DL features, a feature selection technique, and an MKNN classifier to increase the detection and CA of faults in weld GRIs. Nine classes, including eight main defects and the normal one, are considered in the GRI dataset.

The rest of the document is organized as follows: The proposed method and conventional classification techniques are introduced in Section 2. Section 3 presents the experiment findings and comments. The study is concluded in Section 4.

II. The Proposed Approach

An automated detection and classification approach of defects in weld GRIs based on DL features is proposed. This approach used for 9-class classification contains normal and 8 types of defects. Fig. 1 shows the samples of the 9 types of welding joints of GRIs. To assess the effectiveness of the suggested strategy, we compare its results with the classification ability of traditional methods, including 11 different types of extracted features. The welding region is initially separated from the entire GRI, and all irrelevant information is eliminated. The extracted welding area is called the region of interest (ROI). Then, the traditional classification method and the proposed DL feature extraction approach are carried out, and their results are compared.

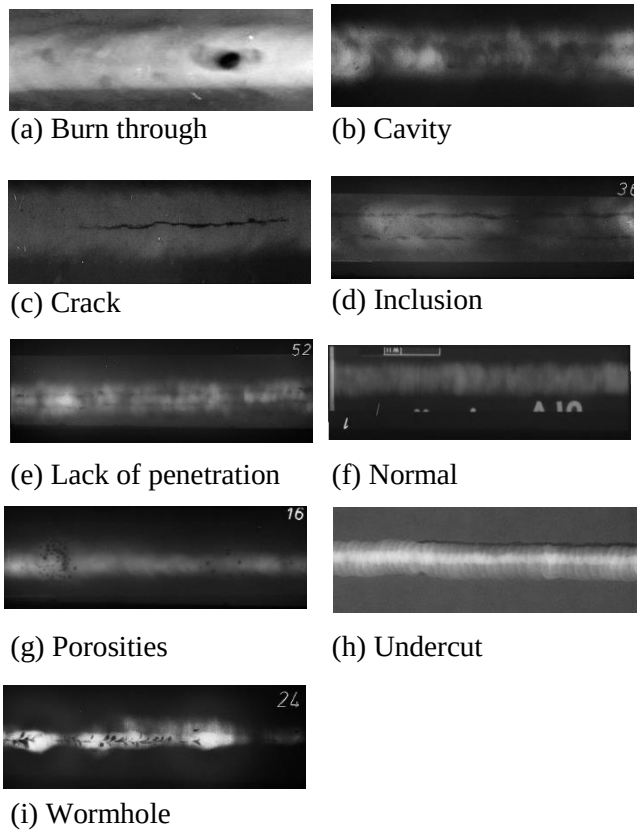


Fig. 1. Samples of the gamma radiography welding images.

A. Traditional Classification Method

Fig. 2 shows the flow diagram of the traditional method. After the ROI step, three steps are performed: preprocessing, feature extraction, and defects classification. For image enhancement, Histogram Equalization (HE), the Retinex algorithm, and bilateral filtering are employed. Several features are extracted and classified.

Image preprocessing: The intensity of each GRI is standardized to produce the same range of grey levels for every image. The image enhancement stage is then carried out utilizing HE. An image processing technique called HE is

employed to boost contrast in the image. Every image that goes through the enhancement process has its contrast adjusted based on the pixels around it. After that, the Retinex algorithm was employed [28]. The image has been modified to have moderate brightness and strong contrast, and it can better retain the image's features. Subsequently, bilateral filtering—a non-linear, edge-preserving, and noise-reducing image smoothing filter—is used [29]. It replaces each pixel's intensity with a weighted average of the intensity values of adjacent pixels. The results of GRIs preprocessing steps were shown in Fig. 3.

Features extraction: The first step before pattern classification is feature extraction. To make the decision-making process for the categorization task simpler, it is mostly used to represent the raw data. Eleven different types of features that are used for classification are introduced in this section. In this step, the features are extracted from the processed weld GRI. First, features are extracted from transform domains like DST, DCT, DWT, and WPT. After that, features may be extracted from transforms of WPT as DCT (WPT), DST (WPT), and DWT (WPT). The PDS feature extraction is also applied. Other features are extracted from the HOS of weld GRIs, such as bi-spectrum (Bic) using the direct method (DM), indirect method (IDM), and ARMA, and tri-spectrum (Tris) using ARMA. Each extracted feature's ability is examined using the SVM or ANN classifier, and their accuracy is compared with the proposed DL approach.

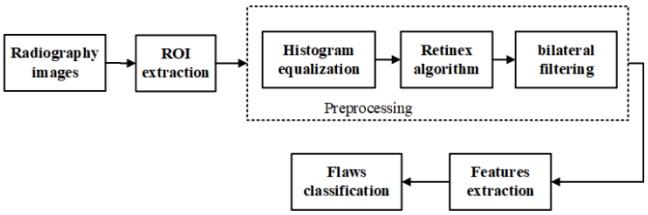


Fig. 2. The flow diagram of the traditional classification method of welding GRIs.

B. The Deep Learning Feature Extraction Approach.

An efficient method for automated weld defects classification in GRI based on fusion and selection of deep-learned features is proposed. Five different DL models called SqueezeNet [30], GoogleNet [30, 31], ShuffleNet, DarkNet19, and MobileNet-V2 are used. SqueezeNet is a CNN that stands out because to its simple network design. Nevertheless, compared to AlexNet, it performs better and provides a smaller model size (less than 50×) [30, 32, 33]. Eight fire modules, three max pooling layers, two convolution layers, one output layer (Softmax), and one global average pooling layer make up SqueezeNet's sixty-eight layers.

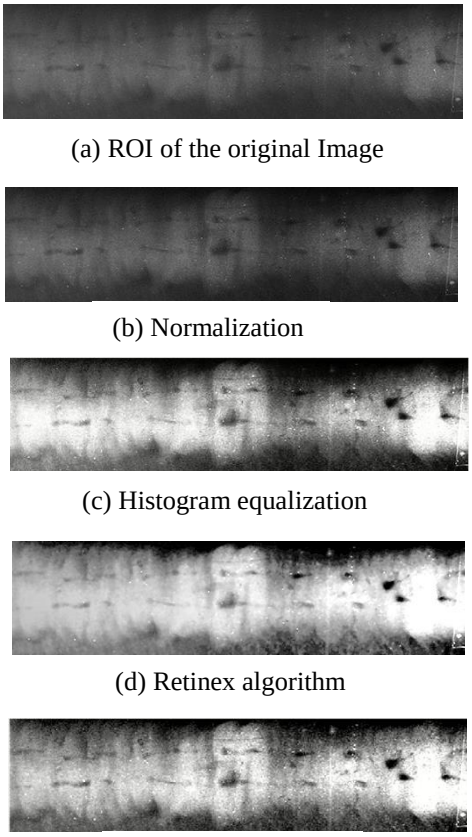
GoogleNet is a complex design because of the inception modules inside its structure. It has 22 layers and achieved a 5.7 percent error rate to win the ImageNet competition. It is widely regarded as one of the first CNN structures to avoid sequentially pooling layers and stacking convolutional layers [30, 31]. ShuffleNet, a CNN, outperforms several networks in speed and accuracy measures at the same computing circumstance [32]. It is made up of 172 layers in total, comprising a convolution layer, a max pooling layer, three stages with ShuffleNet units, one global average pooling, a fully connected layer, and a Softmax output. Darknet19 is a deep CNN with sixteen layers [34]. However, each layer is carefully designed to optimize the network's ability to detect objects in images. It primarily employs 3x3 filters and doubles the number of channels after each pooling phase. This approach, like Network in Network (NIN), employs global average pooling for prediction and 1x1 filters to reduce feature representation between 3x3 convolutions. This approach, similar to Network in Network (NIN), employs global average pooling for prediction and 1x1 filters to reduce feature representation between 3x3 convolutions. Batch normalization improves training stability, accelerates convergence, and ensures model batch consistency. MobileNet-v2 is a 53-layer CNN optimized for mobile devices [35]. It is based on an inverted residual structure, with residual connections emerging at bottleneck layers. The first fully convolution layer with 32 filters makes up MobileNetV2's design. It is followed by 19 residual bottleneck layers.

After that, the extracted features from the 5CNN are fused, and the robust features are selected using the PCC, FS, or RF algorithm, which is fed to an MKNN classifier for automatic classification of weld GRI. The test statistic known as PCC is used to determine the statistical association or link between two continuous variables [36, 37]. It provides details on the direction of the relationship as well as the strength of the association, or correlation. In the end, the characteristics that had the highest separability and lowest correlation were chosen for categorization. FS is a straightforward feature selection filter technique that assesses how well two sets of real numbers can be discriminated from [38, 39]. A feature selection technique based on statistics is the F-score. It evaluates each feature separately to determine which characteristic is relevant. A higher F-score number indicates the most useful feature. RF algorithm One of the most effective feature filtering algorithms, RF was used for multi-class situations [30, 40]. This algorithm can be used to successfully finish feature selection. The RF method is quite effective and places no limitations on the data type attributes. The RF algorithm aids in the solution of a variety of issues by selecting the most nearby samples from each sample in many categories. When y is a categorical variable with multiple classes, RF determines the predictor weights. The strategy encourages forecasters who assign different values to neighbors in separate classes while penalizing those who assign different values to neighbors in the same class [41]. The RF algorithm's formula is Eq. 1.

$$W_f^{i+1} = W_f^i + \sum_{C \neq class(x)} \frac{\frac{P(x)}{1-P(class(x))} \sum_{j=1}^{k} diff(x, Mj(x))}{m*k} - \sum_{j=1}^{k} diff_f(x, H_j, (x))/(m*k) \quad (1)$$

The difference in feature size between the two samples is represented by *diff* (). The neighbor samples of the sample x

are represented by $H_j(x)$. The neighbor samples from neighbors belonging to various classes are represented by Mj(x). The probability of class is given by p(x).



(a) ROI of the original Image

(b) Normalization

(c) Histogram equalization

(d) Retinex algorithm

(e) bilateral filtering

Fig. 3. The results of GRIs preprocessing steps.

Fig. 4 depicts the proposed DL approach's architecture. In the first stage, the prepared dataset's input images are scaled to fit the suggested network (255x255x3). The five pre-trained architectures were then utilized to extract features. Each CNN network generates 1000 features on its last fully connected layers from each input image. In the next stage, all 5000 extracted features from the 5 CNN architectures were collected and sent to the PCC, FS, or RF algorithm for feature selection and compared to obtain the best one.

The 500 RF features that had the best prediction accuracy were selected, and then they were split into 50 subfeatures. Finally, an MKNN classifier was used to evaluate its classification efficiency. Consequently, the RF algorithm decreased a total of 5000 features to 500 features. Based on the experiment findings and comparison with other published approaches, the 450-feature selection yielded the best classification performance.

The five selected CNN architectures—SqueezeNet, GoogleNet, ShuffleNet, DarkNet19, and MobileNet-V2—were chosen due to their lightweight design, architectural diversity, and proven performance in embedded and real-time applications. While deeper networks like ResNet50 and EfficientNet-b0 Feffwere also evaluated (see Table 4 and Table 6), the selected models offered a better trade-off between computational efficiency and feature diversity, which is crucial for feature fusion and selection in our proposed hybrid framework.

III. Experimental Setup

A. Implementation

In this section, the digitized radiographic dataset that was taken from [27] is augmented and used. It obtains 2026 radiographic images and contains normal and eight types of weld defects for welded pipes, and they are obtained using gamma scanning.

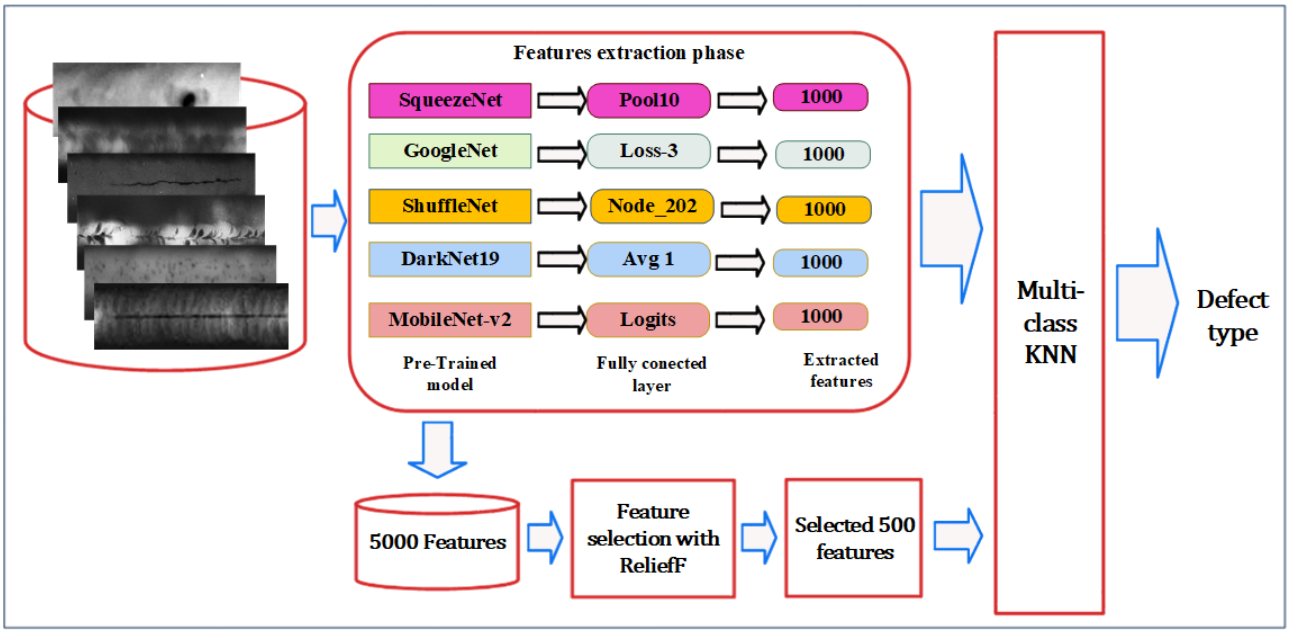


Fig. 4. The proposed DL classification approach.

Fig. 1 shows samples of the weld GRIs dataset that was used in our experiments. Details of the radiography dataset images are depicted in Table 1. The table clearly shows that 80% of the image datasets are utilized as a training set, with the remaining 20% used as a testing set.

The dataset exhibits class imbalance, with some defect types (e.g., burn-through) being underrepresented. To mitigate this, the ReliefF algorithm was chosen for its robustness in multi-class scenarios, as it assigns weights based on the ability of features to distinguish between neighboring samples of different classes. Additionally, the MKNN classifier, which considers multiple neighbors during classification, helps reduce the bias toward majority classes. No oversampling or synthetic data generation was applied to preserve the integrity of the original dataset.

TABLE 1
THE DETAILS OF THE RADIOGRAPHY IMAGES DATASET

No	Class name	Number of images		
		Total	Training	Testing
1	Burn through	64	51	13
2	Cavity	264	211	53
3	Cracks	207	166	41
4	Inclusion	221	177	44
5	Lack of penetration	257	206	51
6	Normal	325	260	65
7	Porosities	388	311	77
8	Undercut	133	105	28
9	Wormhole	167	134	33
Total	**All 9 classes**	**2026**	**1621**	**405**

Three experiments were carried out on GRIs to assess the efficacy of the suggested method. Binary classification is the first experiment; the GRIs database is classified into normal and abnormal images using our proposed method to compare its result with state-of-art. In this case, 325 normal images and 1701 defective images are used in our experiment, including porosities, lack of penetration, cracks, inclusion, undercut, cavity, wormhole, and burn-throughs. MKNN classifier has been built to examine the extracted features from the proposed approach.

The second experiment is the detection of abnormal radiographic images. The traditional and DL approaches are applied to the weld GRIs database to be examined and compared. In this case, all detection and classification methods are used to discriminate between 9 classes. Several features are extracted from the radiographic database, including traditional methods (PDS, DCT, DST, DWT, WPT, LBP, DCT (WPT), DST (WPT), DWT (WPT), Bis (IDM), Bis (ARAMA), and Tris (ARAMA)) and the five DL models (SqueezeNet, GoogleNet, ShuffleNet, DarkNet19, and MobileNet-V2). Each extracted feature of the five DL models has been used to train and test MKNN classifiers with 80% training and 20% testing of all datasets. The training dataset is constructed from 1621 radiographic images of the database containing all weld types, i.e., normal and flaw images, as cleared in Table 1. The remaining database constructs the testing dataset (405 radiographic images).

In the third experiment, the extracted features from 5 CNN networks are fused. Then, the PCC, FS, and RF algorithms are examined and compared to obtain robust features. The selected features are sent into the classification process using MKNN. The proposed method was used to analyze 5000 features from DL models in total. These features were split into 50 sub-features with the highest prediction accuracy. The MKNN classifier was given inputs of 50, 100, 150, and 500 features, respectively.

B. Evaluation Metrics

The Confusion Matrix (CM) and Receiver Operation Characteristics (ROC) curves are used to demonstrate the suggested model's performance. Other measures also used to evaluate the performance of CNNs and proposed networks include sensitivity, specificity, precision, DSC, and accuracy.

$$\text{Sensitivity} = \frac{T_P}{T_P+F_N} \quad (2)$$

$$specificity = \frac{T_N}{T_N+F_P} \quad (3)$$

$$Precision = \frac{T_P}{T_P+F_P} \quad (4)$$

$$DSC = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

$$ClassificationAccuracy = \frac{T_P+T_N}{T_N+T_P+F_N+F_P} \quad (6)$$

Where TN and TP, respectively, represent the total number of true negatives and positives. The total number of false negatives and false positives, respectively, equals FN and FP.

IV. Results and discussion

Table 2 shows the results of our first experiment. The suggested method and other published methods are compared with the CAs on normal and abnormal radiography pictures in the table. The highest CA of 100% is achieved using the proposed detection approach. The results of the second experiment are illustrated in Tables 2 and 3. In the tables, a comparison study between traditional methods, DL models, 5 CNN with feature selection algorithms, and the proposed. The obtained CA of the nine classes of GRI are listed for each method. The highest classification accuracy is achieved using the proposed approach. Figure 5 illustrates the success of the feature selection techniques. A comparison study between PCC, FS, RF, and no feature selection with the 5 CNN is clear in Table 3 and Fig. 5 at different training ratios of 70%, 80%, and 90%.

TABLE 2
COMPARISON OF CLASSIFICATION ACCURACIES RADIOGRAPHIC IMAGES

	Method	Binary CA %	9 class CA%
[1]	WPT+SVM	99.21	69.5
	DST(WPT)+SVM	99.22	76.27
	DCT(WPT)+SVM	99.5	79.66
	DWT(WPT)+SVM	99.5	84.75
[4]	PDS+ANN	99.5	74.57
[27]	DWT+ANN	95	77.97
	DCT+ANN	97	81.36
	DST+ANN	99	79.66
[26]	Bis (ARMA) + ANN	99.5	81.36
	Bis (IDM) + ANN	94	72.88
	Tris (ARMA) + ANN	99.22	71.2
The proposed	5CNN+ReliefF+MKNN	100	99.75

TABLE 3
COMPARISON OF THE HYBRID 5CNN CA WITH VARIOUS FEATURE SELECTION TECHNIQUES AND THE CA OF THE INDIVIDUAL CNN MODELS

Pretrained CNN model	Training ratio			
	60%	70%	80%	90%
SqueezeNet	95.31	95.55	95.80	96.04
GoogleNet	95.56	95.72	95.80	96.04
ShuffleNet	96.18	96.38	96.54	97.03
DarkNet19	94.94	95.93	96.05	97.52
MobileNet-v2	96.92	97.36	97.78	98.02
5 CNN	97.04	97.53	97.86	98.02
5 CNN + PCC	97.87	98.52	98.77	99.01
5 CNN + FS	98.15	99.01	99.26	99.5
5 CNN + RF	98.64	99.34	99.75	99.5

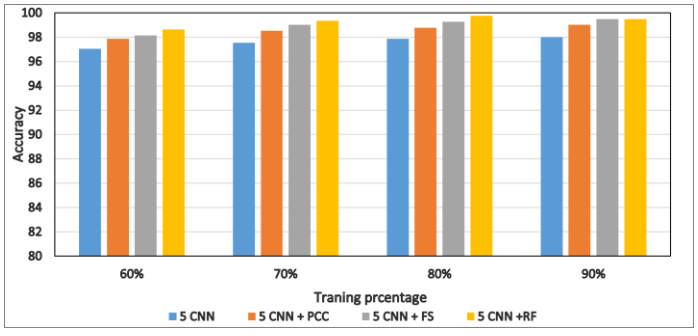


Fig. 5. Comparison of the CA of hybrid 5 CNN with and without feature selection algorithms.

A comparison study between traditional and proposed methods is illustrated in Table 4 at an 80% training ratio. In the table, the obtained CA of the nine classes of GRI are listed for each method. The effects of several chosen features by PCC, FS, and RF feature selection algorithms on the CA are made clear in Table 5 and Fig. 6. It is obvious that 500 features attain higher CA for all feature selection algorithms. As can be shown in Fig. 6, the performance rate rose from 50 to 500 feature sets. Consequently, the combined DL and RF feature selection method performed best with only 450 features, as opposed to the reduced 5000 features.

Specificity, precision, DSC, and accuracy are the main metrics used to judge how effective the suggested approach is. Table 6 displays the typical results of the suggested model and other DL models. The performance of the proposed model outperforms not only the traditional methods but also the DL models. The Confusion Matrices (CMs) and the ROCs of the proposed models are depicted in Fig. 7 and Fig. 8, respectively. The 405 GRI are tested, all classes are correctly classified except one class, and only one image of inclusion images is misclassified into cavity by adding RF to our approach.

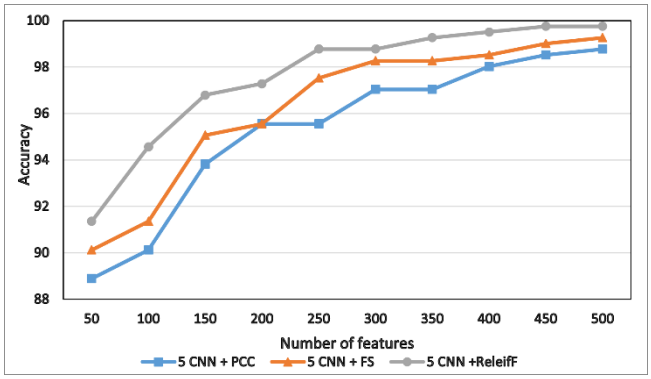


Fig. 6. Accuracy of the hybrid DL models with different sub-feature sets.

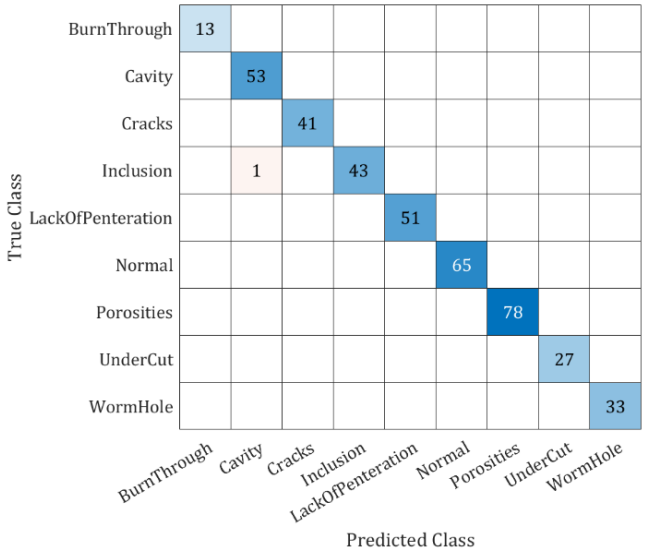


Fig. 7. The proposed model confusion matrix.

TABLE 4
CLASSIFICATION ACCURACIES (%) OF 9 CLASSES OF RADIOGRAPHIC DATABASE

Method	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
WPT [1]	58.62	41.4	51.72	48.3	50	41.4	82.8	91.4	87.93
DPS [4]	96.55	82.76	79.3	91.4	79.3	87.93	82.76	84.5	94.83
DST [27]	87.93	87.93	91.38	93.1	89.93	87.93	91.38	93.1	96.55
Bispc [26]	93.1	79.31	79.31	91.4	91.4	79.31	86.21	84.48	89.65
EfficientNet_b0	99.75	99.01	97.31	98.52	97.53	99.5	98.02	99	99.25
ResNet50	100	99.25	98.27	99.5	99.75	100	98.76	100	100
SqueezeNet	100	97.78	99.26	98.52	98.02	100	98.27	99.75	100
GoogleNet	100	98.77	99.01	97.78	98.02	100	98.52	100	99.51
ShuffleNet	100	98.52	99.26	97.78	99.26	100	98.77	100	99.51
DarkNet19	100	98.52	99.01	98.77	98.27	100	98.77	99.51	99.26
MobileNet-v2	100	98.52	99.01	98.77	98.27	100	98.77	99.51	99.26
5 CNN + PCC	100	99.51	99.51	99.51	99.26	100	100	99.75	100
5 CNN + FS	100	100	99.51	99.51	99.75	100	99.75	100	100
The Proposed	100	99.75	100	100	99.75	100	100	100	100

TABLE 5
THE CA OF HYBRID DL MODELS WITH DIFFERENT FEATURES SUBSETS

Number of features	Accuracy		
	5 CNN + PCC	**5 CNN + FS**	**5 CNN + RF**
50	88.89	90.12	91.35
100	90.12	91.35	94.57
150	93.82	95.06	96.79
200	95.55	95.55	97.28
250	95.55	97.53	98.77
300	97.04	98.27	98.77
350	97.04	98.27	99.26
400	98.02	98.52	99.51
450	98.52	99.01	99.75
500	98.77	99.26	99.75

TABLE 6
A COMPARISON OF THE SUGGESTED MODEL'S AVERAGE PERFORMANCE WITH ALTERNATIVE MODELS.

Method	Sensitivity	Precision	Specificity	DSC	Accuracy
EfficientNet_b0	94.3	93.56	99.1	93.84	94.07
ResNet50	97.63	98.48	99.24	97.95	97.78
SqueezeNet	99.46	96.34	96.34	96.31	95.80
GoogleNet	99.47	96.10	96.44	96.21	95.80
ShuffleNet	99.56	96.98	96.70	96.82	96.54
DarkNet19	99.5	96.16	96.25	96.16	96.05
MobileNet-v2	99.5	96.16	96.25	96.16	97.78
5 CNN + PCC	99.85	98.64	98.87	98.73	98.77
5 CNN + FS	99.9	99.38	99.24	99.29	99.26
The Proposed	99.97	99.79	99.75	99.77	99.75

The ROC curves in Fig. 8 demonstrate that our method produces the best classifier performance. For the model without the RF, a lower performance was attained. Results show that our model performed better than other models. Only when our model is applied do we achieve the best CA of 98.3%. For the DL model without the RF algorithm, a lower performance was attained. Results show that our model performed better than other models. The lower CA is attained in the absence of the RF. Only when our model is applied do we achieve the best CA of 99.75%.

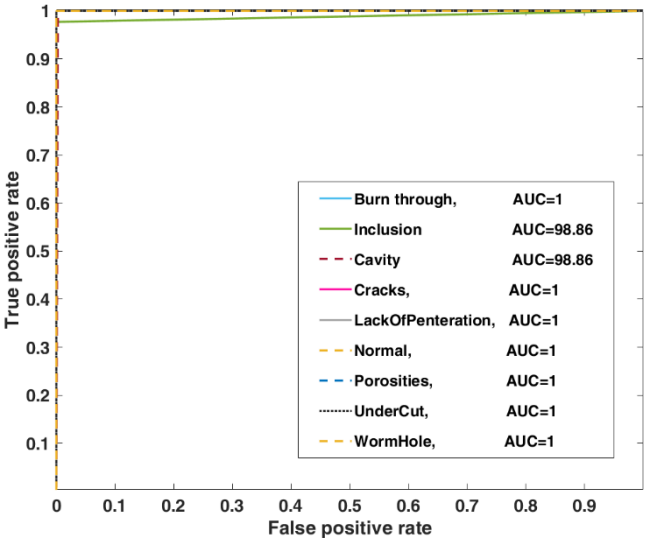


Fig. 8. The ROC curve of our model.

Ablation analysis: We perform ablation experiments to evaluate the effectiveness of the key parameters (RF and CNN networks) in the proposed model. Table 7 presents the results in numerical form. Each ablation study only removes the examined component from the proposed system while keeping the others. We examine the effects of removing each of the five CNN networks. Four CNNs are used in each ablation session, and 450 features are selected by the RF algorithm and sent to MKNN for classification. As seen in Table 7, the CA is lower compared to the proposed network without (W/O) SqueezeNet, GoogleNet, ShuffleNet, DarkNet19, or MobileNet-V2. The largest reduction occurred when MobileNet-V2 was not used. Additionally, the impact of the RF feature selection model is investigated. The MKNN classifier in this study receives all 5000 extracted features from the five CNNs and uses them to carry out the classification process. The worst CA is attained in the absence of the RF, as the table illustrates. Using the proposed model is the only way to achieve the best CA.

V. Conclusion

This paper presents a hybrid DL model used for improving the CA of weld defects of GRIs with the fewest possible features. SqueezeNet, GoogleNet, ShuffleNet, DarkNet19, and MobileNet-V2 DL models, five of the most common CNN architectures, contributed weighted DL features to this model. All features of the five CNN models are collected and fused. Selecting reliable, highly ranked features is done using the RF algorithm. Out of the 5000 features, 450 robust features are selected that affect classification performance. The features with the highest weight—50, 100, 150, ..., 450—were classified using MKNN. For training ratios of 60%, 70%, and 80%, the efficient suggested hybrid technique achieved 98.64%, 99.34%, and 99.75% CA. When compared to traditional approaches, independent CNN models, and other studies in literature, the proposed hybrid approach provides the best performance for weld defects classification. The ablation study highlights the importance of our model's essential elements.

TABLE 7
ABLATION ANALYSIS OF IMPORTANT COMPONENTS IN OUR TECHNIQUE AT DIFFERENT TRAINING RATIO.

W/O	Accuracy at different training ratio			
	60%	70%	80%	90%
SqueezeNet	98.03	98.35	98.52	99.01
GoogleNet	97.78	98.18	98.27	99.01
ShuffleNet	97.65	98.02	98.27	98.51
DarkNet19	97.53	97.86	98.19	98.51
MobileNet-v2	97.29	97.86	98.02	98.51
RF	97.04	97.53	97.86	98.02

Acknowledgments

This work has been supported by the International Atomic Energy Agency (IAEA) within the framework of the IAEA Coordinated Research Project F22069 (Imaging Technologies for Process Investigations and Components Testing), contract number 22610.

References

[1] M. S. El-Tokhy and I. I. Mahmoud, "Classification of Welding Flaws in Gamma Radiography Images Based on Multi-scale Wavelet Packet Feature Extraction Using Support Vector Machine," *Journal of Nondestructive Evaluation*, vol. 34, no. 4, pp. 1-17, 2015, doi: 10.1007/s10921-015-0305-9.

[2] X. Wang, B. S. Wong, and C. S. Tan, "Recognition of welding defects in radiographic images by using support vector machine classifier," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 2, no. 3, pp. 295-301, 2010.

[3] I. Valavanis and D. Kosmopoulos, "Multiclass defect detection and classification in weld radiographic images using geometric and texture features," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7606-7614, 2010, doi: 10.1016/j.eswa.2010.04.082.

[4] O. Zahran, H. Kasban, M. El-Kordy, and F. E. A. El-Samie, "Automatic weld defect identification from

radiographic images," *NDT and E International*, vol. 57, pp. 26-35, 2013, doi: 10.1016/j.ndteint.2012.11.005.

[5] V. A. Golodov and A. A. Mittseva, "Weld Segmentation and Defect Detection in Radiographic Images of Pipe Welds," in *2019 International Russian Automation Conference (RusAutoCon)*, 8-14 Sept. 2019 2019, pp. 1-6, doi: 10.1109/RUSAUTOCON.2019.8867734.

[6] M. M. Ratnam and M. Khalid, "Automatic classification of weld defects using simulated data and an MLP neural network," no. January 2015, 2007, doi: 10.1784/insi.2007.49.3.154.

[7] N. B. Yahia, T. Belhadj, S. Brag, and A. Zghal, "Automatic detection of welding defects using radiography with a neural approach," 2011, vol. 10, pp. 671-679, doi: 10.1016/j.proeng.2011.04.112.

[8] T. W. Liao, "Improving the accuracy of computer-aided radiographic weld inspection by feature selection," *NDT and E International*, vol. 42, no. 4, pp. 229-239, 2009, doi: 10.1016/j.ndteint.2008.11.002.

[9] J. Hassan, "Welding Defect Detection and Classification Using Geometric Features," no. III, pp. 0-5, 2012, doi: 10.1109/FIT.2012.33.

[10] J. Zapata, R. Vilar, and R. Ruiz, "Performance evaluation of an automatic inspection system of weld defects in radiographic images based on neuro-classifiers," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8812-8824, 2011, doi: 10.1016/j.eswa.2011.01.092.

[11] N. N. Bhat, K. Kumari, S. Dutta, S. K. Pal, and S. Pal, "Friction stir weld classification by applying wavelet analysis and support vector machine on weld surface images," *Journal of Manufacturing Processes*, vol. 20, pp. 274-281, 2015, doi: 10.1016/j.jmapro.2015.07.002.

[12] N. Boaretto and T. M. Centeno, "Automated detection of welding defects in pipelines from radiographic images DWDI," *NDT and E International*, vol. 86, no. February 2016, pp. 7-13, 2017, doi: 10.1016/j.ndteint.2016.11.003.

[13] V. Vishal, R. Ramya, P. Vinay Srinivas, and R. Vimal Samsingh, "A review of implementation of artificial intelligence systems for weld defect classification," *Materials Today: Proceedings*, vol. 16, pp. 579-583, 2019, doi: 10.1016/j.matpr.2019.05.131.

[14] N. Nacereddine, A. B. Goumeidane, and D. Ziou, "Unsupervised weld defect classification in radiographic images using multivariate generalized Gaussian mixture model with exact computation of mean and shape parameters," *Computers in Industry*, vol. 108, pp. 132-149, 2019, doi: 10.1016/j.compind.2019.02.010.

[15] W. Hou, Y. Wei, Y. Jin, and C. Zhu, "Deep features based on a DCNN model for classifying imbalanced weld flaw types," *Measurement: Journal of the International Measurement Confederation*, vol. 131, pp. 482-489, 2019, doi: 10.1016/j.measurement.2018.09.011.

[16] X. Liu, J. Liu, F. Qu, H. Zhu, and D. Lu, "A Weld Defect Detection Method Based on Triplet Deep Neural Network," *Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020*, pp. 649-653, 2020, doi: 10.1109/CCDC49329.2020.9164549.

[17] C. Ajmi, J. Zapata, S. Elferchichi, A. Zaafouri, and K. Laabidi, "Deep Learning Technology for Weld Defects Classification Based on Transfer Learning and Activation Features," *Advances in Materials Science and Engineering*, vol. 2020, p. 1574350, 2020/08/14 2020, doi: 10.1155/2020/1574350.

[18] V. A. Golodov and A. A. Maltseva, "Approach to weld segmentation and defect classification in radiographic images of pipe welds," *NDT and E International*, vol. 127, no. November 2021, pp. 102597-102597, 2022, doi: 10.1016/j.ndteint.2021.102597.

[19] Z. Zhang, W. Liu, and X. Sun, "Image recognition of limited and imbalanced samples based on transfer learning methods for defects in welds," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 236, no. 12, pp. 1643-1652, 2022, doi: 10.1177/09544054221082779.

[20] S. Kumaresan, K. S. J. Aultrin, S. S. Kumar, and M. D. Anand, "Transfer Learning with CNN for Classification of Weld Defect," *IEEE Access*, vol. 9, pp. 95097-95108, 2021, doi: 10.1109/ACCESS.2021.3093487.

[21] T. Ting *et al.*, "Deep Learning in Manufacturing: A Focus on Welding Defect Classification with CNNs," *Proceedings of International Conference on Artificial Life and Robotics*, vol. 29, pp. 877-882, 02/22 2024, doi: 10.5954/ICAROB.2024.OS26-9.

[22] C. Ajmi, J. Zapata, S. Elferchichi, and K. Laabidi, "Advanced Faster-RCNN Model for Automated Recognition and Detection of Weld Defects on Limited X-Ray Image Dataset," *Journal of Nondestructive Evaluation*, vol. 43, no. 1, p. 14, 2023/12/16 2023, doi: 10.1007/s10921-023-01032-x.

[23] N. Thi Hoa, T. Ha Minh Quan, and Q. B. Diep, "Weld-CNN: Advancing non-destructive testing with a hybrid deep learning model for weld defect detection," *Advances in Mechanical Engineering*, vol. 17, no. 5, 2025, doi: 10.1177/16878132251341615.

[24] S. Roman, V. Viacheslav, G. Nikolay, Z. Aleksandr, and P. Aleksandra, "Automatic detection of welding defects using the convolutional neural network," in *Proc.SPIE*, 2019, vol. 11061, p. 110610E, doi: 10.1117/12.2525643.

[25] D. Stephen and D. L. P.P, "Development of Radiographic Image Classification System for Weld Defect Identification using Deep Learning Technique," *International Journal of Scientific & Engineering Research*, vol. 12, no. 5, pp. 390-394, 2021, doi: 10.14299/ijser.2021.05.01.

[26] S. Saber and G. I. Selim, "Higher-Order Statistics for Automatic Weld Defect Detection," Journal of *Software Engineering and Applications*, vol. 06, no. 05, pp. 251-258, 2013, doi: 10.4236/jsea.2013.65031.

[27] H. Kasban, O. Zahran, H. Arafa, M. El-Kordy, S. M. S. Elaraby, and F. E. Abd El-Samie, "Welding defect detection from radiography images with a cepstral approach," *NDT and E International*, vol. 44, no. 2, pp. 226-231, 2011, doi: 10.1016/j.ndteint.2010.10.005.

[28] A. S. Parihar and K. Singh, "A study on Retinex based method for image enhancement," *Proceedings of the 2nd International Conference on Inventive* Systems and Control, ICISC 2018, no. Icisc, pp. 619-624, 2018, doi: 10.1109/ICISC.2018.8398874.

[29] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications,"

Foundations and Trends in Computer Graphics and Vision, vol. 4, no. 1, pp. 1-73, 2009, doi: 10.1561/0600000020.

[30] F. Özyurt, "Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures," *Journal of Supercomputing*, vol. 76, no. 11, pp. 8413-8431, 2020, doi: 10.1007/s11227-019-03106-y.

[31] J. Z. C. Ajmi, S. Elferchichi, A. Zaafouri, and K. Laabidi, "Deep Learning Technology for Weld Defects Classification Based on Transfer Learning and Activation Features," *Advances in Materials Science and Engineering*, vol. 2020, p. 1574350, 2020.

[32] A. S. Elkorany and Z. F. Elsharkawy, "COVIDetection-Net: A tailored COVID-19 detection from chest radiography images using deep learning," *Optik*, vol. 231, no. January, pp. 1-10, 2021, doi: 10.1016/j.ijleo.2021.166405.

[33] A. S. Elkorany and Z. F. Elsharkawy, "Efficient breast cancer mammograms diagnosis using three deep neural networks and term variance," *Scientific Reports*, no. 0123456789, pp. 1-12, 2023, doi: 10.1038/s41598-023-29875-4.

[34] Z. Ahmadchoudhry, H. Shahid, S. Z. H. Naqvi, S. Aziz, and M. U. Khan, "DarkNet-19 based Decision Algorithm for the Diagnosis of Ophthalmic Disorders," *4th International Conference on Innovative Computing, ICIC 2021*, no. November, 2021, doi: 10.1109/ICIC53490.2021.9693030.

[35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520, 2018, doi: 10.1109/CVPR.2018.00474.

[36] Y. Zhang *et al.*, "A Feature Selection Method based on the Pearson's Correlation and Transformed Divergence Analysis," *Journal of Physics: Conference Series*, vol. 1284, no. 1, 2019, doi: 10.1088/1742-6596/1284/1/012001.

[37] I. M. Nasir *et al.*, "Pearson correlation-based feature selection for document classification using balanced training," *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1-18, 2020, doi: 10.3390/s20236793.

[38] N. Sevani, I. Hermawan, and W. Jatmiko, "Feature selection based on f-score for enhancing CTG data classification," *Proceedings: CYBERNETICSCOM 2019 - 2019 IEEE International Conference on Cybernetics and Computational Intelligence: Towards a Smart and Human-Centered Cyber World*, pp. 18-22, 2019, doi: 10.1109/CYBERNETICSCOM.2019.8875656.

[39] S. Ding, "Feature selection based F-score and ACO algorithm in support vector machine," *2009 2nd International Symposium on Knowledge Acquisition and Modeling, KAM 2009*, vol. 1, pp. 19-23, 2009, doi: 10.1109/KAM.2009.137.

[40] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 784 LNCS, pp. 171-182, 1994, doi: 10.1007/3-540-57868-4_57.

[41] T. Tuncer and F. Ertam, "Neighborhood component analysis and reliefF based survival recognition methods for Hepatocellular carcinoma," *Physica A: Statistical Mechanics and its Applications*, vol. 540, pp. 123143-123143, 2020, doi: 10.1016/j.physa.2019.123143.