

Summer 2023

Deep CNN-Based Automated Optical Inspection for Aerospace Components

Shashi Bhushan Jha

Embry-Riddle Aeronautical University, JHAS1@my.erau.edu

Follow this and additional works at: <https://commons.erau.edu/edt>



Part of the [Computational Engineering Commons](#), [Industrial Engineering Commons](#), and the [Other Computer Engineering Commons](#)

Scholarly Commons Citation

Jha, Shashi Bhushan, "Deep CNN-Based Automated Optical Inspection for Aerospace Components" (2023). *Doctoral Dissertations and Master's Theses*. 757.

<https://commons.erau.edu/edt/757>

This Dissertation - Open Access is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Doctoral Dissertations and Master's Theses by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

Deep CNN-Based Automated Optical Inspection for Aerospace Components

by
Shashi Bhushan Jha

A dissertation submitted to the Faculty of
Embry-Riddle Aeronautical University in partial fulfillment
of the requirements for the degree of Doctor of Philosophy in
Electrical Engineering and Computer Science

Embry-Riddle Aeronautical University
Daytona Beach, Florida
June 2023

Deep CNN-Based Automated Optical Inspection for Aerospace Components

by
Shashi Bhushan Jha

This dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Dr. Radu F. Babiceanu, and has been approved by the members of the dissertation committee. It was submitted to the College of Engineering and accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Electrical Engineering and Computer Science.

Radu F. Babiceanu, Ph.D.
Committee Chair

Eduardo A. Rojas-Nastrucci, Ph.D.
Committee Member

Laxima Niure Kandel, Ph.D.
Committee Member

Kenji Yoshigoe, Ph.D.
Committee Member

Prashant Shekhar, Ph.D.
Committee Member

Radu F. Babiceanu, Ph.D.
Interim Chair, Electrical Engineering and Computer Science

Date

James W. Gregory, Ph.D.
Dean, College of Engineering

Date

Christopher D. Grant, Ph.D.
Associate Provost of Academic Support

Date

ACKNOWLEDGEMENTS

First and foremost, I express my gratitude to my Ph.D. advisor, Dr. Radu Babiceanu, for providing guidance throughout my doctoral program. Dr. Babiceanu has been instrumental in offering valuable insights in research, paper writing, and career advancement. Completing this work would have been significantly more challenging without his assistance. Additionally, I extend my thanks to all the dissertation committee members for their valuable suggestions and comments in my Ph.D. journey. I would like to acknowledge Dr. Prashant Shekhar for the fruitful discussions on data science and algorithm development. I am immensely grateful to Dr. Sirish Namilae, Mr. Deepak Kumar, and Mr. Samarth Motagi for generously sharing the initial composite materials image dataset from the Composites Laboratory.

Furthermore, I would like to express my heartfelt appreciation to my parents, late Shyam Kishor Jha and Mohni Devi, for their unwavering support and unconditional love throughout my educational pursuits. Despite my mother not having had the opportunity to pursue higher education, she recognized its significance. I want to express my gratitude to my brother, Mr. Shubham Jha, whose unwavering support has greatly contributed to my educational journey. I would like to take this moment to offer special thanks to my master's advisor and mentor, Dr. Manoj Kumar Tiwari, for consistently guiding me with his extensive research expertise. I also want to extend my deepest appreciation to my wife, Anuradha Kumari, who has been a constant source of support during challenging times, helping me overcome various obstacles. I dedicate this work to my family and friends, as their love and support have been invaluable to me.

ABSTRACT

The defect detection problem is of utmost importance in high-tech industries such as aerospace manufacturing and is widely employed using automated industrial quality control systems. In the aerospace manufacturing industry, composite materials are extensively applied as structural components in civilian and military aircraft. To ensure the quality of the product and high reliability, manual inspection and traditional automatic optical inspection have been employed to identify the defects throughout production and maintenance. These inspection techniques have several limitations such as tedious, time-consuming, inconsistent, subjective, labor intensive, expensive, etc. To make the operation effective and efficient, modern automated optical inspection needs to be preferred. In this dissertation work, automatic defect detection techniques are tested on three levels using a novel aerospace composite materials image dataset (ACMID). First, classical machine learning models, namely, Support Vector Machine and Random Forest, are employed for both datasets. Second, deep CNN-based models, such as improved ResNet50 and MobileNetV2 architectures are trained on ACMID datasets. Third, an efficient defect detection technique that combines the features of deep learning and classical machine learning model is proposed for ACMID dataset. To assess the aerospace composite components, all the models are trained and tested on ACMID datasets with distinct sizes. In addition, this work investigates the scenario when defective and non-defective samples are scarce and imbalanced. To overcome the problems of imbalanced and scarce datasets, oversampling techniques and data augmentation using improved deep convolutional generative adversarial networks (DCGAN) are considered. Furthermore, the proposed models are also validated using one of the benchmark steel surface defects (SSD) dataset.

CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	x
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: RESEARCH OBJECTIVES AND CONTRIBUTIONS	4
CHAPTER 3: BACKGROUND	6
3.1. Defect Inspection Methods Classification	6
3.1.1. Traditional Automatic Optical Inspection Methods	6
3.1.2. Modern Automatic Optical Inspection Methods	10
3.1.3. Architecture of Automatic Optical Inspection Methods	15
3.2. Deep CNN-Based Defect Detection	17
3.2.1. Basic Structure of CNN	17
3.2.2. The Emergence of CNN	18
3.2.3. Supervised CNN	20
3.2.4. Unsupervised CNN	24
3.2.5. Object Detection	27
3.2.6. Pixel Level Segmentation	30

3.2.7. Applications of Deep CNN-Based Defect Detection.....	32
3.3. Open Challenges	35
3.3.1. Challenges In Algorithms	36
3.3.2. Challenges In Applications.....	40
3.3.3. Challenges In High Performance Computing.....	41
CHAPTER 4: METHODOLOGY	43
4.1. Acquisition Of Aerospace Composite Material Image Dataset.....	43
4.2. Classical ML Models	46
4.2.1. SVM.....	47
4.2.2. RF.....	47
4.3. Deep CNN-Based Models.....	48
4.3.1. Improved Resnet50 Model.....	49
4.3.2. Improved MobilenetV2 Model	50
4.4. Hybrid Model.....	51
4.4.1. Improved Resnet50 with Classical ML Models.....	51
4.4.2. Improved MobilenetV2 with Classical ML Models	52
4.5. Imbalanced Dataset Modeling	54
4.5.1. Oversampling.....	55
4.5.2. Data Augmentation using Enhanced DCGAN Model	56
CHAPTER 5: EXPERIMENTAL SETUP AND RESULTS	60

5.1. Datasets	60
5.1.1. ACMID Dataset	60
5.1.2. ACMID Imbalanced Dataset.....	61
5.1.3. ACMID Oversampling Dataset.....	61
5.1.4. Augmented ACMID Dataset using Enhanced DCGAN Model	62
5.1.5. SSD Dataset	63
5.2. Performance Measures.....	64
5.3. Implementation Details.....	65
5.4. Results.....	66
5.4.1. Comparison with Classical ML Models using Distinct Datasets.....	66
5.4.2. Comparison with Deep CNN Based Models using Distinct Datasets	71
5.4.3. Comparison with Hybrid Models using Distinct Datasets.....	76
5.4.3.1. Improved ResNet50 Model with Classical ML Models using Distinct Datasets.....	76
5.4.3.2. Improved MobileNetV2 Model with Classical ML Models using Distinct Datasets.....	81
5.5. Research Implications.....	86
CHAPTER 6: CONCLUSION AND FUTURE WORK.....	88
REFERENCES	90

LIST OF FIGURES

Figure 1: Overall framework of the proposed model..... 3

Figure 2: Generic architecture of an automated optical system..... 15

Figure 3: Basic structure of CNN network 18

Figure 4: Top 1 accuracy of various CNN models in ImageNet challenge over the years20

Figure 5: Number of parameters of various CNN models in ImageNet challenge over the years 21

Figure 6: Error rate of various CNN based model in ImageNet challenge over the years 22

Figure 7: Equipment used for acquiring the images of aerospace components. (a) Autoclave with 3D DIC and (b) autoclave viewports and DIC cameras 44

Figure 8: Layup over cylindrical tool 44

Figure 9: Samples images of aerospace composite components with defects..... 45

Figure 10: (a) Images with defects; (b) Images without defects..... 46

Figure 11: Improved ResNet50 architecture for the defect detection problem 49

Figure 12: Improved MobileNetV2 architecture for the defect detection problem. 50

Figure 13: Proposed hybrid approach including improved ResNet50 and SVM for an AOI problem. 52

Figure 14: Proposed hybrid approach including improved ResNet50 and RF for an AOI problem. 52

Figure 15: Proposed hybrid approach including improved MobileNetV2 and SVM for an AOI problem. 53

Figure 16: Proposed hybrid approach including improved MobileNetV2 and RF for an AOI problem. 53

Figure 17: A generative adversarial network for ACMID Dataset..... 56

Figure 18: Enhanced DCGAN generator block for ACMID Dataset. 57

Figure 19: Enhanced DCGAN discriminator block for ACMID Dataset..... 58

Figure 20: Generated samples of enhanced DCGAN model - (a) with defects and (b) without defects 59

Figure 21: Samples of six kinds of typical steel surface defects on NEU surface defect database. It shows one example image from each of 300 samples of a class 64

LIST OF TABLES

Table 1: Benefits And Drawbacks Of Defect Inspection Methods.....	8
Table 2: The Emergence Of CNN Models Using Imagenet Database	20
Table 3: Advantages And Disadvantages Of CNN-Based Classifiers And Segmentation Networks For Defect Detection	25
Table 4: Advantages And Disadvantages Of Anomaly Detection, GAN-Based, And Hybrid Approach For Defect Classification And Defect Detection	26
Table 5: Performance Comparisons Of State-Of-The-Art Pixel Level Segmentation Model For Automatic Optical Inspection	31
Table 6: Categorization of Defect Detection Application	33
Table 7: All 14 models including classical ML models, deep CNN-based models, and hybrid models.....	54
Table 8: ACMID dataset.....	63
Table 9: SSD dataset.....	64
Table 10: Defect classification results of classical ML models using ACMID Dataset, the best results are highlighted in red color.	67
Table 11: Defect classification results of classical ML models using Imbalanced ACMID Dataset, the best results are highlighted in red color.	68
Table 12: Defect classification results of classical ML models using Over Sampled ACMID Dataset, the best results are highlighted in red color.	69
Table 13: Defect classification results of classical ML models with Augmented ACMID Dataset using enhanced DCGAN model, the best results are highlighted in red color. ...	70

Table 14: Defect classification results of classical ML models using SSD Dataset, the best results are highlighted in red color. 71

Table 15: Defect classification results of deep CNN based models using ACMID half and full dataset, the best results are highlighted in red color..... 72

Table 16: Defect classification results of deep CNN based models using Imbalanced ACMID Dataset, the best results are highlighted in red color. 73

Table 17: Defect classification results of deep CNN based models using Over Sampled ACMID Dataset, the best results are highlighted in red color. 74

Table 18: Defect classification results of deep CNN based models with Augmented ACMID Dataset using GAN, the best results are highlighted in red color..... 75

Table 19: Defect classification results of deep CNN based models using SSD Dataset, the best results are highlighted in red color. 76

Table 20: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF); the best results are highlighted in red color. 77

Table 21: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) using Imbalanced ACMID Dataset, the best results are highlighted in red color. 78

Table 22: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) using Over Sampled ACMID Dataset, the best results are highlighted in red color. 79

Table 23: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) with Augmented ACMID Dataset using GAN, the best results are highlighted in red color. 80

Table 24: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) using SSD Dataset, the best results are highlighted in red color. 81

Table 25: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF), the best results are highlighted in red color. 82

Table 26: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) using Imbalanced ACMID Dataset, the best results are highlighted in red color. 83

Table 27: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) using Over Sampled ACMID Dataset, the best results are highlighted in red color. 84

Table 28: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) with Augmented ACMID Dataset using GAN, the best results are highlighted in red color. 85

Table 29: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) using SSD Dataset, the best results are highlighted in red color. 86

CHAPTER 1: INTRODUCTION

The defect detection problem is extensively applied in automated industrial quality control systems. It is broadly used for quality control to monitor the manufactured components to ensure product quality is maintained. Traditionally, quality inspection in the industry is usually performed by human workers. However, the worker-based inspection operation is tedious, time consuming, and requires experienced workforce. In the current times of cyber manufacturing, smart manufacturing, and industry 4.0, new solutions such as automatic optical inspection (AOI) are essential to enable real-time quality assessment and monitoring. Automatic inspection significantly reduces the workload of human experts, as well as the needed labor cost. In addition, the literature reports it improves the quality of the product [1].

This study focuses on the inspection of aerospace components to identify defects. To identify and evaluate the possible defects, the aerospace components are examined manually by human experts in practice. The inadequacy of this operation comprises that it is tedious, subjective, labor-extensive, inconsistent, and even-biased. To make the operation more effective and efficient, automated optical inspection (AOI) system is preferred to evaluate the aerospace composite components (ACC). Ideally, the AOI system should deliver more consistent, accurate, and unbiased assessment results than manual inspections.

Even though the models and systems have been developed in the past to classify the defects in the aerospace or other manufacturing industry, most of them are traditional

automated optical inspection, namely, statistical methods [2]–[5], spectral methods [6]–[8], model-based methods [9], [10], and learning-based methods [11]–[16]. The first limitation of these approaches is to depend on the expert views and the other drawback is to employ the methods in a two-stage manner. To extract the features, skilled laborers are required to design specific rules and adjust many parameters. In the next stage, classical machine learning or other models are used to identify the defects. These challenges can be overcome by using the recent advancement of computer vision techniques. Deep learning-based models for AOI problems are widely recognized in several domains such as steel surface defects [17], [18], aerospace welding defects [1], [19]–[21], pipe welding [22], laser welding [23], cracks and wrinkle formation [24]–[28], wafer defects [29], [30], concrete structures or building cracks [31], [32], etc. In the existing literature, not many pieces of research have been devoted to detect the defects for the aerospace composite components.

The aerospace composite component is an essential part of the military or civilian aircraft. In fact, 50 percent of the materials are used as composite materials in aerospace and aviation industry. Therefore, this research primarily concentrates on the composite components of aerospace and aviation industry. We propose a novel AOI method, and an efficient defect detection technique based on deep CNN and classical machine learning model to inspect the aerospace composite components automatically using composite images. The overall framework of the proposed model can be noticed in Figure 1. First, the setup of acquiring the sample images of composite components is depicted where the uncured composite materials are laid up over cylindrical tool inside the autoclave. In the top of autoclave, there are two viewports and top of that two 3D DIC cameras are set to

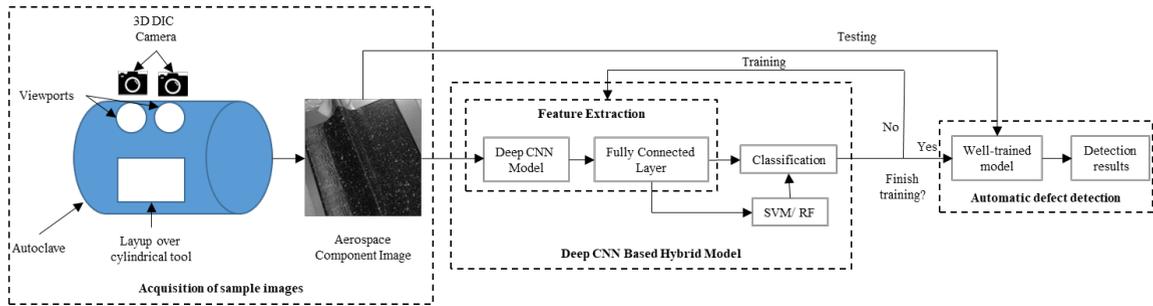


Figure 1: Overall framework of the proposed model

capture the sample images. Second, the acquired aerospace component images are passed to deep CNN based hybrid model to extract the features and classify the defects including classical ML models. Once the training of the model is completed, the well-trained model can detect the defects automatically. To the best of our knowledge, our proposed method that combines the features of deep learning and classical machine learning model has not yet been explored for aerospace composite components.

The organization of the dissertation proposal is outlined here. Research objectives and contributions are presented in Chapter 2. In Chapter 3, background of this research is described where defect inspection methods are categorized, deep CNN-based defect detection methods are described, and open challenges are identified in subsections. Next, methodology is presented where methodology of acquiring images of aerospace composite components, classical ML models, deep CNN-based models, proposed hybrid models, and imbalanced dataset modeling are demonstrated in Chapter 4. Experimental setup and results of this dissertation is provided in Chapter 5. Lastly, conclusion and future works are presented in Chapter 6.

CHAPTER 2: RESEARCH OBJECTIVES AND CONTRIBUTIONS

In this chapter, the research objectives and contributions are described. The main objective of the dissertation is to investigate the literature on defect inspection methods and propose modern automatic optical inspection methods for aerospace components. This dissertation primarily concentrates on deep CNN-based automated optical inspection methods. Major research objectives are as follows:

- (i) Acquire sample images of aerospace composite components for automatic optical inspection.
- (ii) Identify the aerospace composite component defects using classical machine learning models, namely, support vector machine and random forest model.
- (iii) Develop classification models using improved deep CNN-based models such as ResNet50 and MobileNetV2 models.
- (iv) Propose deep CNN-based hybrid approaches that combine the features of deep learning and classical machine learning models for aerospace components.
- (v) Investigate the scenario when defective and non-defective samples are scarce and imbalanced.

Key research contributions are as follows:

- (i) This dissertation investigates existing work on defect inspection methods. In addition, it explores the literature of deep CNN-based defect detection techniques.

- (ii) Acquisition of defective and non-defective samples of aerospace composite components.
- (iii) Hybrid novel approaches are proposed for composite materials image dataset to classify the defects.
- (iv) The defect detection experiments are examined using five-fold cross validation.
- (v) The results are tested considering a two set of datasets for all the models.
- (vi) This work also examines the scarce and imbalanced dataset of aerospace composite components.

CHAPTER 3: BACKGROUND

In this chapter, defect inspection methods are categorized and discussed, along with traditional and modern automatic optical inspection (AOI) methods. Then, the deep CNN-based defect techniques are described, followed by the open challenges of AOI systems presented in the subsequent section.

3.1. Defect Inspection Methods Classification

The past literature covering defect inspection methods can be broadly divided into three parts: (i) manual inspection, (ii) traditional computer vision, and (iii) modern computer vision. To be more general, it can be divided into human and machine inspection. For each defect inspection method, the type of product defects identified, and benefits and drawbacks of the inspection methods are presented in Table 1. Furthermore, traditional AOI, modern AOI, and structure of the AOI are also discussed in the following subsections.

3.1.1. Traditional Automatic Optical Inspection Methods

Traditional AOI technique includes four steps, namely, (a) data acquisition, (b) preprocessing, (c) feature extraction, and (d) defect detection. The first step, data acquisition, is needed to collect the data for the model. It requires lighting and camera for gleaning the data from the production line, if available datasets are not used for the modeling. Secondly, preprocessing step removes the unnecessary parts of the dataset, and the process uses methods such as noise reduction and filtering. For feature extraction and defect detection, there are several methods that can be applied for the automated defect detection. These

methods can be classified into four groups: (1) statistical methods, (2) spectral methods, (3) model-based methods, and (4) learning-based methods [33]. The goal of the statistical methods is to find regions with distinct spatial distribution on the input image. To extract defect features, statistical models based on spatial distribution employ the first order (one pixel), second order (two pixels) statistics and higher order with multiple pixels. Some statistical methods used are co-occurrence matrix [2], autocorrelation [3], histogram properties [4], and edge detection [5]. Spectral approaches transform the signals from the spatial domain to the frequency domain to identify the defect through wavelet transform [6], Fourier transform [7], and Gabor filters [8]. The objective of the model-based approach is to capture the basic characteristics and detect defects by making an image model. Some standard model-based methods for defect identification are autoregressive models [9] and the Markov random field (MRF) models [10]. Learning based approaches first train the model to detect defects and then determine the defects using pattern recognition algorithms, namely, support vector machines (SVMs) [11], k-nearest neighbors (kNNs) [12] and artificial neural networks (ANNs) [13].

Table 1: Benefits And Drawbacks Of Defect Inspection Methods

Defect Inspection Methods	Defects	Benefits	Drawbacks
Manual Inspection	Cracks and wrinkle formation, die casting, scratches, floaters, open circuits, welding defects, steel surface defects, etc.	Diverse set of defects can be identified.	Tedious, time consuming, inconsistent, subjective. Labor intensive, which leads to more mistakes and requires huge human efforts, thus also costly.
Traditional AOI	Steel surface defects, fabric defects, material surfaces, architectural glass materials, etc.	Lessens the labor cost and human errors and reduces the time.	Poor feature extraction, high time complexity. Requires skilled labor for the feature selections and extractions. Requires more processing time and is less reliable and robust.
Modern AOI	Aerospace welding defects, steel, machined, and composite material defects; mobile screen defects, wafer defects, small and large foreign matters, cracks, wrinkle formation, scratches, abrasions, oil stains, dent, chips, light and severe strains, etc.	Offers high detection accuracy, real-time, high speed, more robust. Reduces labor costs significantly and can extract the features by themselves.	Requires large dataset, high-speed processor, GPU, and TPU.

The above-mentioned approaches were researched extensively in literature. To consider a method or combination of these methods to apply on AOI problems depends on

expert views, success results and the dataset availability. For instance, spectral or learning-based methods can be applied to obtain a defect in a patterned surface. To determine deformation on the steel surfaces, statistical or model-based methods can be used. However, these methods are generally employed in a two-stage manner, explicitly, feature extraction and defect detection, and implemented together in a hybrid style model.

For example, [14] considered a fabric defect detection problem and employed AOI methods to extract features using wavelet transform. Then, in the next stage, it classifies the defects using neural network and co-occurrence matrix. In another work, [15] discussed a weld defect detection problem and solved again using the hybrid style model. First, the features of X-ray images were extracted using multiple thresholds, and then the SVM algorithm was employed to classify the defective and non-defective features. Further, Hough transform was also applied for removing the noisy pixels in the defective region and later, the defect was isolated. In another research, principal component analysis (PCA) and ANNs were applied to detect and classify the real time arc welding defects [16]. Using ANN to train the plasma spectrum dataset is reported to have been difficult because of huge number of spectral lines. Thus, PCA was employed first to remove the redundant information and reduce the dimensionality, and then processed data was used to train the ANN model to detect and classify the defects.

To extract features in the traditional AOI methods, a prime role is assigned to human experts who design specific rule and adjust several parameters. Thus, the success behind these methods is highly dependent on experts [34]. Moreover, these methods can perform well under certain conditions, but are sensitive to changes in real world conditions. These drawbacks can be easily overcome by using deep learning. The recent advancements in deep

learning can extract the high-level features from given inputs and can classify the defects without any involvement of human expertise to design features sets manually [35]. Furthermore, these models are highly robust to variations, adaptable, and can also allow detection of several types of defects in various applications.

3.1.2. Modern Automatic Optical Inspection Methods

Deep learning-based methods for automatic optical inspection problems are widely accepted in the research community. Deep learning networks can be primarily divided into two parts: dense networks, if the model is based on fully connected feed forward network, such as deep neural network (DNN), and (2) sparse networks, if the model is sparsely connected, such as deep convolutional neural network (DCNN). Moreover, dense or sparse networks of deep learning methods are mainly classified into three paradigms: supervised learning, semi-supervised learning, and unsupervised learning. Supervised based learning has been the most widely used model, with convolutional neural networks for defect classification and segmentation being employed on numerous occasions. Given a large training dataset, the supervised based models can attain substantial defect detection accuracy.

For example, [36] addressed a surface quality inspection problem of LED chips using computer vision techniques. This work proposed parallel DCNN model for labelled LED chip defects to classify the defects with considerable detection accuracy. In another work, a commutator surface defect detection problem was considered with several defects such as abrade, dark-spot, scratch, and others [37]. To solve the problem, a separable residual CNN-based model has been developed to recognize the defects in a faster way with shallow layers. The solution also reduced the number of parameters of the model due to smaller model size.

The proposed model achieved reasonable accuracy, around 93%. In another research, [24] collected a large dataset of mobile phone glass cover with 16,800 images with labels of dent, scratch, chips, and other defects. The work employed a multi-DCNN to solve the cover glass defect detection problem and attempted to allow manufacturers to set up a fully automated inspection system operated at a high detection accuracy (99%). Nevertheless, collecting a large training dataset and labeling the data requires huge manpower and makes the model expensive. The scarcity of large-labeled datasets can be mitigated by semi-supervised and unsupervised learning models.

Semi-supervised methods can obtain similar or even better results than supervised methods requiring fewer labeled training datasets. In the following four reported research works employing semi-supervised machine vision techniques, the detection accuracy varies from 92% to 99%. In [38], authors considered the automated optical inspection problem with the objective to detect the anomaly. To solve this problem, semi-supervised anomaly detection using dual prototypes autoencoder model was proposed. The model is trained with Aluminum Profile Surface Defect (APSD) dataset and obtained reasonable accuracy. In addition, the results are compared with state-of-the-art algorithms considering four different publicly available datasets, namely, Magnetic-Tile (MT) defect dataset, Road Surface Defect (RSD) dataset, Carpet Surface Defect (CSD) dataset, and APSD dataset. In another study of automated surface inspection problem [18], a generic semi-supervised model is developed considering two public datasets (DAGM and NEU) and an industrial dataset (CCL). The proposed model outperforms the several benchmark algorithms with 95% accuracy.

Two weak-supervision computer vision detection methods were developed with considerable accuracy in [25], where a synthesis algorithm was proposed to simulate a large

dataset of mobile phone screen defects such as light strains, severe strains, scratches, and floaters, so that the challenges of insufficient amount of training dataset can be overcome. Then the model was trained, fine-tuned, and used for defect recognition. In another research [39], automated surface inspection problem was studied, and a model was developed with a new loss function and trained with a small number of defect dataset comprising around 25 defect samples. The approach can identify the anomaly regions at image levels and can address imbalanced data at the pixel level using collaboration learning strategy by utilizing the loss function. The reported detection accuracy of both models with small real-world datasets is 95% and 99%, respectively. Nevertheless, semi-supervised models still need label training samples.

Unsupervised learning is currently one of the most attractive research directions in the machine-learning domain. Unsupervised based models work on unlabeled training samples, and thus do not require manual labeling, which reduces labor cost. The literature suggests that the widely recognized models of unsupervised learning for automatic optical inspection are based on deep autoencoders (AE) and generative adversarial network (GAN). The AE model is a distinctive unsupervised model for high dimensional data comprised of two neural networks namely, encoder and decoder. The encoder extracts the latent features from the input images, while the decoder reconstructs the input image with some loss. GAN is another typical unsupervised learning model consisting of a generative and a discriminative stage. The following paragraphs present a series of research models from the available literature, which are based on AE and GAN approaches.

First, three literature models are explained using both dense and sparse networks, both networks employing autoencoder models. In [40], process pattern recognition problem

was discussed and a deep autoencoder feature learning approach based on stack de-noising AE models (SDAE) was developed for manufacturing processes to learn important features from the process signals. Moreover, the robustness of the model was checked using a large, simulated dataset and Tennessee Eastman process. To extend the work, a multivariate manufacturing process using DNNs was examined to detect other types of patterns such as cycle, trend, etc. In another work [41], wafer defect detection problem was investigated and an improved SDAE-based feature learning approach to recognize the defects was proposed. The detection accuracy of this improved model was reportedly enhanced to 97%. All previous reported research mentioned so far in this literature review were using normal and abnormal samples to train the model, either with supervised or unsupervised learning. But research work [42] studied the automatic optical inspection problem concentrating on solder images of integrated circuits (IC) manufacturing that used only normal samples where one-class based feature learning method was developed to recognize the defects using deep autoencoder. The results of the experiments showed that both sensitivity and specificity were reasonable, around 85% and 7%, respectively. Another one-class-based research was conducted on surface inspection problem focusing on decorated plastic parts to detect the fault with improved area under receiver operator characteristic curve (AUROC) reported around 98% using small datasets [43].

After the widely gained success of AE models, there were several variants of AE models proposed for defect detection problem. One such example is the simple AE model which resulted in an overfit for complex problem models, but the variational autoencoder (VAE) models turned out to perform well on the same complex problems. For example, machine vision inspection system for anomaly detection was examined aiming to identify

the abnormality of Electric Cathode Metal Coating (KTL Coating) [44]. To solve this problem, VAE model was trained with KTL coating datasets at pixel-level and the results of the model showed more robustness than those obtained using simple methods. This type of VAE model can be also used for generating the datasets to work with segmentation-based models for defect detection.

Besides AE based models, GAN based models, such as pixelGAN and cycleGAN, are also currently popular in unsupervised learning. For example, [45] studied the adversarial defect detection problem with the objective of isolating the defects. The research proposed a pixelGAN based model concentrating on semiconductor manufacturing process data to expedite the process. The results outperformed the baseline model such as CenterNet on a real industry dataset. In another research, machine surfaces and medical acne patches inspection problem were examined to detect the surface defects [46]. This work developed a two-phase deep learning model to isolate the faults at pixel level without human annotation. In the first phase, it synthesized the defects and annotated fault pixels in the input image using cycleGAN model. Then, the resultant dataset was used to train the model using U-Net semantic network. The results showed that it can be applied to distinct set of surface inspection problems with considerable detection accuracy, in the range of 95%.

The main drawback of unsupervised models is that it is not as accurate or reliable as supervised learning. However, it significantly reduces the effort for labeling the image and manual annotation at pixel-level in the image dataset, since it does not require a large number of defective samples with the semantic network. In some cases, unsupervised anomaly detection models can be trained with only one-class of dataset, thus, only small size of defective dataset can be required during the testing of the model.

3.1.3. Architecture of Automatic Optical Inspection Methods

Figure 2 describes a generic automated visual inspection system for the external surface defect detection problem. The batches of products move on a conveyor belt with suitable lighting system. In addition, the architecture includes industrial camera sets with proper angles to capture the image data of the product and store on the image captured card, so that it can be transferred on the industrial computer.

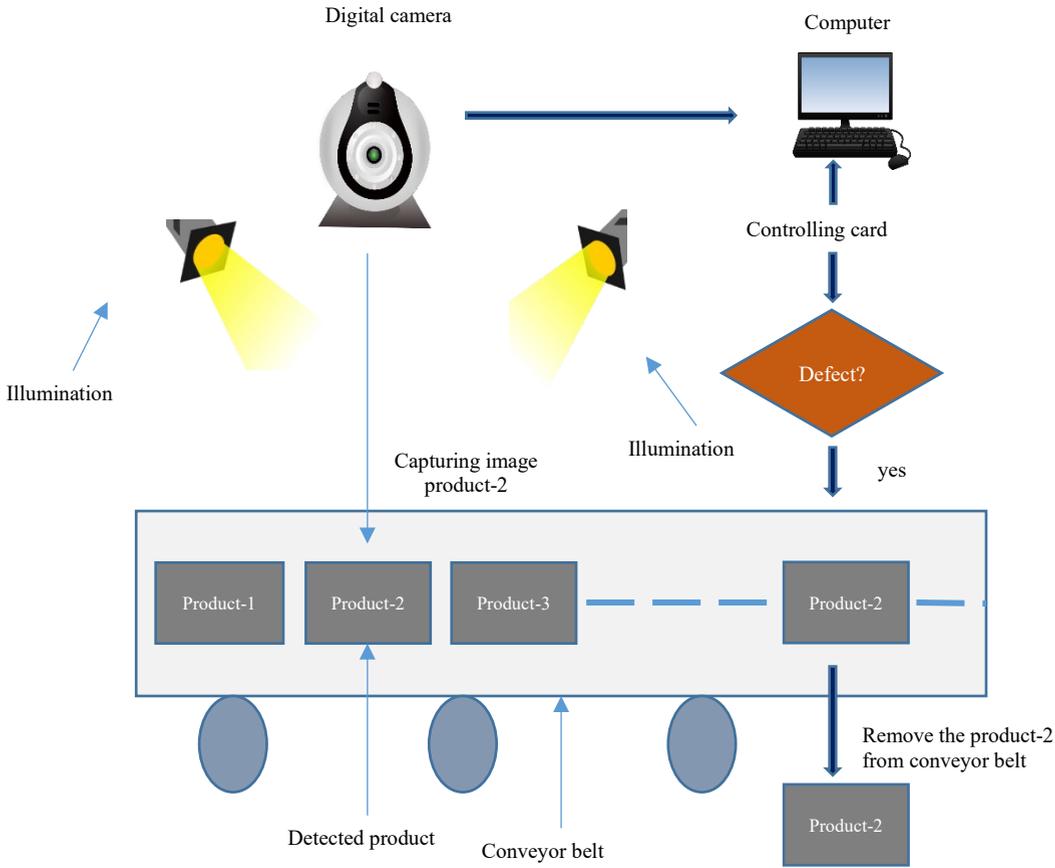


Figure 2: Generic architecture of an automated optical system

The choice of camera depends on the specific application and the requirements for resolution, speed, and accuracy. Factors such as lighting conditions, object size, and

processing requirements also need to be considered when selecting a camera for a particular application. There are several types of cameras that are commonly used in computer vision systems for conveyor applications. These include: (1) area scan camera – it captures an image of a 2D area and is typically used for applications where a relatively high resolution is required; (2) line scan camera – it captures an image of a single line at a time and is commonly used in applications where high-speed inspection is required; (3) 3D camera – it captures depth information, allowing for the creation of 3D images and used in applications where the shape and size of objects are important; (4) smart camera – this camera has built-in processing capabilities and can be used for applications where real-time processing is required. Then, direct communication interfaces between a camera and a computer network interface card (NIC) are commonly used in machine vision applications. These interfaces allow for real-time streaming of image data from the camera to the computer for processing. The most common direct communication interfaces are GigE Vision and USB3 Vision. GigE Vision uses Ethernet technology to transmit image data over a standard network connection, while USB3 Vision uses USB 3.0 technology to transmit data over a USB connection. In addition, an image data buffer, a temporary storage space, is also used to ensure that real-time streaming of image data is reliable and efficient.

Next, the visual system processing continues with the modern machine vision model that is stored on the computer system, and which processes the captured image to make the decision whether the product is defective or defect-free. If the product is defective, then the system sends a signal to the sensor to sideline the product from conveyor belt. Otherwise, it moves the product forward. For internal defect detection, the set up would be different for scanning the product and capturing the X-ray image of the product for decision making.

3.2. Deep CNN-Based Defect Detection

Since the early 2010s, solving computer and machine vision problems using CNN techniques has been gaining momentum. Several computer vision problems are image classification, image segmentation, object detection, feature extraction, and object tracking. However, instead of only classifying the image as normal or defective, the topic of interest here is to localize the defect. Thus, image segmentation and object detection techniques are considered to solve computer vision problems. In this section, the basic structure of CNN is discussed, along with the emergence of CNN models in the past decade. Next, supervised and unsupervised learning of deep CNN-based models are explained with diverse set of computer vision techniques. Then, object detection and pixel level segmentation techniques are specifically reviewed since these techniques cover the state-of-the-art methodologies for automatic optical inspection. In addition, computer vision applications using deep CNN-based models are also surveyed.

3.2.1. Basic Structure of CNN

The CNN architecture has three key components, namely, convolutional layer, pooling layer, and full connected layer that can be observed in Figure 3. Convolutional layers are the essential parts of CNN networks. The function of this layer is to extract high-level features from the input image. Convolutional layers also comprise distinct set of filters that produce a set of feature maps after convolving the kernel over the input image. Convolutional networks stack up the convolutional and pooling layers, with and fully connected layer to complete the construction of the model. Finally, the last layer is output layer that classifies the image or pixel in the classification section.

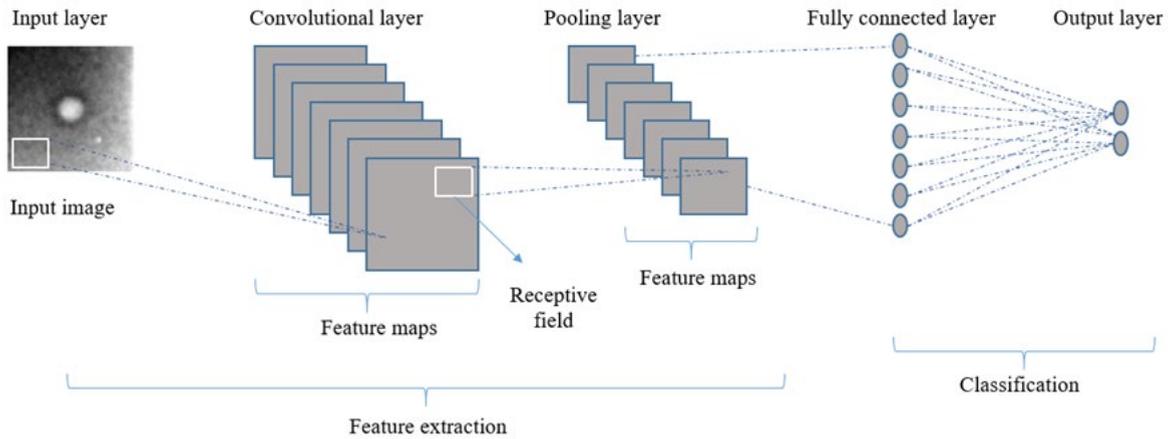


Figure 3: Basic structure of CNN network

3.2.2. The Emergence of CNN

Since 1998, LeCun [47] introduced a CNN architecture based on gradient learning which was implemented on hand digit recognition. Since CNN is a sparsely connected network, it has a couple of benefits such as parameter sharing, and less trainable parameters than traditional fully connected feed forward networks. It primarily includes few basic concepts: shared weights, local acceptance, and pooling layers. Particularly, the shared parameters of CNN reduce the degree of freedom parameters without degrading the solutions quality. It also allows CNN to be implemented by normal gradient decent approach. Thus, the CNN-based model emerges as one of the key algorithms to solve any computer vision problems, and it also has become one of the most active research domains in the machine vision community.

LeNet-5, described in [47] was the first CNN architecture with five convolution layers to recognize the handwritten digits. AlexNet, described in [48] is another CNN

architecture with a few more convolutional layers, which gained higher accuracy in ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Later, GoogleNet [49], VGGNet [50], and ResNet [51] models obtained better accuracy in the ImageNet challenge with thousands of classes and millions of training samples, including millions of parameters. Initially, the researchers were adding more layers and achieving higher accuracy. However, ResNet found that adding more convolutional layers can increase the complexity of the model, while not always achieving higher accuracy. Thus, ResNet introduced the residual concept and found better results. Table 2 shows the Top 1 accuracy, model parameters, and error rate of the CNN model changes in the ImageNet challenge from year 2010 to 2021. Over the years, the percentage of Top 1 accuracy and model parameters in millions of CNN architectures increased tremendously in the ImageNet Challenge, which can be noticed in Figure 4 and Figure 5, respectively. On the other hand, the percentage of error rate of the model decreased significantly, result illustrated in Figure 6. Besides supervised CNN models, convolutional networks also worked with AE and GAN models, namely, convolutional autoencoder (CAE) and deep convolutional GAN (DCGAN), respectively. These algorithms are widely accepted in solving unsupervised tasks in computer vision. CAE was first introduced by [52] to extract the high-level features and for dimensionality reduction. DCGAN was initially proposed by [53] to generate new data with the same distribution of training dataset at image level.

Table 2: The Emergence Of CNN Models Using Imagenet Database

Year	Model	Top 1-Accuracy (%)	Parameters (in Million)	Error Rate (%)	Network
2010	ILSVRC'10 [54]	52.9	-	28.2	shallow
2011	ILSVRC'11 [55]	54.3	-	25.8	shallow
2012	AlexNet [48]	63.3	60	16.4	deep
2013	ZFNet [56]	64.0	-	11.7	deep
2014	VGG19 [50]	74.5	144	7.3	deeper
2014	GoogleNet [49]	74.8	11.2	6.7	deeper
2015	ResNet [51]	81.2	25	3.57	deeper
2016	GBDNet [57]	66.3	-	2.81	deeper
2017	SENet [58]	80.9	1.23	2.25	deeper
2018	MobileNet-V2 [59]	74.7	3.4	-	deeper
2019	FixResNeXt-101 [60]	86.4	829	-	deeper
2020	EfficentNet-L2 [61]	88.4	480	-	deeper
2021	ViT-G/14 [62]	90.5	1843	-	deeper

3.2.3. Supervised CNN

Supervised CNN is one of the most widely used models for defect detection problems. In the past decade, CNN techniques have shown the most promising results for several tasks. In literature, supervised CNN techniques are primarily used for two tasks: defect classification and defect segmentation.

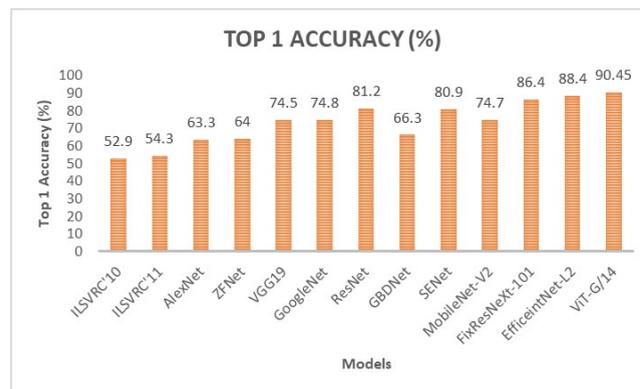


Figure 4: Top 1 accuracy of various CNN models in ImageNet challenge over the years

Defect classification is a classification technique that works at image level and seeks to recognize the type of object in the image. Defect segmentation is a segmentation technique, which works at pixel level and seeks to find the object type in each pixel of the input image. Both the techniques can be applied on the defect detection problem, and the model can be trained either from the scratch or using pre-trained models. Ref. [63] proposed a CNN-based model with Naïve Bayes data fusion technique aiming to classify cracks in components' surfaces of nuclear power plants, where regular inspections are required for safety.

For the surface defect detection problem, [64] developed a deep learning approach with a segmentation network. In the first stage, a segmentation CNN model was trained, and in the next stage, the features extracted from the previous stage were used to train a classification CNN-based model. The final detection task was implemented as a classification task to identify whether an image is normal or abnormal. Another research [65] studied the underwater pipeline damage detection problem using pre-trained MobileNet model for defect classification task. Ref. [66] studied a Mura defect classification problem for a thin film transistor liquid crystal display. To ensure the quality of the displays, the research work developed a new method that blended a CNN feature extractor with a

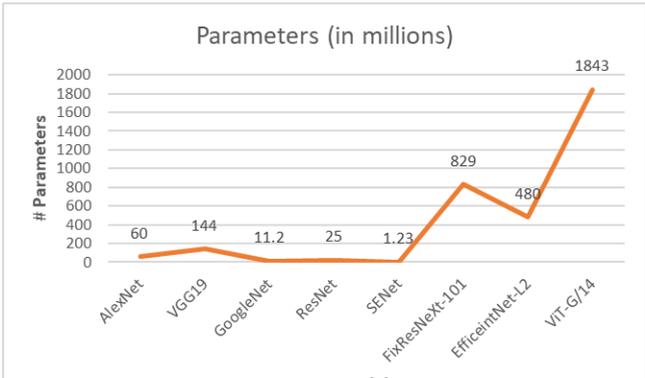


Figure 5: Number of parameters of various CNN models in ImageNet challenge over the years

sequential extreme learning classifier. In another study, X-ray images of castings were considered to identify the defects by employing CNN-based spatial attention bilinear network [67].

End-to-end deep learning approach can be used for both defect classification and defect segmentation. This approach is exemplified in [68] for a steel surface defect detection problem. To identify the exact class and precise location, a standard CNN model combined with multilevel feature fusion network was proposed to accomplish robust classification ability.

Similarly, [34] used a pre-trained CNN-based model for both classification and segmentation task to extract the features from the patches for automated surface inspection. Additionally, an image segmentation model is trained in [69] by using the segmentation network along with random forest techniques.

In 2015, one popular model for semantic segmentation, fully convolutional network (FCN), was proposed with effective inference and learning [70]. Along these lines, [71] developed an FCN architecture to generate the defect segmentation map in one step, identifying the localization of the defects precisely. In the same year, [72] proposed U-Net, which is a CNN-based network, for biomedical image segmentation with different network

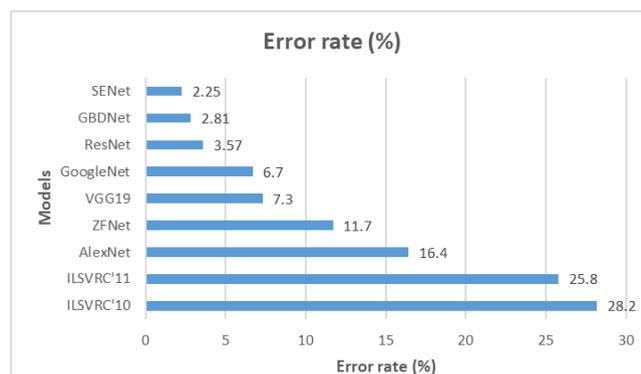


Figure 6: Error rate of various CNN based model in ImageNet challenge over the years

and training strategy relying on vigorous data augmentation techniques. Later, several researchers used this technique in their studies for diverse set of applications and obtained efficient results. In [73], an automated fiber placement defect detection was proposed with end-to-end fashion formulated as a defect segmentation problem. Initially, the artificial training data was generated using a probabilistic model and then, a CNN network motivated by U-Net architecture was trained to identify the defects at pixel level. In another research [74], a novel end-to-end trainable segmentation-based CNN model was developed for crack detection problem using multi-scale feature learning and results were compared with the benchmark U-Net model. For polycrystalline solar cell defect inspection problem, [75] proposed an improved U-Net architecture considering multi-attention networks to classify and segment the complex defects using real photovoltaic images.

In another U-Net based architecture, [76] developed a model, named as MCuePush U-net, comprising of primarily three components: (1) MCue that creates three channel inputs, (2) U-net that learns the informative regions, and (3) Push network that identifies the specific region with bounding boxes. A hybrid approach of regional proposal networks (RPN) to identify defect regions, and modified U-Net architectures to segment the defects at pixel level, were employed for the silicon wafer defect segmentation problem [77]. In most of the research works surveyed, it has been shown that U-Net performs better than FCN for segmentation networks, which is portrayed in Table 3. Furthermore, Table 3 also presents the advantages and disadvantages of CNN-based classifiers and semantic networks for defect classification and defect segmentation, respectively.

3.2.4. Unsupervised CNN

The growing demand of unsupervised CNN techniques for defect detection problems is unparalleled. Unsupervised CNN models have the potential to overcome the challenges of supervised models such as labeling images and annotating pixels and are also capable of isolating the defects with reasonable accuracy, without having the limitations of supervised CNN model. The models can work for image labeling as well as pixel level. In addition, they can classify the defective image and can localize the defective area. Unsupervised CNN can be applied in several machine vision applications such as internal defect detection (non-destructive testing, radiography images, etc.) and external defect detection (steel surface, mobile phone screen, etc.). The literature mainly categories unsupervised CNN technique for defect detection in three types: (1) anomaly detection, (2) GAN-based model, and (3) hybrid models.

Anomaly detection is one of the most common techniques in unsupervised CNN-based models to detect abnormalities. In the specific domain literature, anomaly detection often uses the CAE model and variants of AE model. As an example, [78] studied the machine vision inspection of surface defects problem considering only defect-free samples for model training. To localize the defective area fast and with accuracy, the research proposed a multi-scale fully convolutional autoencoder (FCAE).

Table 3: Advantages And Disadvantages Of CNN-Based Classifiers And Segmentation Networks For Defect Detection

Category	Method	Advantage	Disadvantage
Defect Classification (CNN-based)	CNN + logistic/classifier/attention [53] [54] [55] [56] [57]	Require only image-based labeling, high speed in training and testing.	Require large training samples (defective and defect-free samples); no defect segmentation.
Defect Segmentation (Semantic Networks)	FCN; U-net [58] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68]	Fast training and evaluation; identify and isolate the defective area.	Requisite of pixel-level human annotation; tiresome work, labor-intensive.

Another CAE based model for anomaly detection was developed to identify the abnormality in concrete structures [31]. The study solved the civil infrastructure inspection problem, requiring no labeled images for training, thus it highlighted processing time savings for data labeling. In [79], authors presented a winner-take-all AE method to learn the shift-invariant sparse representations including lifetime and spatial sparsity in each feature map. For semi-supervised and unsupervised anomaly detection, [80] proposed a generalization of deep support vector description (D-SVDD) model. For several years, AE and variants of AE based model were predominant, yet GANs eventually gained the lead in machine vision domain.

The second category of unsupervised CNN models for defect detection is the GAN-based method. The variants of GAN are further classified as GAN synthesis, GAN scoring, and AnoGAN (anomaly detection with GAN). For GAN synthesis, [81] proposed a defect exaggeration model, where GAN is combined with CNN network to generate flawless image and identify tiny surface defects. To improve the defect recognition process, [82] developed a surface defect-generation adversarial network (SDGAN) and applied it on defect-free

images to enlarge the defective dataset. For GAN scoring, [83] trained a GAN-based model using score function to make the anomaly detection more efficient on high-dimensional dataset. In another research for high-dimensional spaces, [84] proposed a novel method for anomaly detection considering GAN networks to search a good representation of the sample in the generator. Research on AnoGAN is advanced in [85], where a fast AnoGAN method is developed to identify defective images and segment the anomalous area.

Table 4: Advantages And Disadvantages Of Anomaly Detection, GAN-Based, And Hybrid Approach For Defect Classification And Defect Detection

Category	Method	Advantage	Disadvantage
Anomaly Detection (CAE-based)	CAE reconstruction; FCAE; D-SVDD [78][31][79][80]	Require only defect-free samples to train the model; no need of pixel-level annotation.	Defect localization is hard; detection of tiny defects is difficult.
GAN-based	GAN synthesis; GAN scoring; AnoGAN [81][82][83][84][85]	Defect synthesis to train the model; no annotation is needed at pixel-level; AnoGAN requires only defect-free images.	Training and validation are slow; require human screening in defect synthesis; imprecise defect segmentation.
Hybrid Models	Bilinear model; CycleGAN; U-net; [86][46]	CNN-based; require only image labelling; no need of human annotated pixels; real-time defect inspection; defect localization.	Requires a substantial amount of dataset; blurry defective area; slow training.

The third category of unsupervised CNN models for defect detection is the hybrid approach. A generic defect detection problem was explored to classify surface defects by extracting features locally and globally using bilinear based model constructed as two symmetric sub-networks based on visual geometry, labeled Double-VGG16 [86]. However,

the proposed model has some limitations, especially on tasks such as localization of defects in complex textures. Ref. [46] developed a novel hybrid approach using CycleGAN and U-Net semantic networks with the objective to detect the pixel-wise defect. More detailed information, including the advantages and disadvantages of anomaly detection, GAN-based, and hybrid approach for defect classification and defect segmentation are presented in Table 4 presented in this section.

3.2.5. Object Detection

In this section, one of the most recent computer vision methods for automatic optical inspection is discussed. In recent years, there has been a significant increase in scholarly research on visual defect detection problems using object detection techniques, namely, YOLO and RCNN based models.

In [87], steel strip production can result in surface defects due to mechanical forces and environmental factors. Identifying these defects is crucial for producing high-quality products, as their presence can cause significant economic losses for the high-tech industry. To address the limitations of current algorithms, researchers developed an end-to-end defect detection model based on YOLO-V3, utilizing an anchor-free feature selection mechanism and specially designed dense convolution blocks to improve feature reuse, feature propagation, and network characterization. Experimental results showed that the proposed model outperformed other comparison models, achieving 71.3% mAP on the GC10-DET dataset and 72.2% mAP on the NEUDET dataset. In [88], the aerospace industry involves assembling many fastening elements, such as bolts, washers, and nuts, which are currently identified manually by humans. However, human error can have a significant impact on efficiency and safety. To address this, a deep learning and image processing approach using

the YOLO-v5 algorithm was proposed to classify these components based on their shape, along with an image processing method to estimate their spatial dimensions, including thread pitch. Despite the challenges, the proposed system achieved promising results. Concrete is a common building material, but strong wind erosion in Northwest China causes damage to its surface, affecting both the appearance and safety of buildings [89]. To identify erosion areas in concrete, a deep learning dataset was established through erosion tests and an improved YOLO-v3 algorithm model was proposed. The model demonstrated more accurate recognition of erosion damage to concrete, achieving accuracy, precision, and map of 96.32%, 95.68%, and 75.68%, respectively.

Ref. [90] introduced a deep learning-based automatic defect detection system called YOLO-attention, which was specifically designed for wire and arc additive manufacturing (WAAM) processes. YOLO-attention incorporates improvements in three object detection models and achieves both speed and accuracy in defect detection. The evaluation on the WAAM defect dataset showed that the model achieved a mean average precision of 94.5 and a frame rate of at least 42 frames per second, demonstrating its feasibility in practical industrial applications. A computer vision pipeline was developed to rapidly analyze electroluminescence (EL) images of solar photovoltaic (PV) modules and identify defects using machine learning models such as Random Forest, ResNet models, and YOLO [91]. The developed models were tested on a dedicated testing set, resulting in macro F1 scores of 0.83 (ResNet18) and 0.78 (YOLO), and were used to analyze 18,954 EL images of a PV power plant damaged in a vegetation fire, finding increased frequency of certain defects on the edges of the solar module closest to the ground after fire.

In [92], Wheel hub defects have complex types, different location and size, making it difficult to establish an accurate detection model. To address this, a wheel nuclear hub defect detection method based on the DS-Cascade RCNN was proposed that uses spatial attention mechanism, deformable convolution, and pruning algorithm to optimize the model and compress the model space without losing accuracy. Experimental results show that the proposed method can effectively detect six kinds of wheel hub defects, and the mean Average Precision (mAP) is 95.49%. In [93] The safety of pipeline transportation relies on Non-Destructive Testing (NDT) to detect weld joint defects. However, traditional manual inspection of X-ray images suffers from accuracy and efficiency issues. To address this, a model integrating Feature Pyramid Network FPN, and a new visual attention mechanism SPAM was proposed, along with a data augmentation method based on geometry transformation. Experimental results show that the proposed model outperforms Faster-RCNN in detecting defects, with a 4.0% increase in mAP value. Damage to metro tunnel surfaces caused by environmental changes, train-induced vibration, and human interference can lead to accidents if not adequately and efficiently maintained [94].

The inspection of these surfaces is challenging due to harsh conditions, such as low light and limited inspection time. To address this, an automatic Metro Tunnel Surface Inspection System (MTSIS) has been developed, consisting of hardware and software components, including a high-speed image capture system, image pre-processing methods (contrast enhancement and stitching), and a defect detection method based on a multi-layer feature fusion network. Practical experimental results demonstrate the effectiveness of the proposed MTSIS in detecting defects on metro tunnel surfaces.

3.2.6. Pixel Level Segmentation

This section reviews the latest developed computer vision techniques for automatic optical inspection. Scholarly contributions for the pixel level segmentation have been significantly growing in the last few years. The models developed in the articles referenced in the previous sections are mostly evaluated through accuracy metrics. The newly developed pixel level segmentation models are highly encouraged to be compared with the mean intersection over union (MIoU) or Dice coefficient, instead of accuracy metrics. The performance comparisons of recently developed pixel level segmentation models for AOI are presented in Table 5 and include MeanIoU and Dice Coefficient evaluations. The reviewed articles are mainly categorized as supervised and unsupervised deep CNN based segmentation models. In addition, a few papers focus on small or micro defects datasets.

Using RSDD dataset, [95] and [96] presented segmentation-based model. Ref. [95] developed a pixel level segmentation network including deep feature fusion, multi-level feature aggregation module, and multi-branch decoder. Ref. [96] proposed an NDD-Net model to create an end-to-end defect segmentation scheme comprising of attention fusion block to obtain discriminative features and improve the performances. The performance achieved by these two models show a MIoU of 0.85 for [95] and a Dice Coefficient of 0.835 for the model presented in [96], respectively. Furthermore, [97] introduced the UCF EL defect dataset and proposed a semantic segmentation model to classify the five defects with MIoU of 0.573 and pixel-level accuracy of 95.4%.

Table 5: Performance Comparisons Of State-Of-The-Art Pixel Level Segmentation Model For Automatic Optical Inspection

Model	Dataset	Accuracy (%)	MIoU	Dice Coefficient
Pixel level segmentation on RSDD [95]	RSDD	-	0.850	-
NDD-Net [96]	RSDD	0.997	-	0.835
Semantic Segmentation of EL Images [97]	UCF EL Defect Dataset	95.4	0.573	-
U-Net GMP + SCL _{Dice} [98]	Kolektor	-	0.56	-
Regression based pixel segmentation [99]	DAGM	-	0.845	-
TAS2-Net [100]	DAGM	-	0.869	-
Automatic deep segmentation [101]	GDXray	0.998	-	0.854
Improved super-pixel segmentation model [102]	Machine surfaces	91.11	-	-
CycleGAN [46]	Machined surfaces	95	0.71	-
Pixel-wise semi-supervised model [103]	FID	91.85	0.825	-

In [98], authors presented a U-Net GMP method comprising of SCL_{Dice} using Kolektor dataset with MIoU of 0.56. Ref. [99] developed a regression-based pixel segmentation model using DAGM dataset to localize the defects with MIoU of 0.845. In another work, [100] proposed a TAS2-Net model for small surface defects with same dataset but slightly better MIoU of 0.869. To address the class imbalanced or micro defects issue, an automatic deep segmentation model is proposed [101] with an attention-guided segmentation network for pixel level welding defects with decent Dice Coefficient of 0.854. To classify the machined surface defects at pixel level, [102] developed an improved super-pixel segmentation model. In contrast, [41] proposed the unsupervised segmentation CycleGAN model to segment the machined surface defects with MIoU of 0.71. Ref. [103] authors

developed pixel-wise semi-supervised segmentation model with multi-task mean teacher using fabric image dataset. The MIoU performance metric reported is 0.825. Future research in this area should be coordinated to compress models in such a way so that to yield a more lightweight model while ensuring high detection accuracy.

3.2.7. Applications of Deep CNN-Based Defect Detection

Computer vision applications have been widely adopted in quality inspection problems and mostly solved using deep CNN-based model. They can be mainly categorized into internal surface inspection and external surface inspection, as portrayed in Table 6. Internal defect detection is mostly used in aerospace welding, pipe welding, laser welding, and other similar welding operations. Welding is needed in various manufacturing industries to join two distinct parts into one component, many of these processes being used in aerospace industry. In certain unanticipated cases, defects might occur in aerospace-welded components that can increase the risk of accidents. To ensure the quality and safety for aerospace industry components, several researchers solved this problem using deep CNN-based techniques, such as X-ray images of aerospace welds [1],[19],[20].

In addition, [21] considered X-ray images of aerospace composite materials to recognize the defects using transfer learning model. Covering different industries, [22] studied the petroleum pipelines welding defect detection problem using conditional GAN and transfer learning with augmenting the X-ray images, while [104] presented the Keyhole Tungsten Inert Gas (TIG) welding type inspection using ResNet models to identify the different states of welding. To inspect the laser welding defects of safety vents on power battery, [23] used a pre-trained SqueezeNet model to identify the abnormalities in the images.

The SqueezeNet model is a CNN architecture that was reported to attain a high accuracy on ImageNet challenge, even though it uses a small model and low number of parameters.

External surface defect detection is primarily classified into textured and patterned surfaces. First, textured surface defect detection methods are applicable in several domains. This literature review divides textured surface applications into three parts such as 3C products, construction, and miscellaneous. The 3C products include mobile phone-type devices, LCD displays, and printed circuit boards (PCB) components. To improve the quality of 3C products, surface defect detection of mobile phone screens is one of the essential tasks. As an example, [26] studied machine vision problem of mobile phone screens using a novel deep learning algorithm to extract the features and classify the defects. In addition, a weak-supervised defect detection method was proposed for mobile phone screen defects such as scratches, floaters and strains [25]. Ref. [24] addressed the smart factory display manufacturing for mobile screens and developed a multi-deep learning neural network to identify the defects.

Table 6: Categorization of Defect Detection Application

		Defect Detection		
		Internal Surface	External Surface	
		Radiographic Images of	Textured	Patterned
		3C Products	Construction Materials	Miscellaneous
Aerospace welding [1], [19]–[21]				
Pipe welding [22]				
Laser welding [23]				
Mobile phone screen [24]–[26]				
LCD [27], [28]				
PCB [22], [104]				
Glass panels [108]				
Concrete structures / building cracks [31], [32]				
Machined surfaces [46], [109]				
die casting [110]				
Steel [17], [18], aluminum profile [38], polycrystalline alloy [111]				
Bottle [112], wood [27]				
Wafer surface [29], [30], [41]				
Fabric [105]–[107]				

With the goal to enhance the quality of the LCD display of 3C products, [27] addressed the homogeneously structured LCD display defect detection problem using autoencoder based anomaly detection technique. While the developed model includes a small size dataset, the experimental results are significant (100%). In another study covering small industrial image dataset of LCD, [28] developed a deep CNN-based method for defect detection. To ensure the quality of 3C products, one of the key issues is to improve the external surface defect detection of PCB. Within the defect detection domain, [113] addressed an automatic inspection system for PCB board using effective self-adaption methods to identify the PCB defects with significant detection rate. In another study, also covering PCB defects, [114] proposed a deep learning method using faster region-based CNN (R-CNN) and feature pyramid network to recognize the PCB surface defects with mean average precision (mAP) of 95%. The mAP is calculated as the average AP values of the number of defects in a candidate area.

Defect detection of construction materials in the civil infrastructure domain is another vital application of machine vision. Construction materials and their typical defects include glass panels that exhibit scratches and concrete structures that exhibit cracks, among others. Glass pieces are the key components of building materials. For quality assurance of glass products, [108] studied the automated scratch detection of transparent glass components. In order to identify the scratches on the surface, this study developed a deep learning approach using mask and region-based CNN (Mask R-CNN) with a significant reported accuracy of 94%. Another important computer vision application in the construction materials domain is the study of concrete structures defects such as building cracks. The data of building surfaces can come from structures such as bridges, houses, roads, and dams. In [32], the building

cracks defect detection using FCN, R-CNN, and richer fully convolutional networks (RFCN) were compared and evaluated for picture performance detection and comprehensive assessment, with RFCN found to exhibit the best outcomes. Ref. [31] attempted to tackle the anomaly detection problem of concrete structures using CAE with defect-free images. Other machine vision applications for textured surface defect detection are reported for machined surfaces [46], [109], die casting [110], steel surfaces [17], [18], aluminum profiles [38], polycrystalline alloys [111], bottles [112], wood [27], and wafer surface [29], [30], [41].

The last defect detection application category identified in the survey is the patterned surface defect detection. This is another widely researched area within the larger external surface domain, with the main application being fabric defect detection. Fabric inspection system plays a key role for quality assurance in textile manufacturing. There is an ever-growing demand in the textile factory to substitute the human-intensive quality inspection performed with naked eyes by an automated inspection system. This task compelled researchers and practitioners to develop deep CNN-based approaches to isolate fabric defects [105]–[107]. For all machine vision applications in this area, the results of detection accuracy vary from 88% to 99%.

3.3. Open Challenges

For the last few decades, many researchers and practitioners have studied the defect detection problem for quality control and assurance. As such, the demand of automated inspection systems for quality control in the manufacturing industry grows by the day. Over the years, researchers proposed a diverse set of deep learning techniques to isolate the defects. However, there are still numerous challenges left to tackle in this domain. This section

presents the identified open challenges which are categorized as follows: (1) challenges in algorithms, (2) challenges in applications, and (3) challenges in data processing on high performance computing systems.

3.3.1. Challenges in Algorithms

This section covers the many challenges explained at the algorithmic level, such as defect inspection methods, supervised CNN-based methods, unsupervised CNN-based methods, etc. Defect inspection methods include manual inspection, traditional AOI, and modern AOI. Each inspection method brings forth various challenges. Manual inspection has specific well-known drawbacks. First, it requires large number of human experts for the inspection, which significantly increases the labor cost. Being labor-intensive work, manual inspection leads to many mistakes. Also, manual inspection is time-consuming, inconsistent, and subjective. To overcome these disadvantages, traditional AOI methods have been developed where image processing techniques and shallow ML algorithms were used to reduce the labor cost, human errors, and inspection time. However, many challenges still exist when employing the above-named methods. For example, traditional computer vision techniques have poor feature extraction and huge time complexity. Thus, the shallow ML techniques require human experts to find the specific features to feed the model.

Though some of the traditional machine vision methods achieve detection accuracy for one defect pattern as high as 99%, they still do not work as expected with multiple patterns. In contrast, modern AOI methods have been consistently used by researchers and manufacturing industries for addressing defect detection problems. These techniques are known to attain high detection accuracy, increased real-time model robustness, reduced labor

costs, and good performance for high-level feature extractions. Still, these traditional machine vision methods were identified to have challenges such as the need for substantial amount of data to train the model, performance challenges for hyper-parameters tuning, and limitations on the high-speed processor, GPU, and TPU for execution of complex models.

Modern computer vision techniques for defect detection are deep CNN-based models. For supervised deep CNN models, a large, labeled image dataset is required to train the model to attain considerable detection accuracy, since the detection accuracy of the model highly depends on the quality of the dataset. Thus, acquiring and labeling datasets are essential, but it still carries a series of challenges. First, acquiring a large dataset is one of the major challenges for researchers and practitioners. Capturing the images in the industry could run into limitations such as non-uniform illumination, motion blur, and camera noise. The quality of the images also depends on the manufacturer's standards. Then, the challenges also depend on the image dataset acquired in different applications. As an example, obtaining images by X-ray can run into issues such as noise and defects in the images are very subtle, so background differentiation makes it difficult to process the images.

For steel surfaces, usually the occurring of defects has a very low probability as well as they are visually indistinctive. Therefore, the defective samples are limited in number, which makes it hard to represent the distribution of dataset, which in turn brings even more challenges. Generally, scarce training samples can and many times will lead to a poor generalization ability of the model. One of the major challenges with deep learning models is the problem of generalization. While deep learning models are known to perform well on training data, they may not generalize well to new, unseen data. This is because deep learning models often have a large number of parameters that can be tuned to fit the training data very

closely, sometimes resulting in overfitting. Overfitting occurs when the model fits the noise and idiosyncrasies of the training data rather than the underlying patterns and relationships. Another issue related to generalization is the problem of dataset bias. If the training data is not representative of the target population or contains systematic biases, the model may not be able to generalize well to new data. To address these issues, various techniques have been developed, such as regularization, early stopping, data augmentation, and transfer learning. These techniques aim to improve the generalization performance of deep learning models by reducing overfitting and increasing the robustness of the model to new, unseen data.

Secondly, labelling images is another difficult task to complete and as an undesired consequence it increases the labor cost. In addition, it can only classify images without being able to locate the defective area. To alleviate this issue, several scholars developed CNN-based semantic networks to account for defect localization, such as FCN and U-Net networks. While addressing the defect localization, these methods also have challenges. One first challenge comes from the labor-intensive process to perform pixel-wise human annotation, which again increases the labor cost. A second labor- and time- intensive task is to annotate the data at pixel-level. To overcome these challenges, computer vision researchers proposed unsupervised deep CNN-based models that do not require any human labeled or annotated data and used them to identify the image and/or localize the defective area. Unsupervised anomaly detection techniques and the GAN-based models have the advantage not only to require normal samples to train the model, but also there is no need of pixel-level annotation. Even with these advantages, there are challenges still yet to be addressed, such as improvements in the accuracy of defect localization and in the reliability of unsupervised models, which are lower than that of supervised models.

Other challenges related to algorithms are the complicated process of selection of specific algorithms for certain defect detection problems. The size of training samples influences many aspects of the deep CNN models during learning process, such as the detection accuracy to be achieved, training time of the model, number of features, distribution of the dataset, and the computing system required for the model. Besides algorithm selection, tuning the hyper-parameters is another key challenge for deep CNN-based defect detection algorithms. Some of the hyper-parameters' tasks include determining the number of network layers, finding the proper number of filters, and setting the filter size for each layer, selecting the stride size and the pooling type, and choosing the number of neurons and activation functions for each neuron. To achieve reliable results for large models, the deep CNN-based models challenge list is completed by another series of tasks that need to be addressed, namely selecting an optimizer for the model from stochastic gradient descent (SGD), performing the adaptive momentum estimation (Adam), setting the adaptive gradient learning algorithm (Adagrad), and training a significantly large number of parameters. These challenges require an efficient hardware infrastructure to execute the complex model.

The challenge with the explainability of CNN network decisions is that they operate as black boxes, meaning that it can be difficult to understand how they arrive at their conclusions. CNNs learn to recognize patterns in data through layers of abstraction, and the final decision is often made based on a combination of these patterns. However, it can be challenging to determine which patterns are being used in the decision-making process and why they were given importance. This lack of transparency is particularly problematic when it comes to safety critical applications such as medical devices, aircraft systems, autonomous

vehicles, etc., where it is crucial to understand how a decision was made. Without a clear explanation, it may be challenging to trust the results of a CNN network, and it could lead to serious consequences if incorrect decisions are made. Several techniques have been developed to improve the explainability of CNN network decisions, such as visualization of feature maps and saliency maps, or the use of attention mechanisms to focus on specific regions of an image. However, further research is needed to fully address this challenge and make CNNs more transparent and interpretable.

3.3.2. Challenges in Applications

This section discusses the key applications' challenges, such as those encountered in databases for automated optical inspection. The process of acquiring a database is an essential step for all deep CNN-based models. But, as in any other application, databases come with a series of challenges such as data privacy, data cleanness, data labeling and annotation, and data sharing. Data privacy is one of the major challenges for machine vision-based defect detection methods. Regulations in force in several industries, such as aerospace manufacturing domain, ask that the dataset is kept private and not uploaded in the public domain due to security reasons. Cleaning the dataset is very much a requirement before feeding it into a model. Particularly, cleaning the noises in the dataset improves the quality of the data and makes it easier to explore and understand the model. Also, before training the model, anomalous data needs to be removed, the images need to be resized, and image resolution needs to be fixed.

Data labelling and annotation is another essential step in preparing the dataset for the model and it requires significant human efforts and expertise. Lastly, but not the least

important, researchers and practitioners do not share data easily, which makes it difficult for other interested parties to collect the dataset and train their models. In certain manufacturing industries, it is relatively easy to gather non-defective samples in the early phase of production but collecting enough defective samples for the robust model is hard, since occurrence of defective samples is rare compared to non-defective ones. Automated machine vision-based non-destructive testing (NDT) methods usually face two challenges: low availability of defective images and lack of precise annotation of defective samples. Insufficient abnormal images create an imbalanced dataset and result in inaccurate representation of the distribution of all defective samples, which makes the training process difficult. Furthermore, the low number of defective samples, with low contrast in many cases, causes ambiguity in both defective samples and normal samples.

3.3.3. Challenges in High Performance Computing

To solve the complex defect detection problems using deep CNN-based models, high performance computing (HPC) is essential. Most of the deep CNN models researchers use high-speed processors, substantial amount of graphic processing unit (GPU), and tensor processing unit (TPU) via cloud computing. Cloud computing facilitates HPC by providing huge computational capabilities to individual researchers and organizations who might have insufficient hardware infrastructure in-house to train complex models. Amazon web services (AWS) provides researchers and practitioners with the power to create HPC clusters on demand, train and test ML models, gain valuable insights on complex models, and improve their productivity. But since nothing is free, these services come with external constraints and limitations such as cost, security, data transfer, performance, to name a few. Cost-

management to build and use cloud HPC systems is a major concern for most organizations. Data security on clouds is another major concern for deep learning researchers and practitioners. In addition, to run models on the cloud, proprietary data must be moved into the cloud, which is many times a no-solution challenge for many organizations. Lastly, performance in the cloud is another major concern since most of the deep learning scholars expect high performance from HPC systems, but the performance may be reduced due to inter-connect latencies and outside network limited capabilities for data transfer.

Furthermore, federated learning has emerged as a promising approach for defect detection in manufacturing, as it allows multiple parties to collaborate and jointly train a machine learning model without sharing their private data. However, there are several challenges that need to be addressed when applying federated learning to defect detection, including data heterogeneity, communication overhead, privacy concerns, and quality control. While federated learning holds great promise for defect detection in manufacturing, these challenges need to be carefully addressed to ensure the performance, privacy, and reliability of the system [115].

CHAPTER 4: METHODOLOGY

In this chapter, the acquisition of aerospace composite components image dataset is explained. In addition, the classical ML models and deep CNN based models are described for the collected dataset. In last, the proposed hybrid models including deep learning and classical ML models are presented.

4.1. Acquisition of Aerospace Composite Material Image Dataset

The aim of this research is to classify the defective aerospace components automatically, thus in accordance with the objective of this work is to acquire the ACMID images to train the proposed deep CNN based hybrid model. Prepreg composites are being rapidly used not only in aerospace but also in automobile and civil applications. It decreases the risk of poor impregnation in deformation mechanisms and manufacturing processes and is usually cured in the autoclave under high pressure and temperature conditions. However, the kinematics of aerospace composite layer interfaces significantly impact the manufacturing process. And the formation of defects, namely, delamination and wrinkles are frequently encountered in aerospace manufacturing. Particularly, the wrinkles are occurred due to inter-ply restricted motion and shear deformation [116]. The equipment (Figure 7) located in the Embry-Riddle's Composites Laboratory is used to acquire the images of aerospace composite components. In Figure 7(a), one can note that autoclave is set up to cure the composite plates. To acquire the images for the deep learning model, there are two viewports and for each viewport a 3D DIC camera is placed on the top of the

autoclave that can be seen in Figure 7(b). Furthermore, actual composite layup over cylindrical tool with vacuum source and vacuum probe can be noted in Figure 8.

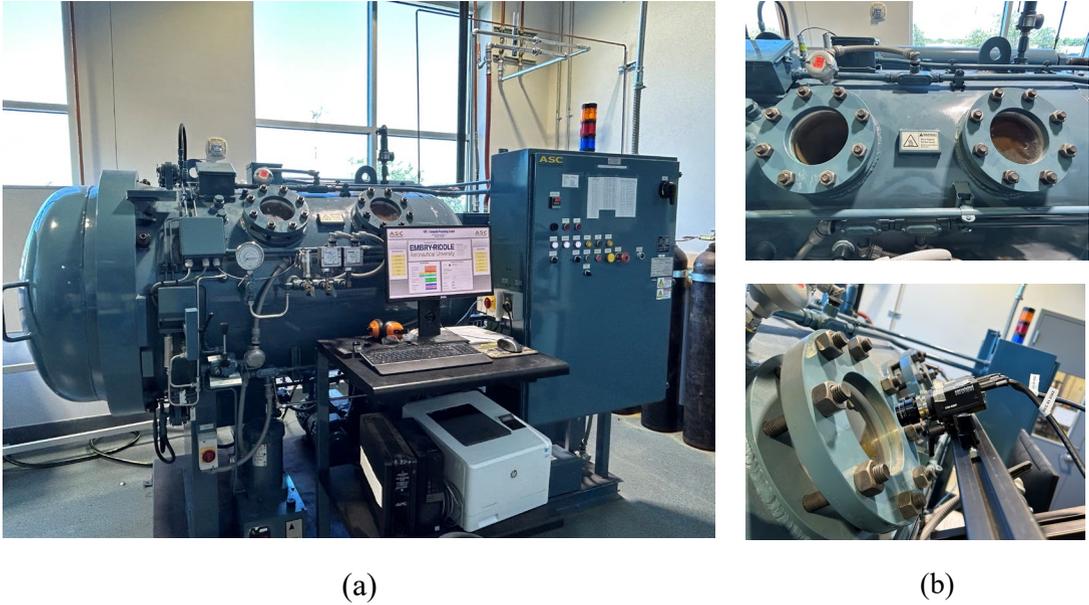


Figure 7: Equipment used for acquiring the images of aerospace components. (a) Autoclave with 3D DIC and (b) autoclave viewports and DIC cameras

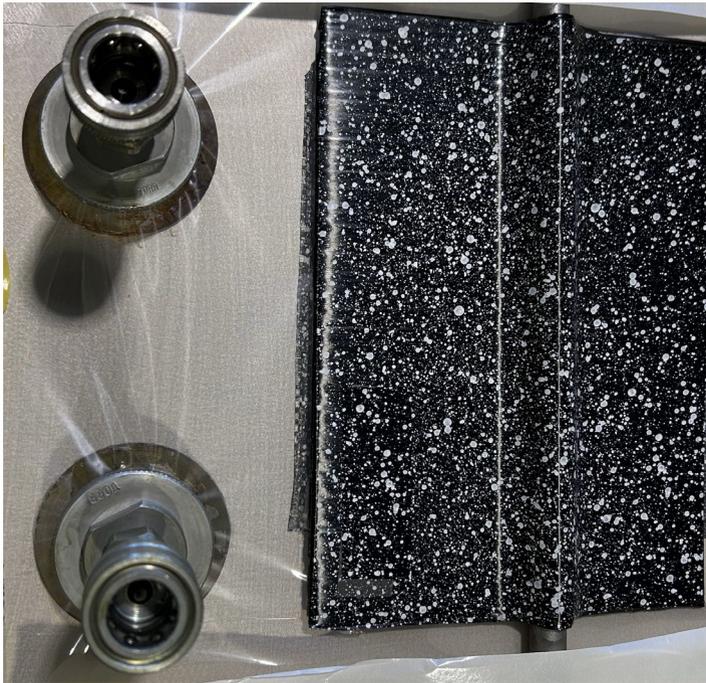


Figure 8: Layup over cylindrical tool

This research focuses on the supervised learning model, freezing most of the learned weights of the parameters and considering less parameters as trainable of the deep CNN based model since the dataset size is not large. The proposed model can work with decent amount of dataset size around 800 images, including 331 of defective images and 460 of non-defective images. The samples of the defective and non-defective images can be seen in Figure 9 and Figure 10. In Figure 9 (a) and (b), the defective areas are highlighted with rectangular boundary. Similarly, the third sample of the defective images is highlighted in a square boundary, but the tiny region is enlarged at the top of the image that is illustrated in Figure 9 (c). A greater number of samples are shown in Figure 10 (a) images with defects and in Figure 10 (b) images without defects.

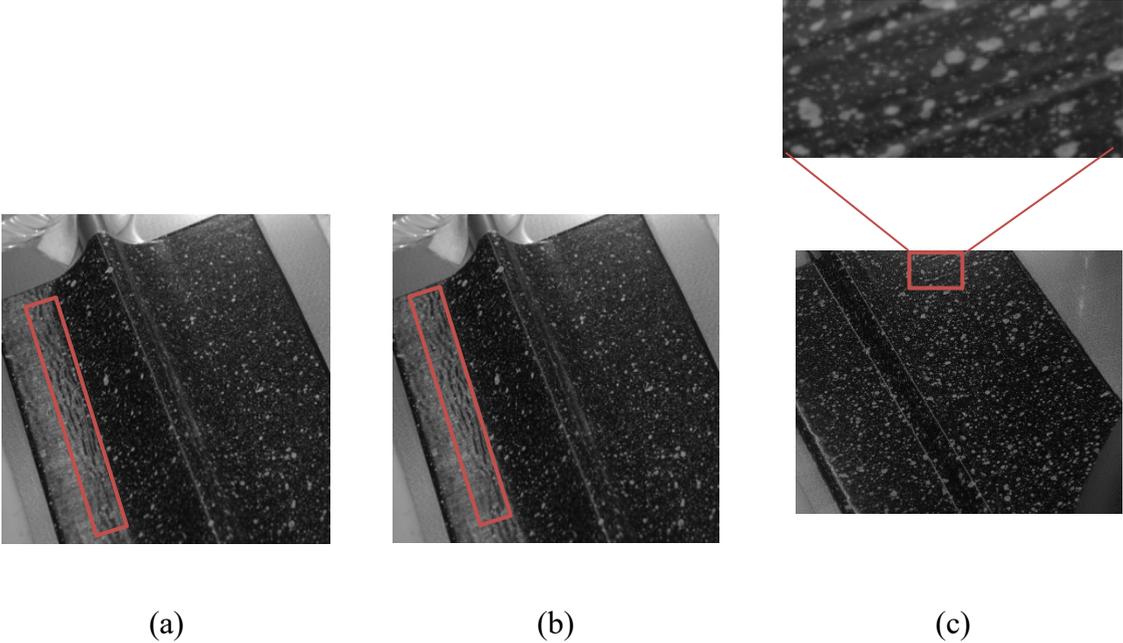


Figure 9: Samples images of aerospace composite components with defects

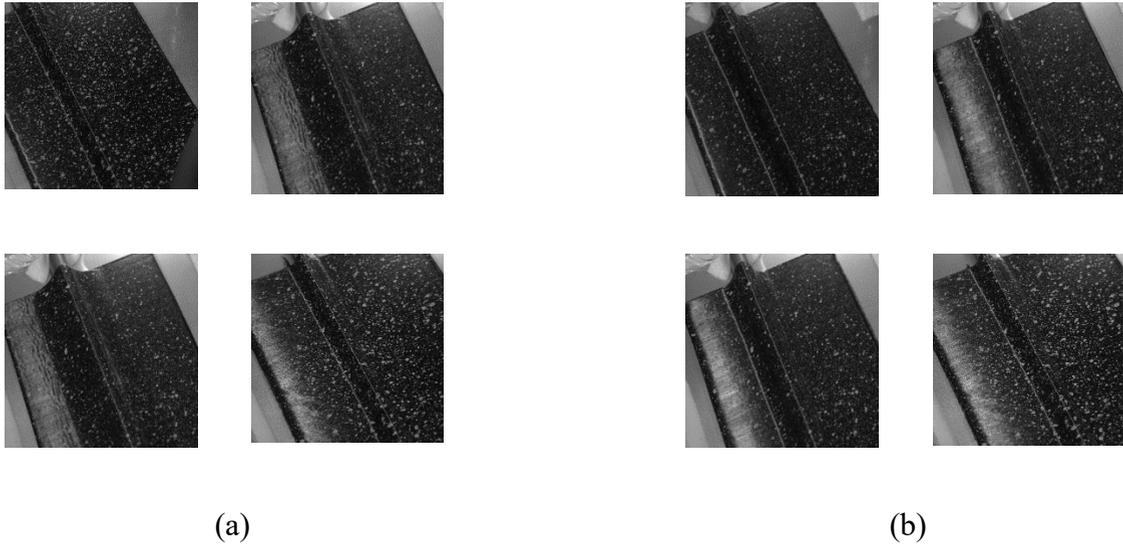


Figure 10: (a) Images with defects; (b) Images without defects

4.2. Classical ML Models

Classical machine learning models are gradually evolving where the models learn to solve a problem by given samples with desired outputs for each instance. In this work, the defect detection problem is considered as a supervised classification task aiming to classify the defective and non-defective images. Several established methods for supervised classification learning are Logistic Regression, Support Vector Machine (SVM), Artificial Neural Network, Decision Tree, Random Forest (RF), etc. However, SVM and RF are well-recognized ML models to identify the defects [11], [19]. In addition, the ACMID dataset has never been used before for any classical machine learning models or deep learning models. Thus, the generated aerospace component images need to be used for well-established classical ML models before the proposed model. And their results are also expected to be compared with advanced deep learning models as well as proposed hybrid models.

4.2.1. SVM

SVM [117] is one of the classical ML models that examine the dataset for classification as well as regression analysis. Moreover, SVM can perform linear and non-linear classification with the help of kernel trick. SVM is effectively used and widely accepted in various domains such as text and defect detection, text, and hypertext categorization, recognizing hand-written characters, biological and other sciences. For this research, SVM is applied for defect detection problem using ACMID dataset and the model is trained with distinct kernels, kernel coefficients, regularization parameters, etc. Furthermore, the grid search cross validation is also used to determine the best hyper parameters for this SVM model.

4.2.2. RF

RF [118] is one of the most important classical ML models and its essential unit is decision tree including various good features, namely, fast prediction, less computational complexity. To build each tree, samples are drawn from the training examples randomly. In addition, the model works on feature randomness to construct a tree; thus, forest of trees can be uncorrelated, and the outcome can be superior to any single decision tree. To enhance the randomness in the model, a subset of features is split at each node of the tree [119]. Although, the single decision tree can render a high variance and leads to overfitting of the model, RF is endowed with two layers of randomization to reduce the variance issue significantly.

Though the classical ML models are reliable and efficient for traditional AOI problems, it requires human experts to design a specific rule and adjust several parameters

to extract the features. In addition, these models can work in certain condition, but it is highly sensitive to changes in a real-world scenario, thus the success of these methods highly relies on experts [34]. These shortcomings can be easily overcome by deep learning methods. The advancement of DL methods can extract high-level features from given samples and recognize the defects without manually designing the features [35].

4.3. Deep CNN-Based Models

In the past decade, computer vision has been transformed by the emergence of deep learning algorithms. The advancement of hardware (CPU, GPU, TPU) enables powerful and large-scale computations and makes it possible to train complex models. Deep CNN based models are primarily designed for the image dataset thus it highly fits for ACMID dataset to address the defect detection problem. However, these deep learning models require large amounts of image samples to train the models otherwise it often raises the overfitting issue. Particularly, acquiring the huge number of aerospace composite images is highly unlikely in aerospace manufacturing since it is expensive and time consuming. In addition, obtaining defective images in the initial phase of the project is a more difficult task. Nevertheless, transfer learning addresses such issues thus this research employs the pretrained DCNN models that are already trained on ImageNet dataset for different classification tasks. These pretrained weights of parameters are transferred for the target task. However, there are 1000 classes in the ImageNet dataset, therefore, the last layer of the pretrained model is dropped and two new layers are added in the last. Finally, the deep CNN based models, namely, ResNet50 [51] and MobileNetV2 [59] are trained with

updated architecture using ACMID dataset to extract the features at high level and identify the defects in the end.

4.3.1. Improved Resnet50 Model

ResNet50 architecture [51] is widely used in the computer vision community since this model is considered as a benchmark model. In this research, ResNet architecture with 50 layers is considered as it is, except the last layer of the original architecture is dropped and two new layers are included that can be noticed in Figure 11. Therefore, the total number of layers in this updated architecture is 51 and the second last layer is a fully connected layer with 32 neurons which is presented in Figure 11. And the last layer has only one neuron since this defect detection problem addresses a binary classification problem to classify the image as defective or non-defective.

This ResNet model can be trained in several ways: (i) consider only the last layer as trainable but in this case the model regularly underfits since the model has very few parameters to learn; (ii) train the model where all the layers considered as trainable, however, in this situation, the model often overfits as lack of samples; (iii) fine-tuned model that opts a few layers as trainable and others as non-trainable and it also tackles the underfitting and overfitting issue. Therefore, a fine-tuned model and model with all layers is employed to identify the defects for the aerospace component defect detection problem.

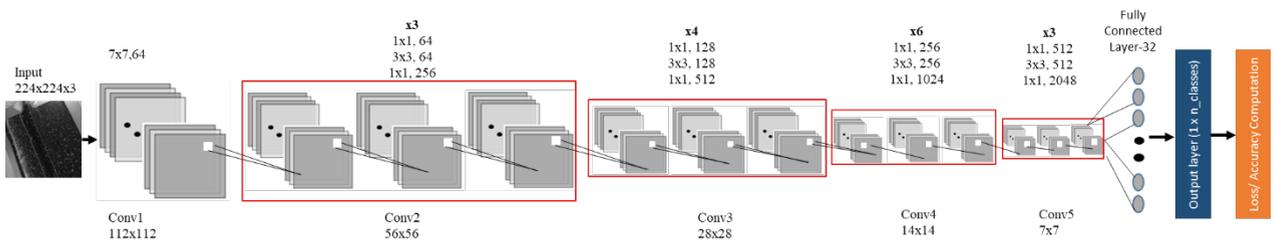


Figure 11: Improved ResNet50 architecture for the defect detection problem

Although the fine-tuned ResNet model is just right model, it requires a lot of computational power because of a huge number of parameters. Thus, this research needs a light-weight model such as MobileNetV2, so that it can also work with less computational resources.

4.3.2. Improved MobilenetV2 Model

Similarly, the updated ResNet architecture, the MobileNetV2 architecture [59] is also modified dropping the last layer and adding two new layers at the end, as shown in Figure 12. Although, MobileNetV2 model has 53 layers, it has significantly less parameters than ResNet model and is a light-weight model that can be used even in mobile devices.

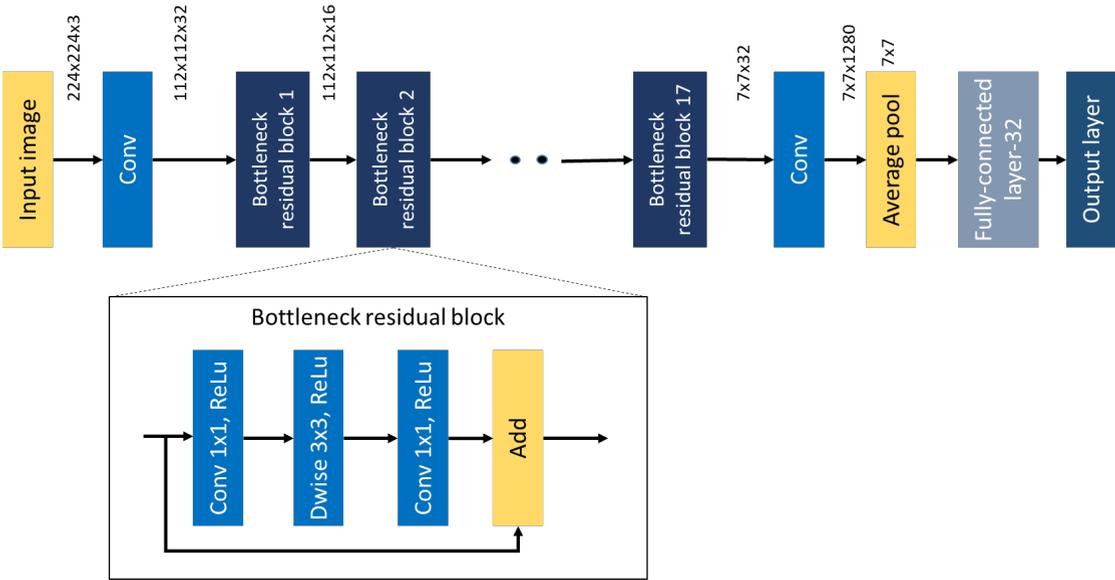


Figure 12: Improved MobileNetV2 architecture for the defect detection problem.

Nevertheless, the DCNN models are effective models and provide decent results, it requires further investigation to improve the results in terms of accuracy, precision, recall, and F1 score. Thus, this work proposes a hybrid model comprising the characteristics of DCNN and classical ML models.

4.4. Hybrid Model

Since the research community relies more on ML models and considers DCNN models as black box that extracts features efficiently, therefore, the present work develops a hybrid model that integrates the qualities of both classical ML model and DCNN models. Therefore, first, Resnet50 with classical ML models are applied on the given ACMID dataset and then, MobileNetV2 models are considered with SVM and RF models.

4.4.1. Improved Resnet50 with Classical ML Models

In Figure 13, it can be noted that first, input images are passing through the updated ResNet architecture similar in Figure 11, to train the model; second, once the fine-tuned ResNet model is trained with the aerospace component images, the last layer is dropped, and it extracts the features using fully connected layer, second last layer with 32 neurons. Then, the extracted features hand it over to classical ML models. In this case, SVM and RF are employed individually to train the model again with the extracted features and classify the defective and non-defective images in the end, as illustrated in Figure 13 and 14. Finally, it determines the performances of the model in terms of accuracy, precision, recall and F1 score to justify the model capability.

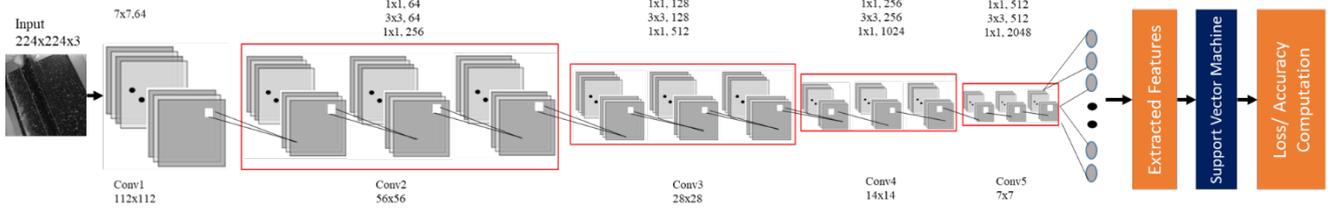


Figure 13: Proposed hybrid approach including improved ResNet50 and SVM for an AOI problem.

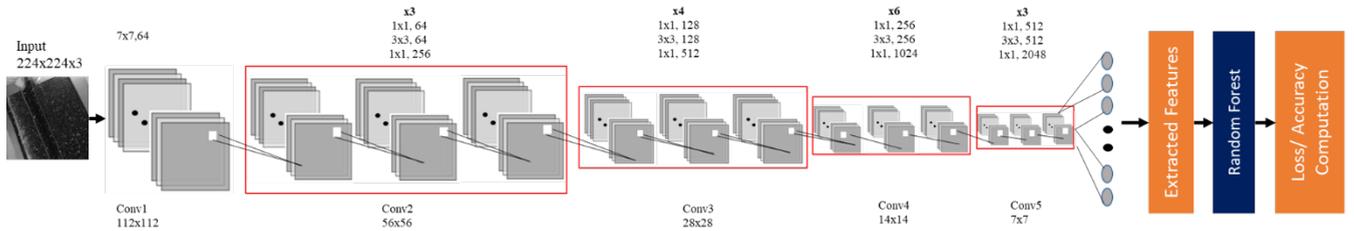


Figure 14: Proposed hybrid approach including improved ResNet50 and RF for an AOI problem.

4.4.2. Improved MobilenetV2 with Classical ML Models

In a similar manner, the proposed architecture of improved ResNet50 with classical ML models for AOI problem in Figure 13 and 14, the enhanced MobileNetV2 with SVM and RF models are proposed to identify the defects automatically. In this case, instead of ResNet architecture, the hybrid model replaces with MobileNetV2 architecture to extract the features and the remaining part is analogous. Proposed model of MobileNetV2 with SVM and RF are displayed in Figure 15 and Figure 16, respectively.

In Table 7, all the models including classical ML models, deep CNN-based models and proposed hybrid models for aerospace composite components are summarized.

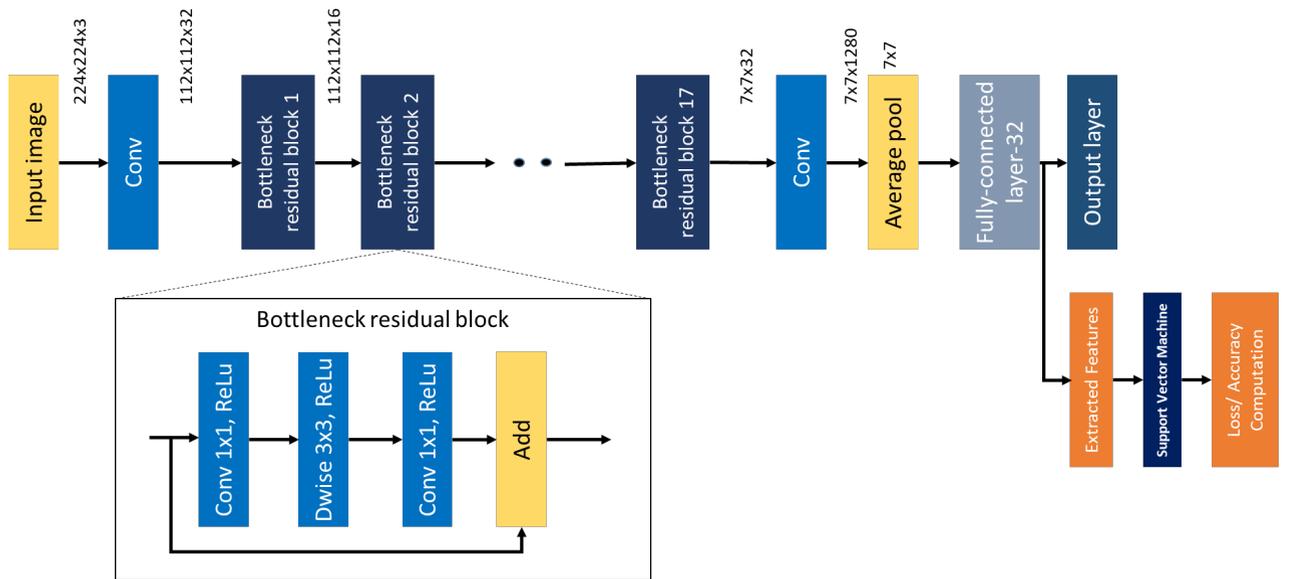


Figure 15: Proposed hybrid approach including improved MobileNetV2 and SVM for an AOI problem.

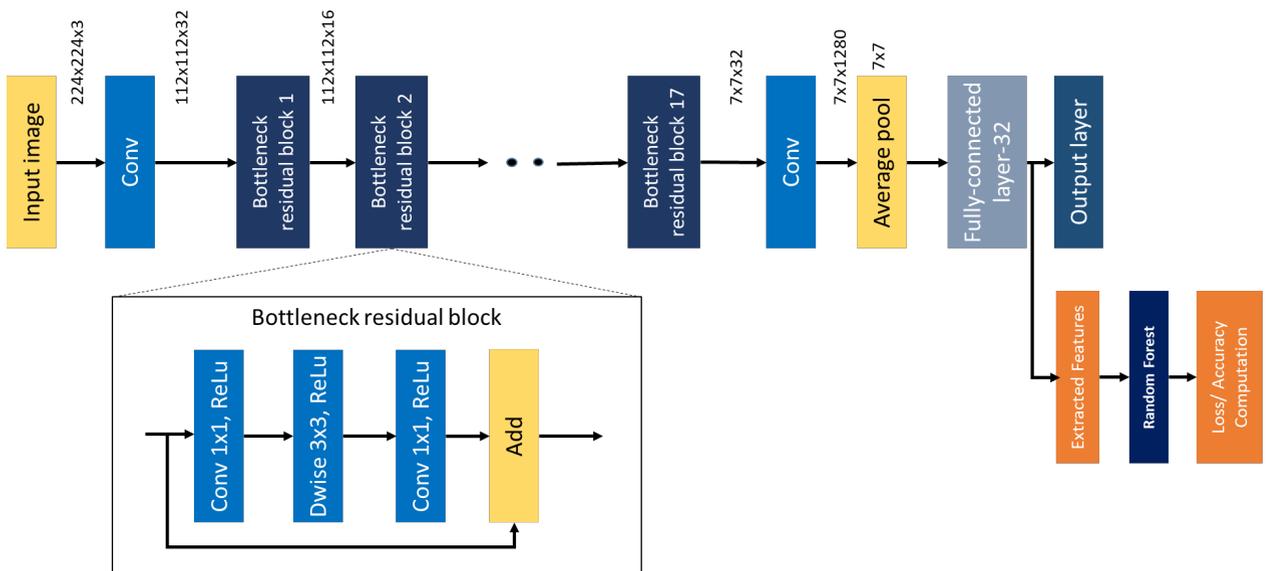


Figure 16: Proposed hybrid approach including improved MobileNetV2 and RF for an AOI problem.

Table 7: All 14 models including classical ML models, deep CNN-based models, and hybrid models.

Model	Trainable Layers				
Classical ML Models	-	Support Vector Machine (SVM)		Random Forest (RF)	
Deep CNN Models	2	ResNet50		MobileNetV2	
	ALL	ResNet50		MobileNetV2	
Hybrid Models	2	ResNet50 + SVM	ResNet50 + RF	MobileNetV2 + SVM	MobileNetV2 + RF
	ALL	ResNet50 + SVM	ResNet50 + RF	MobileNetV2 + SVM	MobileNetV2 + RF

4.5. Imbalanced Dataset Modeling

CNN models have been widely accepted by the computer vision community in several fields. To achieve considerable accuracy of the model, a large-labeled dataset is required to train it for the classification task. However, for the defect classification in automatic optical inspection system, acquiring a large-labeled image dataset is difficult to accomplish. Thus, most of the researchers focused on the data augmentation technique to expand the number of images with the help of shifting, rotating, shearing, and other processes. With this technique, a large training dataset can be achieved from the small number of images. However, in the manufacturing domain, it is hard to collect the same number of defective and defect-free samples in the early phase of production, which leads to an imbalanced dataset. Insufficient abnormal images create an imbalanced dataset and result in inaccurate representation of the distribution of all defective samples, which makes the training process difficult. Furthermore, the low number of defective samples, with low contrast in many cases, causes ambiguity in both defective samples and normal samples. To address this issue, first, imbalanced ACMID dataset is created from the actual ACMID

dataset where 75 percent of the images are normal samples, and the remaining 25 percent of the images are defective samples. To overcome the imbalanced and scarce dataset of defect classification problems, under sampling, over-sampling, and data augmentation techniques can be used to analyze the results. However, in this case, the actual ACMID dataset size is scarce so opting under sampling technique is not suitable. Therefore, this work considers oversampling technique and data augmentation technique using GAN (Generative Adversarial Network) model for imbalanced and scarce samples.

4.5.1. Oversampling

Oversampling is a technique used in machine learning to address the issue of imbalanced datasets. An imbalanced dataset refers to a situation where the classes or categories in the dataset are not represented equally, with one or more classes being significantly underrepresented compared to others. In such cases, a common problem is that machine learning models can be biased towards the majority class, leading to poor performance in predicting the minority class. Oversampling aims to alleviate this issue by increasing the number of instances in the minority class to balance the dataset. There are several methods for oversampling including random sampling, synthetic minority over-sampling technique (SMOTE), adaptive synthetic sampling (ADASYN), etc. In this dissertation work, random sampling is opted.

Random oversampling randomly replicates instances from the minority class until it reaches a desired balance with the majority class. The oversampled dataset will have more instances of the minority class, making it more balanced. This technique aims to balance the dataset by artificially increasing the representation of the minority class. By

doing so, they help machine learning models learn from a more representative training set, improving their ability to predict the minority class accurately. It is important to note that oversampling should be used with caution, as it can introduce some challenges, such as overfitting.

4.5.2. Data Augmentation using Enhanced DCGAN Model

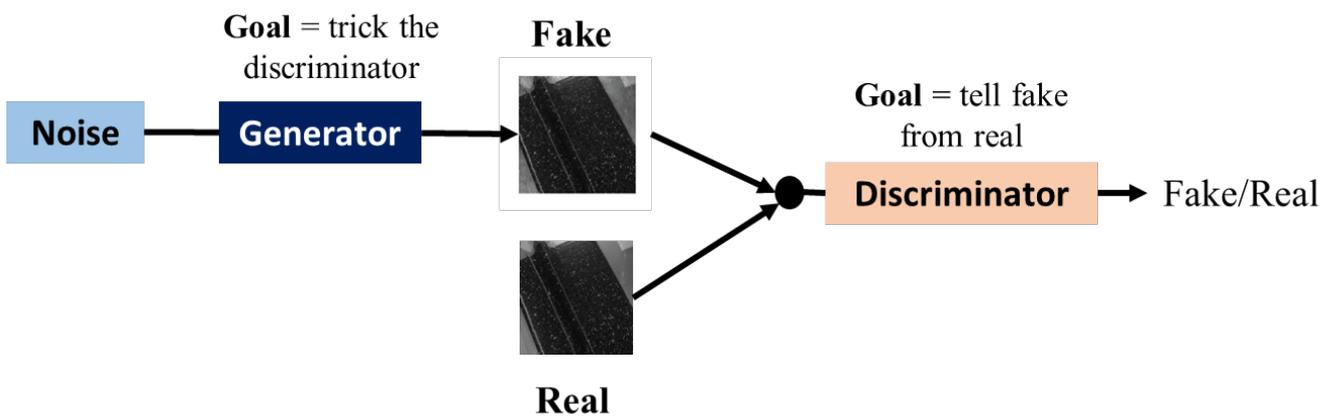


Figure 17: A generative adversarial network for ACMID Dataset.

A Deep Convolutional Generative Adversarial Network (DCGAN) [53] is a class of machine learning models consisting of two neural networks: a generator and a discriminator. Initially, GANs [120] were introduced by Ian Goodfellow and his colleagues in 2014 and have gained significant attention due to their ability to generate realistic synthetic data. A generic generative adversarial network for ACMID dataset is illustrated in Figure 17.

The generator network in an enhanced DCGAN takes Gaussian noise as input and learns to generate synthetic data samples, such as images. It typically starts by producing random and low-quality outputs. Over time, through training, the generator learns to

generate data that becomes increasingly similar to the real data it was trained on. The enhanced DCGAN's generator network begins with Gaussian noise. It then undergoes projection and reshaping operations, specifying the desired dimensions in terms of height, width, and channels. The generator is trained using four convolutional blocks to produce synthetic samples of ACMID, as illustrated in Figure 18.

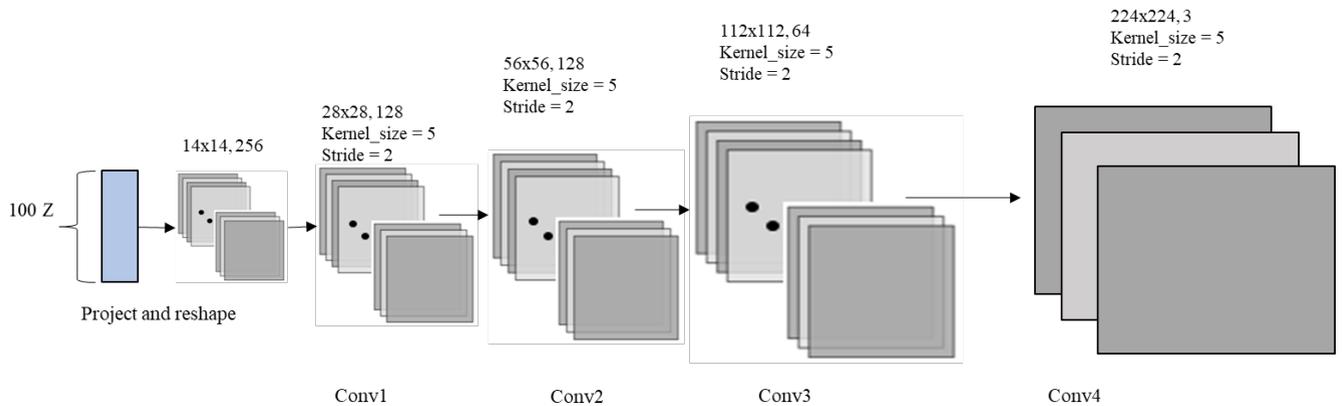


Figure 18: Enhanced DCGAN generator block for ACMID Dataset.

The discriminator network, on the other hand, acts as a binary classifier that learns to distinguish between the real data and the synthetic data produced by the generator. It receives real and synthetic data samples as input and learns to predict whether each sample is real or fake. The enhanced DCGAN utilizes a discriminator network that operates on generated images. It applies convolutional operations on four blocks, flattens the output of the fourth convolutional block, and ends with a binary classifier, as depicted in Figure 19.

During training, the generator and discriminator play a two-player minimax game, competing against each other. The generator's objective is to produce synthetic data that the discriminator classifies as real, while the discriminator aims to correctly classify the

real and synthetic data. As training progresses, the generator improves its ability to generate more realistic data, while the discriminator becomes better at distinguishing between real and synthetic data.

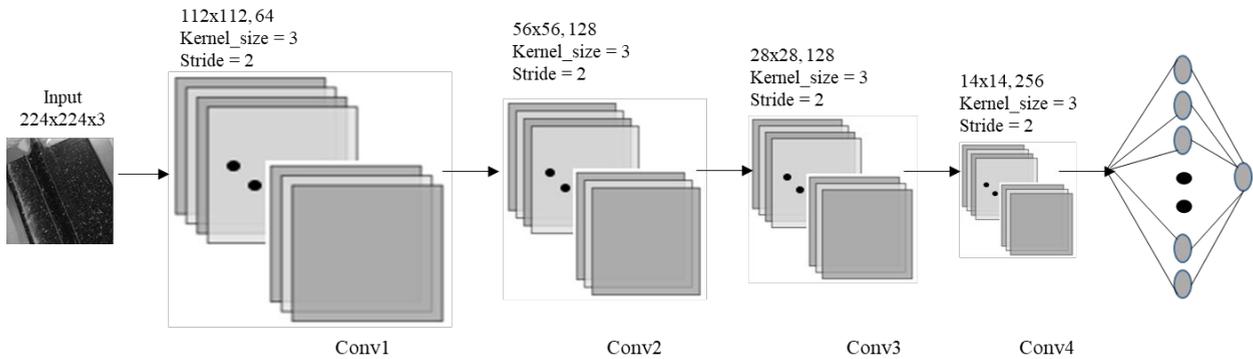


Figure 19: Enhanced DCGAN discriminator block for ACMID Dataset.

The training process involves iteratively updating the weights of both networks based on their performance. The networks are trained using backpropagation and gradient descent techniques, where the generator tries to minimize the discriminator's ability to differentiate real and synthetic data, and the discriminator tries to maximize its accuracy in classifying the samples. The ultimate goal of an improved DCGAN is for the generator to produce synthetic data that is indistinguishable from real data, fooling the discriminator. Enhanced DCGAN generates realistic images, as illustrated in Figure 20.

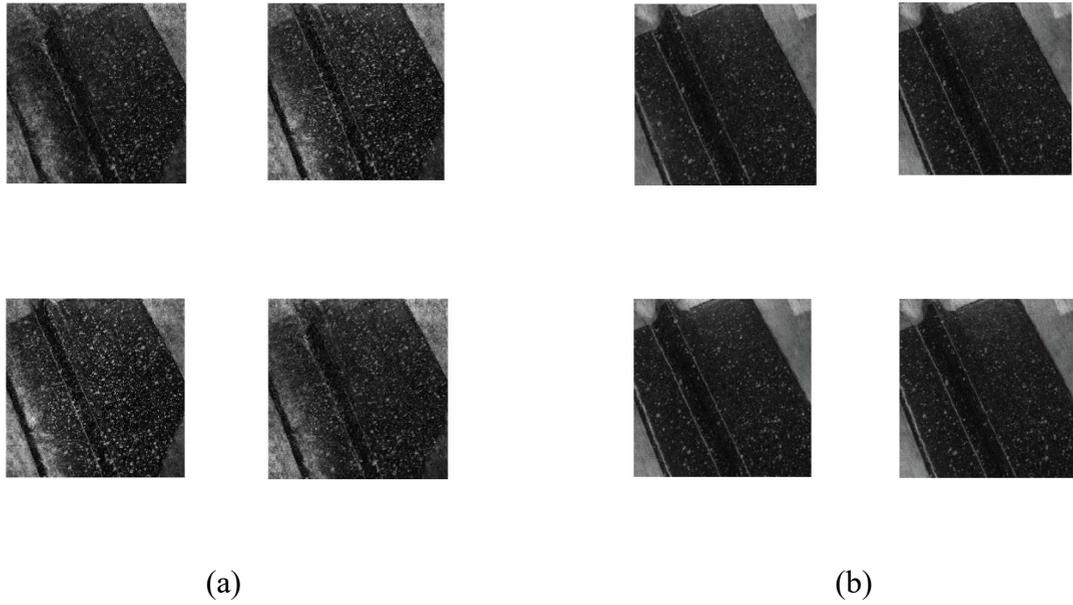


Figure 20: Generated samples of enhanced DCGAN model - (a) with defects and (b) without defects

All the above classical ML models, deep CNN based models and hybrid models will be employed on the generated samples of ACMID dataset using improved DCGAN model to classify the defects.

CHAPTER 5: EXPERIMENTAL SETUP AND RESULTS

In this section, the experimental setup for AOI methods using aerospace component images is described. In addition, experimental results are analyzed, and research implications are discussed.

5.1. Datasets

In this sub section, the datasets used in this dissertation is categorized into five parts such as ACMID dataset, ACMID imbalanced dataset, ACMID oversampled dataset, augmented ACMID dataset using GAN model, and SSD dataset.

5.1.1. ACMID Dataset

The dataset is collected from aerospace composite material lab for automated optical inspection. ACMID dataset comprises of 791 composite component images including 372 defective images and 419 non-defective images. The samples of ACMID dataset are shown in Figure 10. The experiments are extensively conducted on ACMID full dataset and ACMID half dataset using 791 images and 391 images, respectively, as shown in Table 8. However, the resolution of the image is 2736 x 2192 which is quite high, therefore this work converts images into 224 x 224 resolution for all the models. In addition, the defective and non-defective samples are collected from Composite Research Lab, and all the images are normalized before giving as input to the model. The Composites Laboratory is equipped to enable fundamental and applied research and development in emerging composite technologies and is located within Embry-Riddle's Research Park.

Obtaining the ACMID dataset was supported in part by the National Science Foundation (NSF) under Grant 2001038.

5.1.2. ACMID Imbalanced Dataset

In the manufacturing industry, obtaining an equal number of defective and defect-free samples during the initial stages of production is challenging, resulting in imbalanced datasets. This lack of sufficient abnormal images at the beginning creates an imbalanced dataset, which inaccurately represents the distribution of all defective samples and complicates the training process. To tackle this issue, an imbalanced ACMID dataset is generated from the original ACMID dataset, with 75 percent of the images being normal samples and the remaining 25 percent being defective samples. The experiments are carried out extensively on an imbalanced ACMID dataset, which consists of 50 images representing defective samples and 419 images representing non-defective samples. The stats of imbalanced ACMID dataset can be noted in Table 8.

5.1.3. ACMID Oversampling Dataset

To tackle the difficulties presented by the imbalanced dataset, one approach is to utilize oversampling techniques. Specifically, random oversampling is a technique that duplicates instances from the minority class randomly until it achieves a desired equilibrium with the majority class. Consequently, the oversampled dataset contains a greater number of instances from the minority class, effectively creating a more balanced representation. The objective of this technique is to address the dataset's imbalance by artificially augmenting the representation of the minority class. This enables machine learning models to learn from a more representative training set, enhancing their capacity

to accurately predict the minority class. Following the implementation of random oversampling, the size of the defective samples increases to 400, while the non-defective samples remain at 419, as indicated in Table 8.

5.1.4. Augmented ACMID Dataset using Enhanced DCGAN Model

In order to address the challenges posed by the imbalanced and limited dataset, one strategy is to employ data augmentation techniques. An augmented ACMID dataset refers to a modified version of an original ACMID full dataset where additional synthetic or artificially generated images are included. These synthetic images are created by applying enhanced DCGAN model to the original images. A total of four thousand synthetic images are generated in which defective samples are two thousand and same amount is generated for non-defective samples. In addition, oversampling technique is used to increase the size of real images from 791 to 4,000, two thousand for each class. Thus, the total dataset size is eight thousand comprising of four thousand for defective and four thousand for non-defective, as shown in Table 8. Augmentation technique is employed here to increase the size, diversity, and variability of the dataset, which can improve the performance and robustness of machine learning models.

The purpose of using an augmented image dataset is to expose machine learning models to a wider range of variations and scenarios during training. This helps the models to become more robust and generalize better to unseen data. Augmented datasets also address issues such as class imbalance by creating additional samples for underrepresented classes.

Table 8: ACMID dataset

ACMID Dataset	Samples		Dataset
	Defective	Non-defective	Size
Full	372	419	791
Half	191	200	391
Imbalanced	50	419	469
Oversampling	400	419	819
Augmented using enhanced DCGAN model	4,000	4,000	8,000

5.1.5. SSD Dataset

To demonstrate the efficacy of the proposed approach in identifying defects, this work utilizes a Steel Surface Defect (SSD) database originally collected by Song and Yan [121] for detecting defects in hot-rolled steel strips at Northeastern University (NEU). The database comprises six distinct categories of defects, namely crazing, inclusion, patches, pitted-surface, rolled-in scale, and scratches. In total, the database contains 1800 grayscale images, with each defect category consisting of 300 samples, as shown in Table 9. Figure 21 displays sample images of these six typical surface defects, each with an original resolution of 200×200 . This dataset offers a greater variety and more samples compared to the actual ACMID dataset, providing a suitable means to evaluate the robustness of the proposed approaches.

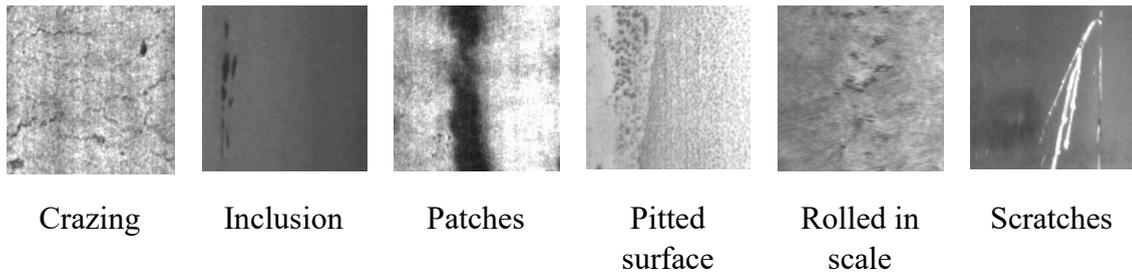


Figure 21: Samples of six kinds of typical steel surface defects on NEU surface defect database. It shows one example image from each of 300 samples of a class

Table 9: SSD dataset

Class	Sample size of each class
Crazing	300
Inclusion	300
Patches	300
Pitted surface	300
Rolled in scale	300
Scratches	300

5.2. Performance Measures

For the defect classification problem, this work evaluates the validation accuracy, precision, recall, and F1 scores for the all the models such as classical ML models, deep CNN based models and proposed hybrid models. For the defect classification task, cross validation technique is used with five folds. In addition, the average validation accuracy, precision, recall, and F1 scores are computed across different folds. For accuracy, the standard deviation is also calculated. These metrics are defined below.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1 score} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative.

5.3. Implementation Details

A total of 14 models including classical ML models, deep CNN-based models, and proposed hybrid models are trained on distinct datasets with different scenarios. First, SVM and RF of classical ML models are trained and parameters of both the models are tuned using grid search technique. Furthermore, enhanced ResNet50 and MobileNetV2 architectures of deep CNN-based models are trained considering last two layers and all layers as trainable layers. Model parameters are initialized with learned parameter values on ImageNet dataset. These models are tested on full, half, imbalanced, oversampling and augmented ACMID dataset, and SSD dataset. Both the ResNet50 and MobileNetV2 architectures are used same as published except the last layer is replaced with two new layers. The second last layer is experimented with a distinct set of neurons and finally it is tuned with a specific number of neurons. Since this is a binary classification problem, the last layer is assigned with one unit. Proposed hybrid models are trained in a similar way to classical ML models and deep CNN models are trained except the architecture is different. In addition, all the models are implemented using Google Colab Pro+ including 51 GB of RAM, NVIDIA P100 or T4 25 GB of GPU, and 250 GB of disk memory.

5.4. Results

To examine the effectiveness of the proposed research, defect classification models are extensively tested using classical ML models, deep CNN-based models, and hybrid models with different datasets and distinct dataset sizes. In the classical ML models, Support Vector Machine and Random Forest model results are compared considering all set of datasets for each model. In addition, a 5-fold cross validation scheme is used for both the models. Then, the generated results for each model are compared with cross validation mean of accuracy, precision, recall, and F1 score. Similarly, the results of deep CNN-based models such as enhanced ResNet50 and MobileNetV2 architecture are analyzed. After assessing classical ML models and deep CNN-based models, this research evaluates the proposed hybrid approach combining enhanced ResNet50 with SVM and RF models, and enhanced MobileNetV2 with SVM and RF models for aerospace components. At the end of the modeling and analysis process, the experiment results of all 14 models considering 70 distinct combinations with ACMID dataset and 14 different combinations with SSD dataset are obtained and compared.

5.4.1. Comparison with Classical ML Models using Distinct Datasets

In this section focusing on classical ML models, the results are obtained for Support Vector Machine (SVM) and Random Forest models using six different datasets: 791 (ACMID full dataset), 391 (ACMID half dataset), 469 (ACMID imbalanced), 819 (ACMID oversampled), 8000 (augmented ACMID), and 1800 (SSD dataset). Both models are evaluated using a 5-fold cross validation scheme. The results, including cross validation accuracy, precision, recall, and F1 score, are compared and presented in Table 10-14. Among the models, Random Forest with the full dataset achieves the highest validation

accuracy, as shown in Table 10. Moreover, the precision and F1 score of the Random Forest model outperforms those of the SVM model, while the recall values are similar for both models. To assess the models' performance with a limited dataset, the experiments are conducted using half of the ACMID dataset size. Once again, the Random Forest model yields better results compared to the SVM model, but it does not surpass the performance achieved with the full dataset, as indicated in Table 10. The parameter values are fine-tuned for each model through multiple experiments with different combinations.

Table 10: Defect classification results of classical ML models using ACMID Dataset, the best results are highlighted in red color.

MODEL	CLASSICAL ML MODEL			
	SVM		RF	
Dataset Size	Full	Half	Full	Half
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	97.85 ± 0.50	96.92 ± 1.03	98.23 ± 0.61	97.69 ± 2.05
PRECISION (MEAN)	0.969	0.963	0.978	0.981
RECALL (MEAN)	0.990	0.981	0.988	0.976
F1 SCORE (MEAN)	0.979	0.971	0.983	0.978

Table 11 presents the outcomes of classical ML models applied to the ACMID imbalanced dataset, consisting of 419 non-defective samples and 50 defective samples. The results reveal that the validation accuracy of the ACMID full dataset surpasses that of the ACMID imbalanced dataset. Furthermore, even though the ACMID half dataset contains

a smaller number of samples, it performs better than the imbalanced dataset because the half dataset is more balanced in terms of class distribution.

Table 11: Defect classification results of classical ML models using Imbalanced ACMID Dataset, the best results are highlighted in red color.

MODEL	CLASSICAL ML MODEL	
	SVM	RF
Dataset Size	469	469
CV	5	5
VAL_ACC (MEAN ± STD)	91.89 ± 0.85	97.23 ± 1.50
PRECISION (MEAN)	0.8977	0.9449
RECALL (MEAN)	0.6376	0.9052
F1 SCORE (MEAN)	0.6903	0.9238

Table 12 displays the outcomes of classical ML models applied to the ACMID oversampled dataset, which consists of 419 non-defective samples and 400 defective samples. It is evident that the validation accuracy of the Random Forest (RF) model using the ACMID oversampled dataset outperforms that of the ACMID full, half, and imbalanced datasets. However, the SVM model yields poor results when utilizing the ACMID oversampled dataset compared to the ACMID full and half datasets. Despite the balanced number of samples for each class, the distribution of defective samples in the oversampled dataset is less diverse than that of non-defective samples, since the defective sample size has been increased through random over-sampling.

Table 12: Defect classification results of classical ML models using Over Sampled ACMID Dataset, the best results are highlighted in red color.

MODEL	CLASSICAL ML MODEL	
	SVM	RF
Dataset Size	819	819
CV	5	5
VAL_ACC (MEAN ± STD)	92.06	99.14
PRECISION (MEAN)	0.9214	0.9914
RECALL (MEAN)	0.9201	0.9916
F1 SCORE (MEAN)	0.9204	0.9914

Table 13 showcases the outcomes of classical ML models applied to the augmented ACMID dataset generated using an enhanced DCGAN model. The dataset consists of 4,000 non-defective samples and 4,000 defective samples. It is evident that the validation accuracy of the Random Forest (RF) model using the augmented ACMID dataset outperforms that of the ACMID full, half, imbalanced, and oversampled datasets. Similarly, the results of the SVM model also surpass those of the ACMID full, half, imbalanced, and oversampled datasets. Furthermore, there is an improvement in precision, recall, and F1 score for both models. This improvement can be attributed to the increased diversity in the distribution of defective and non-defective samples and dataset size compared to previous cases.

Table 13: Defect classification results of classical ML models with Augmented ACMID Dataset using enhanced DCGAN model, the best results are highlighted in red color.

MODEL	CLASSICAL ML MODEL	
	SVM	RF
Dataset Size	8,000	8,000
CV	5	5
VAL_ACC (MEAN ± STD)	98.94	99.37
PRECISION (MEAN)	0.9896	0.9936
RECALL (MEAN)	0.9882	0.9936
F1 SCORE (MEAN)	0.9889	0.9936

Table 14 displays the outcomes of classical ML models applied to the SSD dataset, consisting of 1,800 samples distributed across six classes. One can note that the Random Forest (RF) model outperforms the SVM model in terms of validation accuracy, precision, recall, and F1 score when using the SSD dataset. It is worth noting that the SSD dataset exhibits greater diversity, variation, and an increased number of real samples compared to the ACMID dataset. As a result, there is a significant drop in validation accuracy, precision, recall, and F1 score. However, these results can be improved by employing more powerful models such as deep learning models.

Table 14: Defect classification results of classical ML models using SSD Dataset, the best results are highlighted in red color.

MODEL	CLASSICAL ML MODEL	
	SVM	RF
Dataset Size	1800	1800
CV	5	5
VAL_ACC (MEAN)	80.66	82.16
PRECISION (MEAN)	0.8351	0.8220
RECALL (MEAN)	0.8061	0.8220
F1 SCORE (MEAN)	0.8065	0.8220

5.4.2. Comparison with Deep CNN Based Models using Distinct Datasets

Further, the results of deep CNN-based models such as enhanced ResNet50 and MobileNetV2 architectures are compared using all sets of the dataset. First two models of the deep learning models are employed as a fine-tuned model where few layers are trained layers and other layers are non-trainable. In this case, the last two layers of ResNet50 and MobilenNetV2 models are trainable. Furthermore, these models are also examined with all layers as trainable layers. Cross validation scheme is 5-fold for each model. In the end, a total of four models are tested considering different datasets and distinct dataset sizes. The resulting outcomes are presented in Table 15-19.

The outcomes of deep CNN-based models applied to the ACMID full and half datasets, containing 791 samples and 391 samples respectively, are presented in Table 15. Notably, the enhanced fine-tuned MobileNetV2 model achieves the highest validation accuracy among the models when using the full dataset, and it also exhibits the lowest

standard deviation. Additionally, this model outperforms the others in terms of precision, recall, and F1 score. To evaluate the models' performance with a limited dataset, experiments are conducted using half of the ACMID dataset size. Once again, the fine-tuned MobileNetV2 model produces superior results compared to the ResNet50 model. However, it does not surpass the performance achieved with the full dataset, as indicated in Table 15.

Table 15: Defect classification results of deep CNN based models using ACMID half and full dataset, the best results are highlighted in red color.

MODEL	DL MODEL							
	ResNet50				MobileNetV2			
Trainable Layers	2		ALL		2		ALL	
Dataset Size	Full	Half	Full	Half	Full	Half	Full	Half
CV	5	5	5	5	5	5	5	5
VAL_ACC (MEAN \pm STD)	98.48 \pm 1.02	96.67 \pm 1.03	98.48 \pm 1.24	96.67 \pm 1.72	98.86 \pm 0.92	97.95 \pm 1.52	98.35 \pm 0.94	97.44 \pm 1.40
PRECISION (MEAN)	0.985	0.966	0.9848	0.967	0.988	0.979	0.984	0.974
RECALL (MEAN)	0.984	0.967	0.9847	0.967	0.988	0.979	0.983	0.974
F1 SCORE (MEAN)	0.984	0.966	0.9847	0.967	0.988	0.979	0.983	0.974

Table 16 showcases the outcomes of deep CNN-based models applied to the ACMID imbalanced dataset, which consists of 419 non-defective samples and 50 defective samples. The results demonstrate that the validation accuracy of the ACMID full dataset surpasses that of the ACMID imbalanced dataset. Furthermore, although the ACMID half dataset has a smaller sample size, its results are either superior or comparable to those of

the imbalanced dataset due to the improved balance in class distribution. It is worth noting that the fine-tuned ResNet50 model with two trainable layers performs better than the fine-tuned MobileNetV2 model in this case. However, the results obtained by MobileNetV2 outperform ResNet50 when all layers are trainable.

Table 16: Defect classification results of deep CNN based models using Imbalanced ACMID Dataset, the best results are highlighted in red color.

MODEL	DL MODEL			
	ResNet50		MobileNetV2	
Trainable Layers	2	ALL	2	ALL
Dataset Size	469	469	469	469
CV	5	5	5	5
VAL_ACC (MEAN \pm STD)	97.65 \pm 1.70	97.23 \pm 2.57	97.02 \pm 2.64	97.44 \pm 2.19
PRECISION (MEAN)	0.9591	0.9425	0.9474	0.9433
RECALL (MEAN)	0.9227	0.9194	0.8961	0.9270
F1 SCORE (MEAN)	0.9391	0.9296	0.91734	0.9346

Table 17 presents the outcomes of deep CNN-based models applied to the ACMID oversampled dataset, which includes 419 non-defective samples and 400 defective samples. It is clear that the improved MobileNetV2 model, with all layers trainable, achieves higher validation accuracy when using the ACMID oversampled dataset compared to the ACMID full, half, and imbalanced datasets. However, the fine-tuned ResNet50 model performs poorly when applied to the ACMID oversampled dataset, in contrast to its performance with the ACMID full and half datasets. Despite having a

balanced number of samples for each class, the oversampled dataset exhibits a less diverse distribution of defective samples in comparison to non-defective samples.

Table 17: Defect classification results of deep CNN based models using Over Sampled ACMID Dataset, the best results are highlighted in red color.

MODEL	DL MODEL			
	ResNet50		MobileNetV2	
Trainable Layers	2	ALL	2	ALL
Dataset Size	819	819	819	819
CV	5	5	5	5
VAL_ACC (MEAN \pm STD)	77.53 \pm 3.47	99.02 \pm 0.62	94.02 \pm 1.40	99.14 \pm 0.48
PRECISION (MEAN)	0.7753	0.9902	0.9428	0.9912
RECALL (MEAN)	0.7707	0.9904	0.9389	0.9917
F1 SCORE (MEAN)	0.7700	0.9902	0.9397	0.9914

Table 18 presents the results of deep CNN-based models applied to the augmented ACMID dataset, which was created using an enhanced DCGAN model. The dataset consists of 4,000 non-defective samples and 4,000 defective samples. It is clear that the enhanced MobileNetV2 model, with all layers trainable, achieves a higher validation accuracy when using the augmented ACMID dataset compared to the ACMID full, half, imbalanced, and oversampled datasets. Similarly, the enhanced ResNet50 model, with all layers trainable, also outperforms the ACMID full, half, imbalanced, and oversampled datasets. Moreover, there is an improvement in precision, recall, and F1 score for both models. This improvement can be attributed to the increased diversity in the distribution of

defective and non-defective samples, as well as the larger dataset size, compared to the previous cases.

Table 18: Defect classification results of deep CNN based models with Augmented ACMID Dataset using GAN, the best results are highlighted in red color.

MODEL	DL MODEL			
	ResNet50		MobileNetV2	
Trainable Layers	2	ALL	2	ALL
Dataset Size	8k	8k	8k	8k
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	95.27 ± 4.51	99.48 ± 0.33	98.52 ± 1.12	99.70 ± 0.26
PRECISION (MEAN)	0.9545	0.9950	0.9820	0.9970
RECALL (MEAN)	0.9539	0.9946	0.9897	0.9970
F1 SCORE (MEAN)	0.9526	0.9948	0.9858	0.9970

Table 19 showcases the results of deep CNN-based models applied to the SSD dataset, which consists of 1,800 samples distributed across six classes. Notably, the MobileNetV2 model with all layers trainable outperforms the ResNet50 model in terms of validation accuracy, precision, recall, and F1 score when using the SSD dataset. It is important to highlight that the SSD dataset offers a greater level of diversity, variation, and a larger number of real samples compared to the ACMID dataset. While there was a significant decrease in validation accuracy, precision, recall, and F1 score observed when using classical ML models in Table 14, these results are significantly improved by leveraging more powerful models such as deep CNN-based models.

Table 19: Defect classification results of deep CNN based models using SSD Dataset, the best results are highlighted in red color.

MODEL	DL MODEL			
	ResNet50		MobileNetV2	
Trainable Layers	2	ALL	2	ALL
Dataset Size	1800	1800	1800	1800
CV	5	5	5	5
VAL_ACC (MEAN)	97.27	99.77	97.27	99.88
PRECISION (MEAN)	0.9927	0.9977	0.9927	0.9990
RECALL (MEAN)	0.9927	0.9977	0.9927	0.9986
F1 SCORE (MEAN)	0.9927	0.9977	0.9927	0.9988

5.4.3. Comparison with Hybrid Models using Distinct Datasets

In this section, the results are obtained and compared using proposed hybrid models on different datasets and dataset sizes.

5.4.3.1. Improved ResNet50 Model with Classical ML Models using Distinct Datasets

Following the assessment of classical ML models and deep CNN-based models, this study proceeds to evaluate a hybrid approach that combines the improved ResNet50 architecture with SVM and RF models for aerospace components. Different combinations are examined for proposed hybrid model of ResNet50 and SVM/RF, considering different trainable layers, datasets, and dataset sizes.

Table 20 presents the results of the proposed hybrid models, ResNet50 + SVM/RF, applied to the ACMID full and half datasets, consisting of 791 samples and 391 samples

respectively. It is worth noting that the validation accuracy of ResNet50 combined with RF is generally superior to that of ResNet50 combined with SVM, except when all layers are trainable with the full dataset. Furthermore, the best results obtained with ResNet50 + SVM are comparable to those achieved with ResNet50 + RF.

To assess the models' performance with a limited dataset, experiments are conducted using half of the ACMID dataset size. Once again, the ResNet50 + RF model outperforms the ResNet50 + SVM model. However, it does not surpass the performance achieved with the full dataset.

Table 20: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF); the best results are highlighted in red color.

MODEL	HYBRID MODEL							
	ResNet50 + SVM				ResNet50 + RF			
Trainable Layers	2		ALL		2		ALL	
Dataset Size	Full	Half	Full	Half	Full	Half	Full	Half
CV	5	5	5	5	5	5	5	5
VAL_ACC (MEAN ± STD)	97.97 ± 0.73	97.18 ± 0.51	98.86 ± 0.61	97.44 ± 1.40	98.72 ± 0.80	97.69 ± 0.95	98.10 ± 0.69	97.95 ± 0.62
PRECISION (MEAN)	0.963	0.949	0.988	0.954	0.986	0.971	0.978	0.976
RECALL (MEAN)	1.0	1.0	0.990	1.0	0.990	0.985	0.985	0.985
F1 SCORE (MEAN)	0.981	0.974	0.989	0.976	0.988	0.978	0.982	0.980

The outcomes of the proposed hybrid models, ResNet50 + SVM/RF, applied to the ACMID imbalanced dataset, containing 419 non-defective samples and 50 defective

samples, are presented in Table 21. The results indicate that the validation accuracy of the ACMID full and half datasets exceeds that of the ACMID imbalanced dataset. However, it is noteworthy that the enhanced fine-tuned ResNet50 model with RF outperforms the other three models, as shown in Table 21.

Table 21: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) using Imbalanced ACMID Dataset, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	ResNet50 + SVM		ResNet50 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	469	469	469	469
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	89.33 ± 0.004	89.33 ± 0.004	97.44 ± 0.015	97.23 ± 0.020
PRECISION (MEAN)	0.893	0.893	0.981	0.983
RECALL (MEAN)	1.0	1.0	0.990	0.985
F1 SCORE (MEAN)	0.943	0.943	0.985	0.984

The outcomes of the proposed hybrid models, ResNet50 + SVM/RF, applied to the ACMID oversampled dataset, consisting of 419 non-defective samples and 400 defective samples, are presented in Table 22. It is clear that the validation accuracy of the improved ResNet50 + SVM and RF models, with all layers trainable, using the ACMID oversampled dataset outperforms that of the ACMID full, half, and imbalanced datasets. However, the fine-tuned ResNet50 model with SVM and RF produces poor results when applied to the ACMID oversampled dataset in comparison to the ACMID full and half datasets. Despite

having a balanced number of samples for each class, the distribution of defective samples in the oversampled dataset is less diverse than that of non-defective samples. Furthermore, if the deep learning model fails to extract effective features with a less powerful model in the hybrid model, it becomes challenging for classical ML models to achieve satisfactory results after training on the extracted features.

Table 22: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) using Over Sampled ACMID Dataset, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	ResNet50 + SVM		ResNet50 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	819	819	819	819
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	83.76 ± 0.026	99.26 ± 0.002	91.82 ± 0.025	99.26 ± 0.004
PRECISION (MEAN)	0.8910	1.0	0.9263	0.9952
RECALL (MEAN)	0.7782	0.9856	0.9142	0.9904
F1 SCORE (MEAN)	0.8300	0.9927	0.9191	0.9928

The outcomes of the proposed hybrid models, ResNet50 + SVM/RF, applied to the augmented ACMID dataset generated by an enhanced DCGAN model, are presented in Table 23. The dataset consists of 4,000 non-defective samples and 4,000 defective samples. It is evident that the validation accuracy of the proposed ResNet50 + RF model, with all layers trainable, using the augmented ACMID dataset outperforms the other three proposed models in Table 23. Furthermore, the best result obtained using the augmented ACMID

dataset surpasses that of the ACMID full and half datasets and is comparable to the results achieved with the ACMID oversampled dataset.

Table 23: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) with Augmented ACMID Dataset using GAN, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	ResNet50 + SVM		ResNet50 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	8,000	8,000	8,000	8,000
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	94.33 ± 4.95	98.98 ± 0.84	98.52 ± 1.12	99.12 ± 0.69
PRECISION (MEAN)	0.9790	0.9876	0.9820	0.9883
RECALL (MEAN)	0.9081	0.9926	0.9897	0.9946
F1 SCORE (MEAN)	0.9409	0.9901	0.9858	0.9914

The outcomes of the proposed hybrid models, ResNet50 + SVM/RF, applied to the SSD dataset, consisting of 1,800 samples distributed across six classes, are presented in Table 24. It can be observed that the proposed fine-tuned ResNet50 and SVM/RF model outperforms the ResNet50 and SVM/RF models with all layers in terms of validation accuracy, precision, recall, and F1 score when using the SSD dataset. The results achieved by the fine-tuned ResNet50 and RF model are the best so far in terms of validation accuracy and other performance metrics. It is also noteworthy that the SSD dataset exhibits greater diversity, variation, and a larger number of real samples compared to the ACMID dataset. Although there was a significant decline in validation accuracy, precision, recall, and F1

score when using classical ML models, as shown in Table 14, these results are significantly improved by employing more powerful models such as deep learning models. Furthermore, the ResNet50 + RF model with all layers also produces satisfactory results.

Table 24: Defect classification results of deep CNN based hybrid models (ResNet50 + SVM/RF) using SSD Dataset, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	ResNet50 + SVM		ResNet50 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	1800	1800	1800	1800
CV	5	5	5	5
VAL_ACC (MEAN)	99.77	98.50	99.83	99.66
PRECISION (MEAN)	0.9977	98.57	0.9983	99.66
RECALL (MEAN)	0.9977	98.50	0.9983	99.66
F1 SCORE (MEAN)	0.9977	98.49	0.9983	99.66

5.4.3.2. Improved MobileNetV2 Model with Classical ML Models using Distinct Datasets

After evaluating classical ML models, deep CNN-based models, and ResNet50 + SVM/RF, this study proceeds to assess a proposed lightweight hybrid approach that combines the improved MobileNetV2 architecture with SVM and RF models for aerospace components. Different combinations are examined for proposed hybrid model of MobileNetV2 and SVM/RF, considering different trainable layers, datasets, and dataset sizes.

The outcomes of the proposed hybrid models, MobileNetV2 + SVM/RF, applied to the ACMID full and half datasets, consisting of 791 samples and 391 samples respectively, are presented in Table 25. Significantly, the validation accuracy of the fine-tuned MobileNetV2 with RF model surpasses the other three hybrid approaches, as shown in Table 25, when applied to the ACMID full and half datasets. To evaluate the models' performance with a limited dataset, the results obtained by MobileNetV2 + SVM/RF on the ACMID half dataset are comparable, but they do not surpass the performance achieved with the full dataset.

Table 25: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF), the best results are highlighted in red color.

MODEL	HYBRID MODEL							
	MobileNetV2 + SVM				MobileNetV2 + RF			
Trainable Layers	2		ALL		2		ALL	
Dataset Size	Full	Half	Full	Half	Full	Half	Full	Half
CV	5	5	5	5	5	5	5	5
VAL_ACC (MEAN ± STD)	98.86 ± 0.46	98.46 ± 1.25	98.86 ± 0.47	97.18 ± 0.96	99.37 ± 0.01	98.20 ± 0.62	98.60 ± 0.47	97.70 ± 0.94
PRECISION (MEAN)	0.981	0.981	0.990	0.949	0.992	0.985	0.990	0.971
RECALL (MEAN)	0.997	0.990	0.988	1.0	0.995	0.980	0.983	0.985
F1 SCORE (MEAN)	0.989	0.985	0.989	0.973	0.994	0.983	0.986	0.978

Table 26 presents the results of the proposed hybrid models, MobileNetV2 + SVM/RF, applied to the ACMID imbalanced dataset, comprising 419 non-defective samples and 50 defective samples. The findings indicate that the validation accuracy of the

ACMID full and half datasets exceeds that of the ACMID imbalanced dataset. However, the enhanced MobileNetV2 with RF model, utilizing all layers, outperforms the other three models, as shown in Table 26.

Table 26: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) using Imbalanced ACMID Dataset, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	MobileNetV2 + SVM		MobileNetV2 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	469	469	469	469
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	89.33 ± 0.004	89.55 ± 0.004	97.44 ± 0.018	97.87 ± 0.006
PRECISION (MEAN)	0.893	0.895	0.985	0.983
RECALL (MEAN)	1.0	1.0	0.985	0.992
F1 SCORE (MEAN)	0.943	0.944	0.985	0.988

The results of the proposed hybrid models, MobileNetV2 + SVM/RF, applied to the ACMID oversampled dataset, containing 419 non-defective samples and 400 defective samples, are presented in Table 27. It is clear that the validation accuracy of the enhanced MobileNetV2 + SVM and RF models, with all layers, outperforms the validation accuracy of the ACMID full, half, and imbalanced datasets.

Table 27: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) using Over Sampled ACMID Dataset, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	MobileNetV2 + SVM		MobileNetV2 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	819	819	819	819
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	94.99 ± 0.010	99.63 ± 0.004	99.26 ± 0.004	99.63 ± 0.004
PRECISION (MEAN)	0.9797	1.0	1.0	1.0
RECALL (MEAN)	0.9212	0.9928	0.9856	0.9928
F1 SCORE (MEAN)	0.9495	0.9963	0.9927	0.9963

The results of the proposed hybrid models, MobileNetV2 + SVM/RF, applied to the augmented ACMID dataset generated using an enhanced DCGAN model, are presented in Table 28. The dataset consists of 4,000 non-defective samples and 4,000 defective samples. It is clear that the validation accuracy of the fine-tuned MobileNetV2 + RF model, using the augmented ACMID dataset, outperforms the other three proposed models in Table 28. Additionally, the best result achieved using the augmented ACMID dataset surpasses the results obtained from the ACMID full, half, imbalanced, and oversampled datasets.

Table 28: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) with Augmented ACMID Dataset using GAN, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	MobileNetV2 + SVM		MobileNetV2 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	8,000	8,000	8,000	8,000
CV	5	5	5	5
VAL_ACC (MEAN ± STD)	99.47 ± 0.45	99.29 ± 0.72	99.68 ± 0.47	99.47 ± 0.51
PRECISION (MEAN)	0.9903	0.9938	0.9963	0.9939
RECALL (MEAN)	0.9995	0.9924	0.9975	0.9958
F1 SCORE (MEAN)	0.9949	0.9931	0.9969	0.9948

The outcomes of the proposed hybrid models, MobileNetV2 + SVM/RF, applied to the SSD dataset consisting of 1,800 samples across six classes, are presented in Table 29. It is noteworthy that the hybrid models with MobileNetV2 and SVM/RF, utilizing all layers, outperform the fine-tuned MobileNetV2 and SVM/RF models in terms of validation accuracy, precision, recall, and F1 score when applied to the SSD dataset. The results obtained from the proposed MobileNetV2 and RF model with all layers are the best in terms of validation accuracy and other performance metrics. Additionally, it is important to mention that the SSD dataset exhibits greater diversity, variation, and a larger number of real samples compared to the ACMID dataset. While classical ML models in Table 14 showed a significant decline in validation accuracy, precision, recall, and F1 score, these

results are substantially improved by employing more powerful models such as deep learning models.

Table 29: Defect classification results of deep CNN based hybrid models (MobileNetV2 + SVM/RF) using SSD Dataset, the best results are highlighted in red color.

MODEL	HYBRID MODEL			
	MobileNetV2 + SVM		MobileNetV2 + RF	
Trainable Layers	2	ALL	2	ALL
Dataset Size	1800	1800	1800	1800
CV	5	5	5	5
VAL_ACC (MEAN)	95.33	99.94	97.61	99.94
PRECISION (MEAN)	0.9568	99.94	0.9763	99.94
RECALL (MEAN)	0.9533	99.94	0.9761	99.94
F1 SCORE (MEAN)	0.9534	99.94	0.9760	99.94

5.5. Research Implications

This study focuses on the inspection of aerospace components to identify defects. To identify and evaluate the possible defects, in practice, the aerospace components are examined manually by human experts. The inadequacy of this operation results from the fact that is tedious, subjective, labor-extensive, inconsistent, and potentially biased. To make the operation more effective and efficient, the modern AOI system is preferred to evaluate these aerospace composite components. Ideally, the modern AOI system is expected to deliver more consistent, accurate, and unbiased assessment results than manual inspections. In addition, to lessen the workload of human inspectors and lower the labor cost of the aerospace manufacturing industry, this dissertation work presents a deep

learning-based model for automated optical inspection for aerospace composite components.

CHAPTER 6: CONCLUSION AND FUTURE WORK

To lessen the workload of human inspectors in the aerospace manufacturing industry, this dissertation work presents a deep learning-based model for automated optical inspection for aerospace composite components. Initially, the aerospace composite material image dataset is acquired and classified into five sets of ACMID dataset including full, half, imbalanced, oversampled, data augmentation using DCGAN model. In addition to that SSD dataset is included which has more diversity and variation. Furthermore, the machine vision methods are tested on these two datasets with distinct dataset sizes on three levels including classical ML models, deep CNN-based models, and hybrid approaches.

The hybrid method that combines the feature of deep learning and classical machine learning is proposed for aerospace composite components. The detection accuracy of hybrid method for automated optical inspection improves significantly compared to classical ML models and enhanced deep CNN-based models on ACMID dataset. Although the ACMID dataset contains a small number of images, it still produces encouraging results. This indicates that the proposed approach for AOI is suitable to identify the defects on ACMID dataset as well as on other images.

To improve the diversity and variation of ACMID dataset, the number of samples of ACMID dataset are expanded using improved DCGAN model. In addition to that SSD dataset is considered to validate the model and compare with the results of ACMID dataset. Furthermore, supervised pixel segmentation techniques will be explored on the novel

ACMID dataset. And unsupervised pixel level defect detection can also be considered since manual labeling and annotation of the dataset is challenging and expensive.

REFERENCES

- [1] X. Dong, C. J. Taylor, and T. F. Cootes, “Automatic aerospace weld inspection using unsupervised local deep feature learning,” *Knowledge-Based Syst.*, vol. 221, 2021.
- [2] M. W. Ashour, F. Khalid, A. A. Halin, L. N. Abdullah, and S. H. Darwish, “Surface defects classification of hot-rolled steel strips using multi-directional shearlet features,” *Arab. J. Sci. Eng.*, vol. 44, no. 4, pp. 2925–2932, 2019.
- [3] Y. Huang and K. L. Chan, “Texture decomposition by harmonics extraction from higher order statistics,” *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 1–14, 2004.
- [4] M. Li, S. Wan, Z. Deng, and Y. Wang, “Fabric defect detection based on saliency histogram features,” *Comput. Intell.*, vol. 35, no. 3, pp. 517–534, 2019.
- [5] W. Wen and A. Xia, “Verifying edges for visual inspection purposes,” *Pattern Recognit. Lett.*, vol. 20, no. 3, pp. 315–328, 1999.
- [6] Z. Wen, J. Cao, X. Liu, and S. Ying, “Fabric defects detection using adaptive wavelets,” *Int. J. Cloth. Sci. Technol.*, 2014.
- [7] A. S. Malek, J.-Y. Drean, L. Bigue, and J.-F. Osselin, “Optimization of automated online fabric inspection by fast Fourier transform (FFT) and cross-correlation,” *Text. Res. J.*, vol. 83, no. 3, pp. 256–268, 2013.
- [8] J. Ma, Y. Wang, C. Shi, and C. Lu, “Fast surface defect detection using improved Gabor filters,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 1508–1512.

- [9] R. Kulkarni, E. Banoth, and P. Pal, "Automated surface feature detection using fringe projection: An autoregressive modeling-based approach," *Opt. Lasers Eng.*, vol. 121, pp. 506–511, 2019.
- [10] L. Xu and Q. Huang, "Modeling the interactions among neighboring nanostructures for local feature characterization and defect detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 4, pp. 745–754, 2012.
- [11] Z. Zhang, X. Wang, S. Liu, L. Sun, L. Chen, and Y. Guo, "An automatic recognition method for PCB visual defects," in *2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, 2018, pp. 138–142.
- [12] V. H. Nguyen, V. H. Pham, X. Cui, M. Ma, and H. Kim, "Design and evaluation of features and classifiers for OLED panel defect recognition in machine vision," *J. Inf. Telecommun.*, vol. 1, no. 4, pp. 334–350, 2017.
- [13] L.-F. Chen, C.-T. Su, and M.-H. Chen, "A neural-network approach for defect recognition in TFT-LCD photolithography process," *IEEE Trans. Electron. Packag. Manuf.*, vol. 32, no. 1, pp. 1–8, 2009.
- [14] H. . Çelik, L. C. Dülger, and M. Topalbekiroğlu, "Development of a machine vision system: real-time fabric defect detection and classification with neural networks," *J. Text. Inst.*, vol. 105, no. 6, pp. 575–585, 2014.
- [15] Y. Wang, Y. Sun, P. Lv, and H. Wang, "Detection of line weld defects based on multiple thresholds and support vector machine," *Ndt E Int.*, vol. 41, no. 7, pp. 517–524, 2008.

- [16] J. Mirapeix, P. B. Garcia-Allend, A. Cobo, O. M. Conde, and J. M. López-Higuera, “Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks,” *NDT E Int.*, vol. 40, no. 4, pp. 315–323, 2007.
- [17] G. Fu *et al.*, “A deep-learning-based approach for fast and robust steel surface defects classification,” *Opt. Lasers Eng.*, vol. 121, no. February, pp. 397–405, 2019.
- [18] X. Zheng, H. Wang, J. Chen, Y. Kong, and S. Zheng, “A Generic Semi-Supervised Deep Learning-Based Approach for Automated Surface Inspection,” *IEEE Access*, vol. 8, pp. 114088–114099, 2020.
- [19] X. Dong, C. J. Taylor, and T. F. Cootes, “A Random Forest-Based Automatic Inspection System for Aerospace Welds in X-Ray Images,” *IEEE Trans. Autom. Sci. Eng.*, pp. 1–14, 2020.
- [20] X. Dong, C. J. Taylor, and T. F. Cootes, “Defect Detection and Classification by Training a Generic Convolutional Neural Network Encoder,” *IEEE Trans. Signal Process.*, vol. 68, no. 2, pp. 6055–6069, 2020.
- [21] Y. Gong, H. Shao, J. Luo, and Z. Li, “A deep transfer learning model for inclusion defect detection of aeronautics composite materials,” *Compos. Struct.*, vol. 252, no. June, 2020.
- [22] R. Guo, H. Liu, G. Xie, and Y. Zhang, “Weld Defect Detection from Imbalanced Radiographic Images Based on Contrast Enhancement Conditional Generative Adversarial Network and Transfer Learning,” *IEEE Sens. J.*, vol. 21, no. 9, pp. 10844–10853, 2021.

- [23] Y. Yang *et al.*, “A lightweight deep learning algorithm for inspection of laser welding defects on safety vent of power battery,” *Comput. Ind.*, vol. 123, 2020.
- [24] J. Park, H. Riaz, H. Kim, and J. Kim, “Advanced cover glass defect detection and classification based on multi-DNN model,” *Manuf. Lett.*, vol. 23, pp. 53–61, 2020.
- [25] C. Li, Y. Huang, H. Li, and X. Zhang, “A weak supervision machine vision detection method based on artificial defect simulation,” *Knowledge-Based Syst.*, vol. 208, 2020.
- [26] C. Li, X. Zhang, Y. Huang, C. Tang, and S. Fatikow, “A novel algorithm for defect extraction and classification of mobile phone screen based on machine vision,” *Comput. Ind. Eng.*, vol. 146, no. November 2019, 2020.
- [27] D. M. Tsai and P. H. Jen, “Autoencoder-based anomaly detection for surface defect inspection,” *Adv. Eng. Informatics*, vol. 48, no. October 2020, 2021.
- [28] S. Yang, X. Li, X. Jia, Y. Wang, H. Zhao, and J. Lee, “Deep learning-based intelligent defect detection of cutting wheels with industrial images in manufacturing,” *Procedia Manuf.*, vol. 48, no. 2019, pp. 902–907, 2020.
- [29] T. Nakazawa and D. V. Kulkarni, “Anomaly detection and segmentation for wafer defect patterns using deep Convolutional Encoder-Decoder Neural Network Architectures in Semiconductor Manufacturing,” *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 250–256, 2019.
- [30] S. Cheon, H. Lee, C. O. Kim, and S. H. Lee, “Convolutional Neural Network for Wafer Surface Defect Classification and the Detection of Unknown Defect Class,” *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 2, pp. 163–170, 2019.

- [31] J. K. Chow, Z. Su, J. Wu, P. S. Tan, X. Mao, and Y. H. Wang, "Anomaly detection of defects on concrete structures with the convolutional autoencoder," *Adv. Eng. Informatics*, vol. 45, no. April, 2020.
- [32] M. Zheng, Z. Lei, and K. Zhang, "Intelligent detection of building cracks based on deep learning," *Image Vis. Comput.*, vol. 103, 2020.
- [33] T. Ozseven, "Surface defect detection and quantification with image processing methods," in *Theoretical Investigations and Applied Studies in Engineering*, Ekin Publishing House, 2019, pp. 63–98.
- [34] R. Ren, T. Hung, and K. C. Tan, "A Generic Deep-Learning-Based Approach for Automated Surface Inspection," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 929–940, 2018.
- [35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [36] Y. Shu, B. Li, and H. Lin, "Quality safety monitoring of LED chips using deep learning-based vision inspection methods," *Meas. J. Int. Meas. Confed.*, vol. 168, 2021.
- [37] Y. F. Shu, B. Li, X. Li, C. Xiong, S. Cao, and X. Y. Wen, "Deep learning-based fast recognition of commutator surface defects," *Measurement*, vol. 178, no. February, p. 109324, 2021.
- [38] J. Liu, K. Song, M. Feng, Y. Yan, Z. Tu, and L. Zhu, "Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection," *Opt. Lasers Eng.*, vol. 136, no. March 2020, 2021.

- [39] L. Xu, S. Lv, Y. Deng, and X. Li, "A Weakly Supervised Surface Defect Detection Based on Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 42285–42296, 2020.
- [40] J. Yu, X. Zheng, and S. Wang, "A deep autoencoder feature learning method for process pattern recognition," *J. Process Control*, vol. 79, pp. 1–15, 2019.
- [41] J. Yu, "Enhanced stacked denoising autoencoder-based feature learning for recognition of wafer map defects," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 4, pp. 613–624, 2019.
- [42] A. Mujeeb, W. Dai, M. Erdt, and A. Sourin, "One class based feature learning approach for defect detection using deep autoencoders," *Adv. Eng. Informatics*, vol. 42, no. February, 2019.
- [43] M. Haselmann, D. P. Gruber, and P. Tabatabai, "Anomaly Detection Using Deep Learning Based Image Completion," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 1237–1242, 2019.
- [44] K. Nejc and B. Drago, "Visual Inspection System for Anomaly Detection on KTL Coatings Using VAE," *Procedia CIRP*, vol. 81, pp. 771–774, 2019.
- [45] J. Kim *et al.*, "Adversarial Defect Detection in Semiconductor Manufacturing Process," *IEEE Trans. Semicond. Manuf.*, vol. 34, no. 3, pp. 365–371, 2021.
- [46] D. M. Tsai, S. K. S. Fan, and Y. H. Chou, "Auto-Annotated Deep Segmentation for Surface Defect Detection," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021.
- [47] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

- [48] K. Alex, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [49] C. Szegedy *et al.*, “Going deeper with convolutions,” *2015 IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015.
- [50] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.” 2014.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [52] J. Masci, U. Meier, D. Cireşcan, and J. Schmidhuber, “Stacked convolutional auto-encoders for hierarchical feature extraction,” in *International conference on artificial neural networks*, 2011, pp. 52–59.
- [53] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv Prepr. arXiv1511.06434*, 2015.
- [54] Y. Lin *et al.*, “Large-scale image classification: fast feature extraction and svm training,” in *CVPR 2011*, 2010, pp. 1689–1696.
- [55] J. Sánchez and F. Perronnin, “High-dimensional signature compression for large-scale image classification,” in *CVPR 2011*, 2011, pp. 1665–1672.
- [56] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” *Lect. Notes Comput. Sci.*, pp. 818–833, 2013.

- [57] X. Zeng *et al.*, “Crafting gbd-net for object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2109–2123, 2016.
- [58] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.
- [59] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [60] K. Sun *et al.*, “High-resolution representations for labeling pixels and regions,” *arXiv Prepr. arXiv1904.04514*, 2019.
- [61] Q. Xie, M.-T. Luong, E. Hovy, and Q. V Le, “Self-training with noisy student improves imagenet classification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10687–10698.
- [62] X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, “Scaling Vision Transformers,” *arXiv Prepr. arXiv2106.04560*, pp. 1–31, 2021.
- [63] F. C. Chen and M. R. Jahanshahi, “NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4392–4400, 2018.
- [64] D. Tabernik, S. Šela, J. Skvarč, and D. Skočaj, “Segmentation-based deep-learning approach for surface-defect detection,” *J. Intell. Manuf.*, vol. 31, no. 3, pp. 759–776, 2020.

- [65] J. Shi, W. Yin, Y. Du, and J. Folkesson, "Automated Underwater Pipeline Damage Detection using Neural Nets," in *ICRA 2019 Workshop on Underwater Robotics Perception*, 2019.
- [66] H. Yang, S. Mei, K. Song, B. Tao, and Z. Yin, "Transfer-Learning-Based Online Mura Defect Classification," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 116–123, 2018.
- [67] Z. Tang, E. Tian, Y. Wang, L. Wang, and T. Yang, "Nondestructive Defect Detection in Castings by Using Spatial Attention Bilinear Convolutional Neural Network," *IEEE Trans. Ind. Informatics*, vol. 17, no. 1, pp. 82–89, 2021.
- [68] Y. He, K. Song, Q. Meng, and Y. Yan, "An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, 2020.
- [69] X. Dong, C. J. Taylor, and T. F. Cootes, "Small defect detection using convolutional neural network features and random forests," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, p. 0.
- [70] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proc. IEEE Conf. Comput. Vis. pattern recognition.*, pp. 3431–3440, 2015.
- [71] J. Balzategui, L. Eciolaza, and N. Arana-Arexolaleiba, "Defect detection on Polycrystalline solar cells using Electroluminescence and Fully Convolutional Neural Networks," *Proc. 2020 IEEE/SICE Int. Symp. Syst. Integr. SII 2020*, pp. 949–953, 2020.

- [72] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.
- [73] S. Zambal, C. Heindl, C. Eitzinger, and J. Scharinger, "End-to-end defect detection in automated fiber placement based on artificially generated data," in *Fourteenth International Conference on Quality Control by Artificial Vision*, 2019, vol. 11172, p. 111721G.
- [74] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "Deepcrack: Learning hierarchical convolutional features for crack detection," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1498–1512, 2018.
- [75] M. R. U. Rahman and H. Chen, "Defects Inspection in Polycrystalline Solar Cells Electroluminescence Images Using Deep Learning," *IEEE Access*, vol. 8, pp. 40547–40558, 2020.
- [76] Y. Huang, C. Qiu, Y. Guo, X. Wang, and K. Yuan, "Surface Defect Saliency of Magnetic Tile," *IEEE Int. Conf. Autom. Sci. Eng.*, vol. 2018-Augus, pp. 612–617, 2018.
- [77] H. Han, C. Gao, Y. Zhao, S. Liao, L. Tang, and X. Li, "Polycrystalline silicon wafer defect segmentation based on deep convolutional neural networks," *Pattern Recognit. Lett.*, vol. 130, pp. 234–241, 2020.
- [78] H. Yang, Y. Chen, K. Song, and Z. Yin, "Multiscale Feature-Clustering-Based Fully Convolutional Autoencoder for Fast Accurate Visual Inspection of Texture Surface Defects," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 3, pp. 1450–1467, 2019.

- [79] A. Makhzani and B. Frey, “Winner-take-all autoencoders,” *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 2791–2799, 2015.
- [80] L. Ruff, R. A. Vandermeulen, N. Görnitz, A. Binder, E. M. Uller, and M. Kloft, “Deep Support Vector Data Description for Unsupervised and Semi-Supervised Anomaly Detection,” *Proc. ICML 2019 Work. Uncertain. Robustness Deep Learn. Long Beach, CA, USA*, 2019.
- [81] J. Lian *et al.*, “Deep-Learning-Based Small Surface Defect Detection via an Exaggerated Local Variation-Based Generative Adversarial Network,” *IEEE Trans. Ind. Informatics*, vol. 16, no. 2, pp. 1343–1351, 2020.
- [82] S. Niu, B. Li, X. Wang, and H. Lin, “Defect Image Sample Generation with GAN for Improving Defect Recognition,” *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1611–1622, 2020.
- [83] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, “Efficient GAN-Based Anomaly Detection,” 2018.
- [84] L. Deecke, R. Vandermeulen, L. Ruff, S. Mandt, and M. Kloft, “Image anomaly detection with generative adversarial networks,” *Jt. Eur. Conf. Mach. Learn. Knowl. Discov. databases*, pp. 3–17, 2018.
- [85] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, “f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks,” *Med. Image Anal.*, vol. 54, pp. 30–44, 2019.
- [86] F. Zhou, G. Liu, F. Xu, and H. Deng, “A generic automated surface defect detection based on a bilinear model,” *Appl. Sci.*, vol. 9, no. 15, 2019.

- [87] X. Kou, S. Liu, K. Cheng, and Y. Qian, "Development of a YOLO-V3-based model for detecting defects on steel strip surface," *Meas. J. Int. Meas. Confed.*, vol. 182, no. May, 2021.
- [88] F. Mushtaq, K. Ramesh, S. Deshmukh, T. Ray, and C. Parimi, "Nuts & bolts : YOLO-v5 and image processing based component identification system," *Eng. Appl. Artif. Intell.*, vol. 118, no. September 2022, 2023.
- [89] X. Cui, Q. Wang, J. Dai, R. Zhang, and S. Li, "Intelligent recognition of erosion damage to concrete based on improved YOLO-v3," *Mater. Lett.*, vol. 302, no. June, pp. 1–4, 2021.
- [90] W. Li *et al.*, "Deep learning based online metallic surface defect detection method for wire and arc additive manufacturing," *Robot. Comput. Integr. Manuf.*, vol. 80, no. September 2022, 2023.
- [91] X. Chen, T. Karin, and A. Jain, "Automated defect identification in electroluminescence images of solar modules," *Sol. Energy*, vol. 242, no. June, pp. 20–29, 2022.
- [92] S. Cheng *et al.*, "Wheel hub defect detection based on the DS-Cascade RCNN," *Meas. J. Int. Meas. Confed.*, vol. 206, no. June 2022, 2023.
- [93] C. Ji, H. Wang, and H. Li, "Defects detection in weld joints based on visual attention and deep learning," *NDT E Int.*, vol. 133, no. September 2022, 2023.
- [94] D. Li *et al.*, "Automatic defect detection of metro tunnel surfaces using a vision-based inspection system," *Adv. Eng. Informatics*, vol. 47, no. May 2020, 2021.

- [95] J. Cao, G. Yang, and X. Yang, "A Pixel-Level Segmentation Convolutional Neural Network Based on Deep Feature Fusion for Surface Defect Detection," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021.
- [96] L. Yang, J. Fan, B. Huo, E. Li, and Y. Liu, "A nondestructive automatic defect detection method with pixelwise segmentation," *Knowledge-Based Syst.*, vol. 242, 2022.
- [97] J. Fioresi *et al.*, "Automated Defect Detection and Localization in Photovoltaic Cells Using Semantic Segmentation of Electroluminescence Images," *IEEE J. Photovoltaics*, vol. 12, no. 1, pp. 53–61, 2022.
- [98] E. Versini, L. Snidaro, and A. Liani, "SCL—Segmentation—Classification combined Loss for surface defect detection," *Expert Syst. Appl.*, vol. 198, no. March, 2022.
- [99] Z. He and Q. Liu, "Deep Regression Neural Network for Industrial Surface Defect Detection," *IEEE Access*, vol. 8, pp. 35583–35591, 2020.
- [100] T. Liu and Z. He, "TAS2-Net: Triple-Attention Semantic Segmentation Network for Small Surface Defect Detection," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022.
- [101] L. Yang, S. Song, J. Fan, B. Huo, E. Li, and Y. Liu, "An Automatic Deep Segmentation Network for Pixel-Level Welding Defect Detection," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022.
- [102] W. Chen, B. Zou, J. Yang, C. Huang, P. Yao, and J. Liu, "The machined surface defect detection of improved superpixel segmentation and two-level region aggregation based on machine vision," *J. Manuf. Process.*, vol. 80, no. May, pp. 287–301, 2022.

- [103] L. Shao, E. Zhang, Q. Ma, and M. Li, "Pixel-Wise Semisupervised Fabric Defect Detection Method Combined With Multitask Mean Teacher," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022.
- [104] C. Xia, Z. Pan, Z. Fei, S. Zhang, and H. Li, "Vision based defects detection for Keyhole TIG welding using deep learning with visual explanation," *J. Manuf. Process.*, vol. 56, no. March, pp. 845–855, 2020.
- [105] W. Ouyang, B. Xu, J. Hou, and X. Yuan, "Fabric Defect Detection Using Activation Layer Embedded Convolutional Neural Network," *IEEE Access*, vol. 7, no. 2, pp. 70130–70140, 2019.
- [106] C. Zhang, S. Feng, X. Wang, and Y. Wang, "ZJU-Leaper: A Benchmark Dataset for Fabric Defect Detection and a Comparative Study," *IEEE Trans. Artif. Intell.*, vol. 1, no. 3, pp. 219–232, 2021.
- [107] Z. Zhu, G. Han, G. Jia, and L. Shu, "Modified DenseNet for Automatic Fabric Defect Detection with Edge Computing for Minimizing Latency," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9623–9636, 2020.
- [108] Z. Pan, J. Yang, X. er Wang, F. Wang, I. Azim, and C. Wang, "Image-based surface scratch detection on architectural glass panels using deep learning approach," *Constr. Build. Mater.*, vol. 282, pp. 1–11, 2021.
- [109] Q. Wang, R. Yang, C. Wu, and Y. Liu, "An effective defect detection method based on improved Generative Adversarial Networks (iGAN) for machined surfaces," *J. Manuf. Process.*, vol. 65, no. September 2020, pp. 373–381, 2021.

- [110] J. Lee, Y. C. Lee, and J. T. Kim, "Fault detection based on one-class deep learning for manufacturing applications limited to an imbalanced database," *J. Manuf. Syst.*, vol. 57, no. October, pp. 357–366, 2020.
- [111] H. Alqahtani, S. Bharadwaj, and A. Ray, "Classification of fatigue crack damage in polycrystalline alloy structures using convolutional neural networks," *Eng. Fail. Anal.*, vol. 119, no. July 2020, p. 104908, 2021.
- [112] J. Wang, P. Fu, and R. X. Gao, "Machine vision intelligence for product defect inspection based on deep learning and Hough transform," *J. Manuf. Syst.*, vol. 51, no. March, pp. 52–60, 2019.
- [113] Y. T. Li, P. Kuo, and J. I. Guo, "Automatic Industry PCB Board DIP Process Defect Detection System Based on Deep Ensemble Self-Adaption Method," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 11, no. 2, pp. 312–323, 2021.
- [114] B. Hu and J. Wang, "Detection of PCB Surface Defects with Improved Faster-RCNN and Feature Pyramid Network," *IEEE Access*, vol. 8, pp. 108335–108345, 2020.
- [115] M. Mehta and C. Shao, "Federated learning-based semantic segmentation for pixel-wise defect detection in additive manufacturing," *J. Manuf. Syst.*, vol. 64, no. May, pp. 197–210, 2022.
- [116] S. Chava and S. Namilae, "In Situ Investigation of the Kinematics of Ply Interfaces during Composite Manufacturing," *J. Manuf. Sci. Eng. Trans. ASME*, vol. 143, no. 2, pp. 1–10, 2021.

- [117] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.
- [118] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [119] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.
- [120] I. Goodfellow *et al.*, “Generative adversarial nets,” *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.
- [121] K. Song and Y. Yan, “A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects,” *Appl. Surf. Sci.*, vol. 285, pp. 858–864, 2013.