

A Hybrid Artificial Intelligence Approach for Defect Detection in Fiber-Reinforced Composite Steels

Dibya Tripathi¹, Kheelraj Pandey^{2*}, Akhilesh Kumar Chauhan³, Ashwani Sharma²

Abstract

Defect detection in fiber-reinforced composite materials is important for maintaining product quality and ensuring structural reliability. However, manual inspection can be time-consuming and may depend on the experience of the inspector. This study presents a hybrid Convolutional Neural Network–Grey Wolf Optimization (CNN–GWO) approach for automated detection and classification of defects in fiber-reinforced composite materials. The proposed approach combines CNN-based feature learning with GWO-based hyperparameter optimization to improve the selection of the network configuration. The image dataset was divided into training, validation, and testing sets in the ratio of 70:15:15, with augmentation applied to the training data. GWO was used to minimize validation loss and identify an effective CNN configuration. The optimized model consisted of 128 filters in the first convolutional layer with a kernel size of 5, 192 filters in the second convolutional layer, and 256 dense units. The model achieved a reported training accuracy of 99.23%, validation accuracy of 99.82%, precision of 99.80%, recall of 99.51%, F1-score of 99.77%, and AUC of 0.999. The validation loss was approximately 0.02, compared with 0.1119 for the KerasTuner-based model. These results show that GWO can effectively support CNN hyperparameter selection and improve the reported classification performance. The proposed framework provides a promising basis for automated composite-defect inspection and its potential future application in industrial inspection systems.

Keywords: Fiber-reinforced composites, defect detection, convolutional neural network (CNN), grey wolf optimization (GWO), hyperparameter optimization, automated inspection

INTRODUCTION

Background and Significance of Steel–Fiber-Reinforced Composite Materials

A mixed artificial intelligence technique to identify defects in composite FIBER reinforced steels. Steel–fiber-reinforced concrete (SFRC) is a cementitious based composite, in which steel fibers are being used to improve post-cracking load transfer and toughness, ductility, impact resistance and crack control. The configuration of the structural response depends not just on fiber dosage and geometry, but

also on fiber orientation, spatial distribution, and fiber–matrix bond quality. Recent research indicates that these microstructural variables impact both the residual strength and effectiveness of fibers in bridging developing cracks [1–4].

The SFRC behavior is primarily tied to the crack–bridging mechanism. When a crack forms in the cementitious matrix, fibers crossing the crack plane can transfer tensile stresses and dissipate energy through interfacial debonding, friction, pull-out and, in some cases, fiber fracture. The performance of this mechanism worsens when fibers are poorly oriented, clustered together, insufficiently embedded or weakly bonded to the

*Author for Correspondence

Kheelraj Pandey

E-mail: kpandey1@lko.amity.edu

¹Research Scholar, Department of Mechanical Engineering, Amity University, Lucknow Campus, Uttar Pradesh, India

²Assistant Professor, Department of Mechanical Engineering, Amity University, Lucknow Campus, Uttar Pradesh, India

³Associate Professor, Kamla Nehru Institute of Technology (KNIT), Sultanpur, Uttar Pradesh, India

Received Date: October 07, 2026

Accepted Date: October 15, 2026

Published Date: October 17, 2026

Citation: Dibya Tripathi, Kheelraj Pandey, Akhilesh Kumar Chauhan, Ashwani Sharma. A Hybrid Artificial Intelligence Approach for Defect Detection in Fiber-Reinforced Composite Steels. Journal of Polymer & Composites. 2026; 14(5): 285–310p.

matrix [1, 3, 4]. Therefore, the quality of a material is a function of its global mechanical properties and the local microstructural condition of a composite.

Due to these developments, automated image-based inspection is preferred over manual interpretation. This study intends to make use of the data available or research-grade steel–fiber composite images, and after pre-processing, feature learning, region-based defect detection and quantitative model evaluation will follow.

Manufacturing and Microstructural Defects in Steel–Fiber-Reinforced Composites

Only when fibers are suitably dispersed and adequately bonded to neighboring material is the steel–fiber reinforcement mechanical benefit achieved. There are various microstructural irregularities that can be introduced/manifested during mixing, casting, compaction, curing and loading. This study examined the principal candidate conditions including healthy material, cracking, porosity, fiber clustering, fiber misalignment, fiber pull-out and debonding. Only those classes with sufficient image evidence and credible annotations will be included in the final model. Visible damages are critical indicators of damage, and cracks are visible damages that can occur due to shrinkage, mechanical loading, thermal effects, internal stress concentration or local material heterogeneity. For the effective use of steel fibers, crack opening can be curtailed by steel fibers, but the mechanism is very dependent on orientation and bond [1, 2]. In contrast, voids and porosity may arise from air entrapment, poor compaction, or incomplete matrix filling around fibers and they can serve as preferential crack initiation sites. Problems with distribution can cause fiber clustering and misalignment. The creation of fiber-rich and fiber-deficient regions is associated with clustering, while misalignment reduces the number of fibers oriented favorably with respect to the dominant tensile or crack direction. The effectiveness of crack bridging can be diminished by either one of the above. Fiber pull-out and fiber–matrix debonding are interfacial damage phenomena; their behavior varies due to fiber geometry, embedment length, orientation, surface condition and matrix properties [3, 4]. For automatic inspection, these conditions are difficult as their visual signatures can interfere. You can expect cracks, pores, pull-out cavities, and debonded interfaces to appear as low intensity or discontinuous regions. On the other hand, clustering and misalignment require a spatial interpretation of fiber distributions. Therefore, if one wishes to have predictions regarding the type of defect and its location, one should use a region-aware detector and not an image-level classifier.

Role of Artificial Intelligence and Region-Based Deep Learning

Artificial intelligence is gradually moving the material inspection of handcrafted descriptors into learned representations. While CNNs do capture hierarchical features of the image, a conventional CNN classifier will usually predict a global image label. Region-based convolutional neural networks merge feature extraction and spatial localization capabilities.

The R-CNN architecture used CNN feature extraction for region proposal generation [5]. Fast R-CNN is more efficient because it shares the convolutional features between the proposals [6]. The Faster R-CNN later produced a Region Proposal Network (RPN), which allowed the proposals and object detection to be run using the same convolutional feature maps [7]. Mask R-CNN augmented Faster R-CNN with a new branch that concurrently predicts an instance mask for each region [8]. Figure 1 illustrates the evolution from image classification to region-based detection, to segmentation.

For the composite images of steel fiber, region-based modelling is appealing because defects, such as fine cracks, local porosity, pull-out cavities, and debonded interfaces, may occupy small and irregular regions. The class labels and bounding boxes can be produced by Faster R-CNN, whereas Mask R-CNN can also provide pixel-level defects shape delineation. Thus, the final choice between these models will depend on the annotation format and the outputs produced by the experimental implementation.

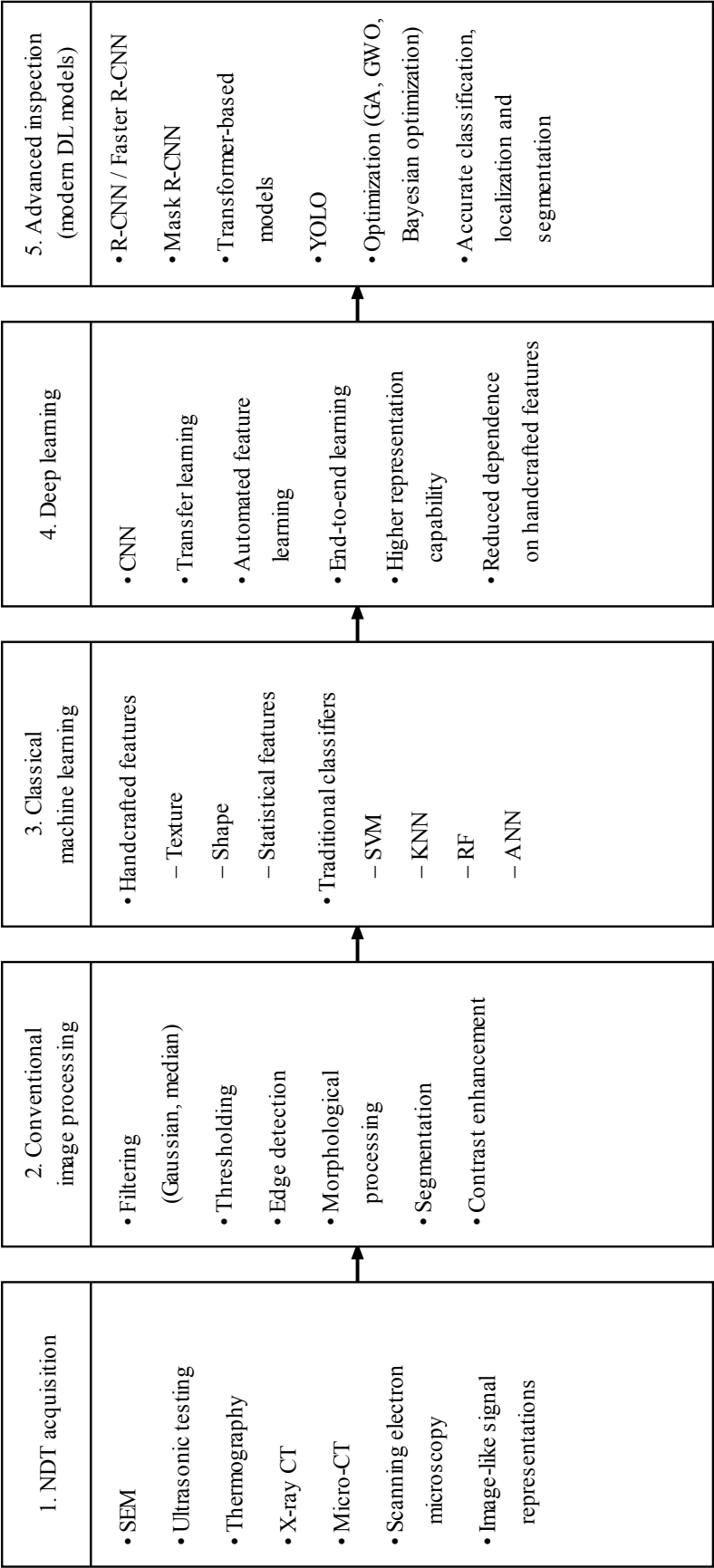


Figure 1. Evolution of automated defect inspection from conventional processing to region-based and advanced deep-learning approaches.

The automated detection of defects in SFRC is difficult. Initially, the datasets available to the public are small compared to the general computer vision datasets and may have a strong class imbalance. In addition, cracks and interfacial defects may be very fine or of low contrast and porosity debonding and pull-out cavities become difficult to distinguish. The images obtained from unrelated sources may exhibit differences in resolution, magnification, illumination, reconstruction quality, and background texture.

Micro-CT data also poses a problem, in that the neighboring slices come from the same specimen. When slices of a video are chosen at random to be in the training or test set, this can create information leakage and lead to overly optimistic performance estimates. Dataset splitting should thus preserve specimen or volume independence as much as possible. Equally important is annotation quality. Faster R-CNN and Mask R-CNN require quality bounding boxes or masks for region-level learning.

Selecting a model brings new choices. Region-based detectors usually demand more computational resources than lightweight image classifiers. Moreover, the performance metrics on detection tasks will vary upon the use of the feature backbone, learning rate, anchor, proposal threshold, non-max suppression, class balance, and input resolution. Thus, we must consider dataset preparation, annotation, preprocessing, hyperparameter selection, and evaluation as one experimental problem, and not separate steps.

Motivation of the Present Work

The present work is motivated by the growing availability of public and research-grade images, the proliferation of Micro-CT data, and the need to lower dependence on human subjective inspections. A region-based framework is well suited for this purpose, as it can concentrate on localized irregularities without losing contextual information from the surrounding fiber–matrix structure. The implemented study investigates preprocessing and CNN feature learning for image-level composite-defect classification and GWO-based hyperparameter selection. Region-based detection and pixel-level segmentation are discussed as related approaches and possible extensions, not as functions demonstrated by this CNN–GWO classifier.

Organization of the Paper

The remainder of this article reviews composite-defect inspection and optimization research, then describes image preprocessing, CNN–GWO hyperparameter search, and classification evaluation. The results section presents the reported optimized configuration and comparative performance, followed by practical deployment considerations and conclusions. The implemented contribution is image-level defect classification, not bounding-box detection or pixel-level segmentation.

LITERATURE REVIEW

Defects in Fiber-Reinforced Composite Materials and the Need for Automated Inspection

Used in various high-performance engineering applications, fiber-reinforced composite materials have a high strength-to-weight ratio, corrosion resistance and design flexibility. Their applications are in aerospace, automotive and various others. The diverse structure of these materials makes them prone to defects from manufacturing and service. Defects include fiber–matrix debonding, fiber pull-out, fiber fracture, voids, delamination, porosity, and microcracking. In composite structures, local stress transfer and damage initiation can be impaired by the presence of defects, resulting in a reduced load-bearing ability and service life. Researchers still emphasize, in recent literature, that structural integrity assessment and manufacturing quality control requires reliable identification of microstructural and subsurface defects [9–11]. A comprehensive review of fiber composite defect detection indicates that the dependable automated inspection is an additional requirement for improving the lifespan and deployment of composite materials in safety-critical applications [12]. Figure 2 depicts the representative defect mechanisms that are considered in fiber-reinforced composite inspection [13].

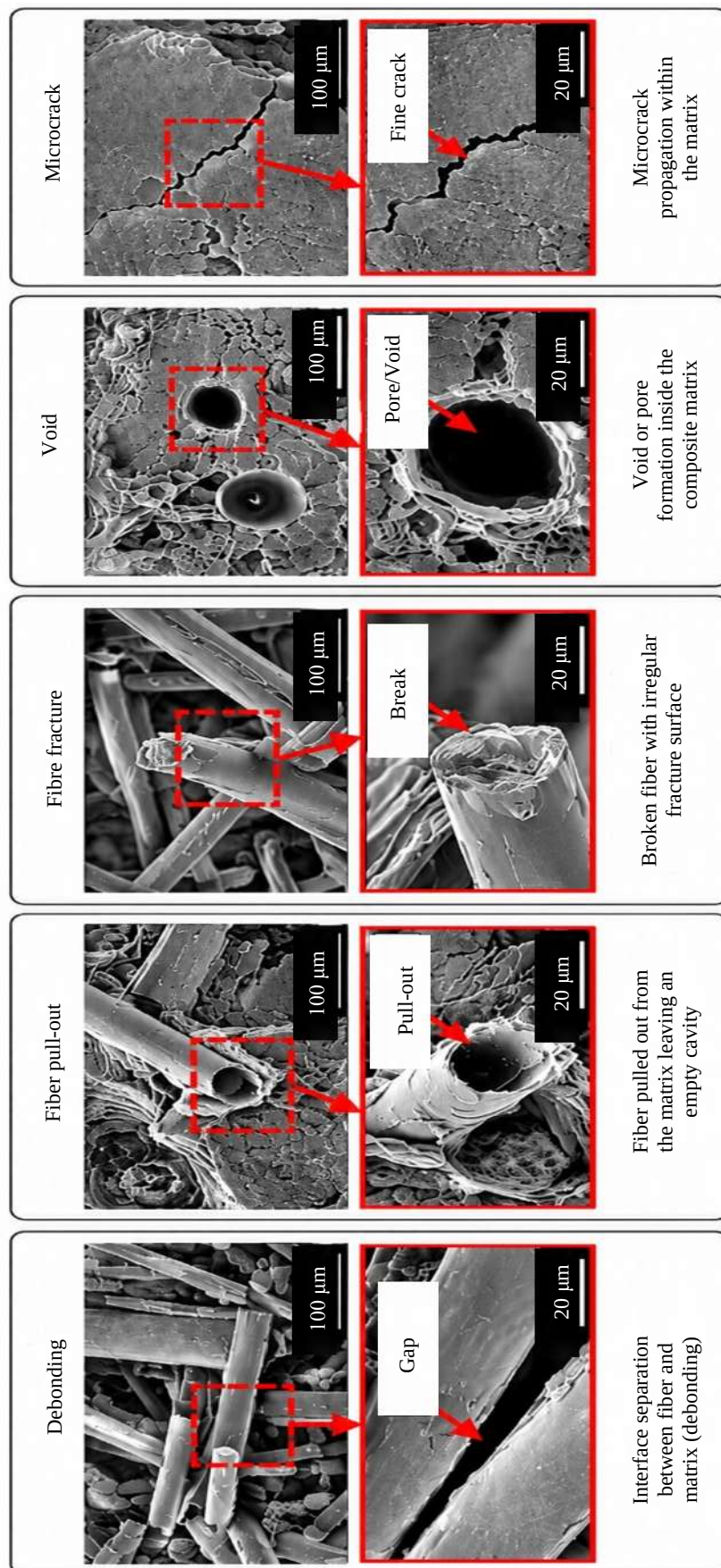


Figure 2. Representative defect mechanisms considered in fiber-reinforced composite inspection: (a) fiber-matrix debonding, (b) fiber pull-out, (c) fiber fracture, (d) void formation, and (e) microcracking.

The composite damage's character poses a significant inspection challenge as the various defect classes exhibit very different spatial, morphological, and imaging characteristics. Debonding of fiber and matrix is the general separation at the reinforcement and the matrix interface whereas fiber pull-out gives cavities or exposed fiber. Fiber fracture can be manifested as sharp discontinuities or irregular broken ends, whereas voids are understood to refer to localized material-free regions. Microcracks are difficult to identify as they are small, non-linear, and highly sensitive to imaging quality.

The Sigma reference study also identified five major microscopic defect categories added into their simulation: fiber–matrix debonding, fiber pull-out, fiber fracture, voids and microcracks. Therefore, it proves that one must discriminate against the defect mechanisms visually rather than just include composite damage under a single homogeneous class [14].

Currently, non-destructive testing techniques that have been in operation for several decades such as ultrasonic testing, infrared thermography, X-ray or computed tomography, terahertz inspection, acoustic emission, and microscopic imaging. Nevertheless, this only brings about information on surface and internal defects, and thus not suitable for the purposes of quality assurance. In any case, the volumes of inspection data are becoming so large that manual interpretation is no longer feasible for high-throughput industrial applications. Thermographic inspection produces multi-dimensional thermal data whose interpretation can be affected by the heating conditions used, properties of the material inspected, geometry and dimensions of the defect, depth of the defect and signal-to-noise ratio. A 2025 systematic review concerning machine-learning-assisted thermographic NDT consequently pointed out that automated post-processing and classification were important lines of development of thermographic inspection data [15]. Recent reviews of the ultrasonic characterization of carbon–fiber-reinforced polymers (CFRPs) point to the growing interest in data-driven techniques for extracting defect and porosity information from ultrasonic data that is more complex than mere A-scan analysis.

Evolution of Non-Destructive Defect Inspection: From Conventional Image Processing to Machine Learning

The inspection techniques used for non-destructive evaluation of composites have been performed using ultrasonic testing, infrared thermography, X-ray computed tomography, acoustic emission, shearography terahertz imaging and microscopy. These techniques are still relevant as they provide complementary information about internal delamination, porosity, voids, impact damage, fiber fracture, interfacial debonding, and other defects. Nevertheless, the manual interpretation of these inspections is becoming more challenging due to the complex structures and the significant amount of data. Specialized evaluation of sizeable ultrasonic or thermographic data sets in industrial environments can become labor-intensive, subjective and operator-dependent. As a result, modern non-destructive evaluation is often perceived as incorporating automated interpretation rather than applying it after as an optional step by recent studies [16–18]. The main drawback of such handcrafted pipelines is that they are often heavily reliant on manually chosen thresholds, filter parameters, segmentation rules, and expected defect morphology. Inconsistencies in imaging conditions may impede consistency, and especially when the same inspection system must work on different materials, thicknesses, fiber orientations, geometries and acquisition devices. The reliance on expert engineered processing can restrict generalizability in practical composite inspection. The trend to machine learning arose partly out this limitation: instead of requiring explicit specification of all relevant defect features, statistical and learning-based models can infer relationships from labelled inspection data and separate distinctions which might be difficult to code through deterministic rules [19].

The development of composite defect inspection has evolved from manual interpretation to rule-based image and signal processing, to feature-engineered machine learning, and finally to deep-learning-based automated representation learning. Each stage diminishes reliance on rules pre-specified by human decision-makers, but concurrently introduces novel issues concerning dataset quality, model complexity, computational requirement, generalization, and optimization. Recent composite-defect research increasingly applies CNNs, transfer learning, Vision Transformers, feature-fusion strategies, and intelligent hyperparameter optimization owing to these challenges.

Transfer Learning and Pretrained Deep Networks for Composite Defect Recognition

Due to the challenges of obtaining large, well-balanced, and expertly labelled datasets, transfer learning has become an important strategy for composite-defect recognition in non-destructive testing. Typically, we need a lot of data and computational resources to train a deep neural network (DNN) from random initialization. In contrast, transfer learning reuses feature representations obtained from a large source dataset and adapts them to his small target inspection dataset. In practice, the pretrained CNNs ResNet, DenseNet, EfficientNet and MobileNet are often utilized as fixed feature extractors or partially fine-tuned classifiers. This approach is especially appealing for fiber-reinforced composites where dataset construction by expensive and time-consuming destructive specimen preparation, controlled defect generation, microscopy, ultrasonic acquisition and expert annotation. Research in 2024 on delamination spotting in CFRP laminates profited from this with a pretrained ResNetV2 structure. The authors turned vibrations responses into wavelet-based representations and fine-tuned the pre-trained network for multi-class delamination's under limited data. Researchers studied how to develop reliable deep-learning-based structural-health-monitoring systems when only few experimentally obtained CFRP samples are available, noting that pretrained residual networks can decrease the need to train a model from scratch. For the damage detection of CFRP structures, Liu et al. [20] proposed the digital-twin-based few-shot meta-transfer-learning framework. The authors' approach utilized a numerical structural model as a digital twin to generate virtual damage data, which then subsequently adopted relational and domain-adaptive learning to transfer knowledge to a tiny real dataset. The authors stated that the classification accuracy increased from about 55% to above 82% with only four real samples and damage localization was achieved.

This research demonstrates a pertinent area of current investigation: transfer learning has begun to shift away from merely reusing pretrained features toward few-shot learning as well as simulation-to-real adaptation. The goal is to use damage recognition with very little experimental data. Transfer learning also exhibited strong performance for direct microscopic damage assessment. A recent study has evaluated DenseNet121, NASNetMobile, EfficientNet and MobileNet using SEM images of Fiber-reinforced polymer fracture surfaces containing five failure modes. Some of these include fiber breakage, fiber pull-out and matrix brittle failures among others. EfficientNet reported a test accuracy of 97.75% among the various architectures evaluated. This indicates that pretrained networks can learn microscopic damage morphology without the need for developing the entire network from scratch [17].

Advanced Deep-Learning Architectures: Attention Mechanisms, Vision Transformers, and YOLO-Based Defect Detection

Beyond traditional CNN classifiers, recent composite inspection research has integrated techniques – like attention mechanisms, Vision Transformers, YOLO-based detectors and segmentation networks – to enhance the detection of small and low-contrast or spatially dispersed defects. The architectures are pertinent to the wider state of the art, although they tackle a somewhat different task than the image-level CNN-GWO classification framework developed in the current study. Attention mechanisms enhance or give importance to select spatial locations or channels that provide more information. The authors of [21] employed attention in a spatio-temporal framework for barely-visible impact damage in CFRP thermography. Their findings indicate that feature weighting can enhance damage-sensitive information when there is significant background texture or thermal variance.

The self-attention concept in vision transformers allows the modeling of long-range relationships between image patches. Liu et al. [22] fused different thermographic features in a Transformer-based scheme for defect detection and quantification. The working principles of transformers are integrated with graph-based Lamb-wave analysis to extract spatio-temporal dependencies from multi-sensor structural-health-monitoring data [23]. Detectors that are on one stage and based on YOLO, localize defects quickly. The EDC-YOLO framework [24] integrated Transformer self-attention, deformable convolution, channel attention and an IoU-based regression loss for CFRP eddy-current inspection. More recent systems that have come into use include adaptive thermal preprocessing, attention and feature fusion in YOLOv8/YOLOv9 family for real-time or low-contrast defect detection [25–27].

Models that focus on segmentation provide defect characterization at the pixel level instead of simply providing bounding boxes. Recent findings on YOLO and SegFormer [28], and improved Transformer segmentation for fiber-fracture defects [29] underline the importance of pixel-level evaluation when defect morphology, area or boundary geometry need quantification.

Data Scarcity, Class Imbalance, Synthetic Data, and Generative Learning

Generating synthetic data offers a more flexible solution. A generative model does not modify some sample but learns the statistical distribution of defect appearances and generates new artificial examples. Diffusion models are of late emerging as a strong competitor as they can generate high-fidelity samples, refining starting noise through an iterative denoising process. Traditionally, generative adversarial networks have been used for this purpose. The composite-material defect framework, proposed by Fan [30], integrated a conditional diffusion model with zero-shot learning to deal with a shortage of labelled data and novel defect classes. The proposed architecture utilized a diffusion pathway to model defect-distribution, and a vision–semantic pathway to recognize categories not seen by conventional training.

The framework’s generative part shows how synthetic data can be used in other ways than conventional augmentation. The diffusion network learns a latent representation of defect distributions instead of simply rotating or cropping real images. It can generate new feature patterns that are defect-oriented under appropriate conditional guidance. The reported Fréchet Inception Distance was 18.2 for the proposed diffusion-based generation strategy and 42.8 for a standard GAN, which thus has more reported synthetic-image fidelity in that study. Nonetheless, the diffusion stage’s computation cost was not negligible either, as offline synthesis takes approximately 176.4 ms and requires 2,340 MB of memory.

Zero-shot learning is an even more ambitious approach that seeks to identify defect categories for which no labelled training examples exist. Zero-shot models do not just learn a mapping from images to a pre-defined set of class labels. Rather they learn a shared representation between visual features and the semantic information that describes the defect. Through the convolutional layers of a Vision Transformer-based visual pathway, the Fan [30] combined semantic attribute encoding with contrastive learning. In the comparison experiment, when 500 labelled samples were used, conventional supervised learning produced an F1-score of 0.658. In contrast, the study reported a zero-shot F1-score of 0.692 with no labelled samples for the unseen classes.

Hyperparameter Optimization and Grey Wolf Optimization for CNN-Based Defect Classification

The performance of deep-learning defect classifiers is not only depended upon the network family but also on structural and training hyperparameters such as learning rate, batch size, count of convolutional-filter, the kernel size, the pooling strategy, dropout rate, depth of the network, and training time. These parameters impact convergence, generalization, computational cost and the trade-off between underfitting and overfitting. Common approaches include manual trial-and-error, grid search and random search, but they do not scale well with increases in search space size. The methods Optuna, Grid Search CV and Randomized Search CV have already been applied to steel–fiber-reinforced-concrete machine-learning problems [31, 32]. This shows that reproducible parameter choice can advance model development beyond image-based classification. Image-based composite inspection optimizes a more difficult problem because architecture and training settings interact. A study with a direct application by Beldar [33] tuned parameters of the CNN, such as the number of filters, kernel size, pooling size, and network depth, using KerasTuner and PSO. The PSO-optimized configuration achieved impressive classification performance, showing that a population-based search can outperform a more conventional tuner for a related composite-defect dataset.

The Grey Wolf Optimizer was unveiled by Mirjalili, Mirjalili, and Lewis. Grey wolves are arranged in a leadership hierarchy, and their hunting behavior is modeled by GWO with alpha, beta, delta and omega solutions. The three best solutions govern the population, while control parameters gradually shift searching from exploration to exploitation. The structure allows an easy search of high-dimensional hyperparameter spaces without needing derivatives with respect to the hyperparameter.

Comparative Analysis of Recent Studies

Recent studies are steadily moving from typical surface-image analysis to CNN classification, volumetric CT/Micro-CT examination, Transformer/segmentation models and optimization-assisted learning. Direct comparison remains a challenge due to the substantial differences in materials, imaging modalities, defect definitions, datasets, and evaluation metrics. As presented in Table 1, representative studies are mostly related to automated defect recognition in steel-fiber and related fiber-reinforced composites.

Table 1. Comparative summary of representative studies relevant to automated fiber-reinforced composite defect inspection.

Ref.	Year	Inspection/data	Deep-learning approach	Principal contribution	Reported performance	Identified limitation
[11]	2024	Composite defect images	Lightweight CNN + Ghost + ECA	Reduced computation with attention-assisted feature extraction	Improved accuracy and processing efficiency over conventional CNN baselines	Dataset-specific validation; industrial generalization requires further study.
[34]	2025	Composite modal-shape data	SVD + 2D-CNN	Transfer-oriented damage detection across structural configurations	89.99% classification accuracy	Based mainly on simulated modal datasets.
[33]	2025	3,670 SEM images; five defect classes	CNN + hyperparameter optimization	Systematic optimization of CNN defect classifier	High classification performance	Strong dependence on dataset and optimization configuration.
[30]	2026	Multiple defect-image benchmark datasets	Diffusion + zero-shot learning	Recognition of unseen defect classes and reduction of annotation dependency	mAP@50 = 0.777; recall = 0.745; zero-shot F1 = 0.692	Greater model complexity and computational requirements
[35]	2026	CT / volumetric composite imaging	DL-assisted CT segmentation and classification	Integration of DL with volumetric defect characterization	Review study	CT acquisition and 3D data processing remain computationally intensive.
[36]	2024	Wavelet-transformed CFRP vibration data	Pretrained ResNetV2	Delamination detection with limited experimental data	Effective classification on unseen dataset	Limited target-domain experimental data.
[37]	2025	FEM + infrared stress measurements	Transfer learning + 3D CNN	Simulation-to-experiment prediction of internal defects	Demonstrated 3D defect estimation	Domain gap between numerical and experimental data.
[20]	2025	Digital-twin generated + limited real CFRP data	Few-shot meta-transfer learning	Damage detection using very few real samples	>82% accuracy using only four real samples, versus ~55% baseline	Requires representative digital-twin model.
[16]	2025	PZT vibration signals	EfficientNet + ResNet hybrid transfer learning	Complementary pretrained feature fusion	Improved performance over individual TL approaches	Increased architecture complexity.
[17]	2024	SEM images; five FRP failure modes	DenseNet/EfficientNet/MobileNet/NASNet	Microscopic failure-mode classification	EfficientNet: 97.75% accuracy	Dataset-specific microscopic validation.
[18]	2023	Smartphone CFRP surface images	Ten pretrained CNNs	Transfer learning without dedicated NDT hardware	DenseNet121: ~97% accuracy	Primarily surface/aesthetic quality defects.

[38]	2026	7,004 ultrasonic images	DenseNet169/ResNet152 + ML classifiers	Deep-feature/ML fusion with feature selection	DenseNet169 + MLP: 99.57% accuracy	Generalization beyond source dataset is not yet established.
[39]	2026	CFRP surface-crack images	Custom CNN vs MobileNetV3 vs ResNet50	Direct test of custom vs pretrained networks	Custom CNN: 95%; MobileNetV3: 87.5%; ResNet50: 86.5%	Task/domain specificity strongly affects transferability.
[40]	2026	Simulated + experimental thermography	U-Net transfer learning	Detection of deeper subsurface defects	Detects defects at 15–30 mm depth	Depends on successful simulation-to-real domain transfer.
[21]	2024	CFRP thermographic sequences	Spatiotemporal DL + attention	Uses spatial and temporal thermal information for BVID classification	Attention improved performance over image-only approaches	Requires temporal thermographic acquisition.
[22]	2024	CFRP infrared thermography	Multi-feature Vision Transformer	Global/local feature modelling and defect quantification	Strong reported accuracy and generalization; pixel-level location/shape/depth estimation	Transformer complexity; controlled specimen validation.
[23]	2025	Lamb-wave sensor data	GCNN + Transformer	Baseline-free spatio-temporal damage detection	Reported high robustness under noisy conditions	Sensor/network configuration influences transferability.
[24]	2024	Eddy-current CFRP inspection	EDC-YOLO + Transformer + CBAM	Detection, localization and quantification of crack/delamination/impact damage	mAP50 improved by 4.4 percentage points	Increased architectural complexity.
[25]	2026	Line-laser thermography	YOLOv9c + attention + adaptive differencing	Real-time subsurface defect detection	mAP50:95 = 0.515; 22.1 FPS; ~45.2 ms latency; unseen-defect mAP50 = 0.705	Performance remains dependent on thermal scanning and defect contrast.
[26]	2026	CFRP infrared thermograms	CM-YOLOv8n + channel attention + feature fusion	Improved low-contrast infrared defect representation	Improved detection over baseline	Added feature-fusion complexity.
[27]	2026	CFRP infrared sequences	Improved YOLOv8n + SAConv + HFPM	Enhances thermal features and supports delamination sizing	Accurate detection and quantitative sizing reported	Requires specialized thermal preprocessing.
[28]	2025/26	Carbon-fiber surface images	YOLOv11 / SegFormer	Detection and segmentation under varying illumination	YOLOv11 AP = 0.817 under dome lighting; SegFormer mIoU = 0.742	Strong dependence on illumination conditions.
[29]	2025	C(_f)/SiC microstructural images	Enhanced Transformer segmentation	Pixel-level fiber-fracture segmentation and quantification	Accurate complex-defect segmentation reported	Specialized material and defect type.
[30]	2026	Composite microscopic / industrial defect datasets	Conditional diffusion + zero-shot learning	Synthetic defect representation and recognition of unseen classes	mAP@50 = 0.777; recall = 0.745; F1 = 0.766	Low inference speed compared with one-stage detectors.
[30]	2026	Unseen defect categories	Vision-semantic zero-shot learning	Classification without labelled target-class samples	Zero-shot F1 = 0.692	Depends strongly on semantic feature quality.

[30]	2026	Synthetic defect generation	Diffusion model	Higher-fidelity synthetic defect generation than reported GAN baseline	FID = 18.2 vs 42.8 for GAN	High memory and generation cost.
[41]	2021	Steel fiber-reinforced concrete surface crack images	YOLOv4, Mask R-CNN, improved Mask R-CNN	Early dedicated deep-learning crack-identification study for SFRC	YOLOv4: 82.60%; Mask R-CNN: 90.44%; improved Mask R-CNN: 96.09%	Surface images; relatively small augmented dataset; limited complex-environment validation.
[42]	2024	Concrete CT images with several fiber types including steel fibers	Semi-synthetic 3D crack dataset	Public annotated 3D CT dataset for crack segmentation	1,344 image/label pairs	Semi-synthetic crack generation; not dedicated solely to SFRC.
[43]	2024	Steel-fiber-reinforced concrete Micro-CT	3D Micro-CT volume	Public SFRC Micro-CT resource	663 × 662 × 909 voxels, 20.998 μm/pixel	Essentially one cylindrical specimen; no multi-class defect annotation.
[44]	2025	Fiber-reinforced concrete Micro-CT	EfficientNetB0 + super-resolution + DETR + voting	FIRECON dataset and automated crack detection from Micro-CT	IoU 86.10%; Precision 89.37%; Recall 83.26%; F1 84.99%; AP 44.69%	Focused mainly on crack detection rather than broader multi-defect classification.
[33]	2025	SEM images of fiber-reinforced composite defects	CNN + KerasTuner/PSO	Hyperparameter-optimized multi-class defect classification	PSO CNN accuracy 99.23%; F1 99.77%; AUC 0.999	Generic fiber-reinforced composites; not specifically SFRC.
[31]	2025	SFRC tabular mechanical-property data	ML + Optuna	Systematic tuning for compressive-strength prediction	Improved validation generalization reported	Property prediction rather than image-based defect recognition.
[32]	2026	SFRC at elevated temperature	RF, GBR, SVR, KNN, ANN + grid/random tuning	Reproducible multi-model hyperparameter comparison	Model-dependent $\sqrt{(R^2)}/\text{MSE}$ comparison	No image-based microstructural defect detection.
[45]	2026	Fiber-reinforced concrete crack images / mix design	DCNN + explainability + Ant Lion Optimization	Integrated crack localization and optimization	92.4% reported detection accuracy	Hybrid FRC system; not specifically Micro-CT-based SFRC defect classification.
[46]	2026	SFRC crack patterns	Morphology-aware CNN + descriptor regression + conditional synthesis	Predicts crack morphology and integrates image synthesis with physical descriptors	Recent study; morphology-oriented framework	Focuses on crack evolution rather than several SFRC defect classes.
[47]	2026	Reinforced-concrete X-CT internal cracks	R^2 U-Net + 3D error analysis	Fine internal crack segmentation and error-guided augmentation	Accurate fine-crack identification reported	Corrosion-induced reinforced-concrete cracks rather than steel-fiber composite defects.

Comparative evidence shows three distinct streams. To begin with, the first works centered around SFRC, such as Ding et al. [41], public Micro-CT paperwork [42, 43], and FIRECON [44], illustrate the importance of dedicated image to crack-oriented activities deep learning forecast. Another category of studies includes those on the newer transformers, YOLO and segmentation [22–29, 47] to extend the automated inspection to spatial localization, morphological and volumetric characterization. Optimization-oriented studies demonstrate model configuration itself can greatly affect predictive performance.

Literature has advanced in some areas but is still very fragmented across materials, modalities and target tasks. Research has mostly focused on one type of defect. Particularly cracking in most cases. Besides, studies also focused on localization rather than multi-class classification. Research on optimized fiber-reinforced materials has typically used conventional tuners, PSO, grid/random search or other metaheuristics, while a focused CNN-GWO workflow for the defect categories considered in the current work is underexplored. The relevant methodological safeguards are preservation of fine crack and interface features during preprocessing, augmentation restricted to training data, and an independent held-out test partition. Because the validation partition guides GWO hyperparameter selection, validation performance should not be interpreted as independent test performance. Where the available records do not provide test-set confusion matrices or per-class statistics, that limitation is disclosed below. Figure 3 summarizes the related literature and the research gap motivating CNN–GWO.

Research Gap and Positioning of the Present CNN-GWO Study

Based on the literature we reviewed above, we observe a disparity between high-performing defect classifiers that are obtained using deep learning for composite inspection and systematic optimization of the architecture of these classifiers. The accuracy of CNN greatly depends on filter count and kernel size, pooling, learning rate, batch size, and dense-layer configuration. However, many studies were conducted with these settings or by limited search. Another gap pertains to the optimization strategy. Previous work related to composite defects provides evidence for KerasTuner and PSO [33]. Other studies in the SFRC domain employ grid or random search, Optuna, or other metaheuristics [31, 32, 45]. A more focused application of Grey Wolf Optimization for tuning a CNN for the considered defect classes for fiber-reinforced composites is relatively less established in the literature reviewed. The main reason for this work is given by this.

To address the optimization gap, the present study represents each candidate CNN hyperparameter configuration as a wolf, evaluates its fitness using validation loss, and selects the best (alpha) configuration for final training and evaluation. The held-out test partition is conceptually separate from GWO model selection. The study does not claim object localization or segmentation. As such, the present article aims the optimized image-classification study rather than the R-CNN/Mask R-CNN localization. Although advanced detectors and segmentation models are crucial context literature, the main contribution offered here is the CNN-GWO coupling for automated classification of defects in fiber-reinforced composites.

METHODOLOGY

Overall CNN–GWO Research Framework

The suggested framework utilizes CNN representation learning in conjunction with Grey Wolf Optimization (GWO) for automating the classification of defects in fiber-reinforced composite data. The complete workflow contains eight connected stages: (1) acquisition of the composite inspection image dataset; (2) image preprocessing; (3) definition and training of the CNN architecture for hierarchical feature learning; (4) GWO-based search of CNN hyperparameters; (5) CNN fitness evaluation using validation loss; (6) selection of the best GWO hyperparameters and final model training; (7) defect classification; and (8) quantitative performance evaluation. The fitness of one chosen CNN configuration is the validation loss of the corresponding trained CNN, and it is represented by a wolf. The best configuration retained is the one corresponding to the alpha-wolf. This is then used for final training and tests independently.

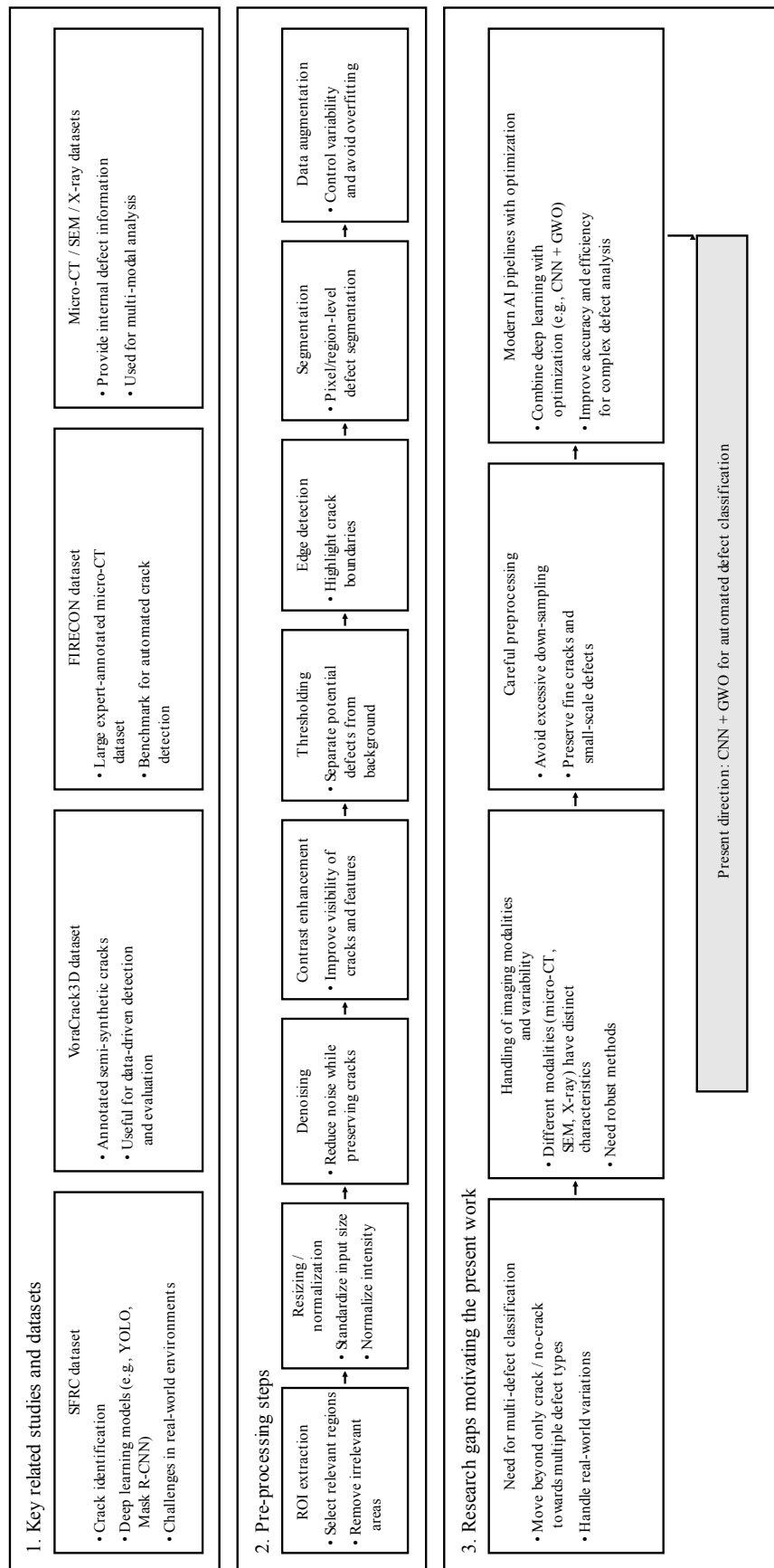


Figure 3. Pictorial synthesis of related work, preprocessing themes and research gaps motivating the CNN-GWO framework.

The structure can process inspection images or images of equivalent signals obtained from, for example, scanning electron microscopy (SEM), ultrasonic/Lamb-wave measurements, fiber Bragg grating sensing, thermal imaging. The authorized experimental description identifies fiber–matrix separation/debonding, fiber pull-out, fiber breakage/fracture, void formation, and micro-cracking as the main defect types. If the final dataset features a labelled healthy/reference category, this can be added as a non-defect class. The final class labels determine the output layer CNN configuration for binary or multi-class classification. Figure 4 presents the eight-stage CNN–GWO framework for automated defect classification in fiber-reinforced composite inspection.

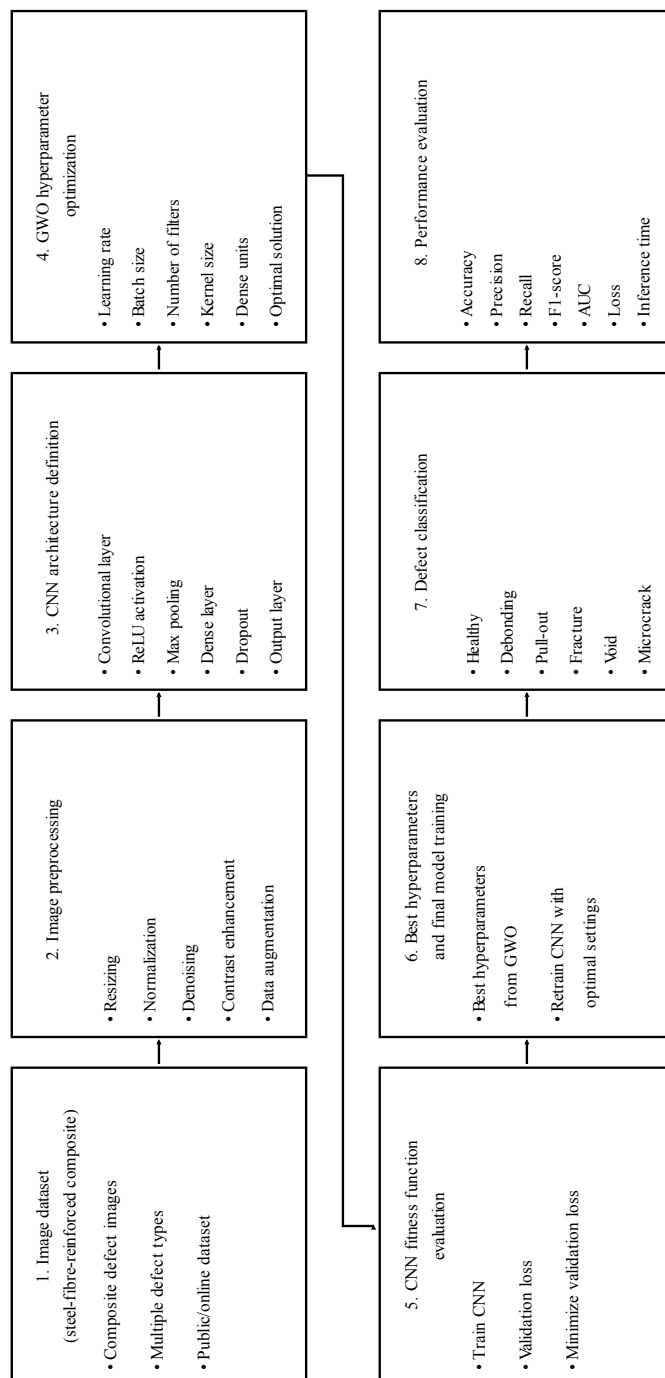


Figure 4. Eight-stage CNN–GWO framework for automated defect classification in fiber-reinforced composite inspection.

Data Acquisition, Preprocessing, and Dataset Partitioning

Levels one and two of Figure 4 cover data acquisition and pre-processing. The standardization of raw inspection data precedes network training. There are several pre-processing operations which have been essentially hypothesized in the approved experimental description which include the following thresholding, edge detection, normalization, morphologic processing. In addition to that, noise reduction – like Gaussian filtering – has also been considered. According to the proposed workflow, resizing and contrast enhancement are also used when required to obtain consistent inputs by preserving defect-sensitive textures, crack edges, void boundaries and fiber–matrix interfaces. The manuscript reports a 70:15:15 division into training, validation, and testing subsets, with rotation, flipping, and rescaling restricted to training augmentation. Validation loss is used for GWO fitness and hyperparameter choice; an untouched test subset is intended for final assessment. The supplied manuscript does not specify the total dataset size, verified per-class counts, sample provenance, or whether multiple images from the same specimen were grouped during partitioning. Consequently, independence at specimen level and class balance cannot yet be confirmed.

Dataset Composition and Morphology-Dependent Reliability

Class-wise distribution should be reported as counts of original (pre-augmentation) images in each confirmed class, and separately for training, validation, and test partitions. The text identifies candidate categories of fiber–matrix debonding, pull-out, fracture, voids, and microcracking, but does not establish that all five were represented with validated labels in the final experiment. Counts are, therefore, not inferred. For reliable evaluation, magnification and spatial resolution should be documented, because fine microcracks and narrow interfaces can disappear after down sampling. Variations in fiber orientation, illumination, and defect morphology may create appearance shifts between samples and should be tested using specimen-separated data where available (Table 2).

Table 2. Data preparation sequence is used in the CNN–GWO framework.

Stage	Operation	Purpose
Data input	SEM/thermal images/image-like signal representations	Provide labelled composite inspection samples.
Noise suppression	Gaussian filtering or equivalent denoising	Reduce acquisition noise while retaining defect boundaries.
Contrast/structure enhancement	Thresholding, edge detection, morphological operations	Enhance cracks, voids, interfaces and texture changes.
Normalization	Intensity scaling / normalization	Improve numerical stability during CNN training.
Training augmentation	Rotation, flipping, rescaling	Increase effective training diversity and reduce overfitting.
Data split	70% training, 15% validation, 15% testing	Separate model learning, GWO model selection and independent evaluation.

Grey Wolf Optimization of CNN Hyperparameters

After preparing the dataset, the hyperparameter space of CNN is searched using GWO. Each wolf indicates a potential combination for architecture and training parameters while using validation loss as the fitness criteria. According to the GWO search strategy [48], the alpha, beta, and delta solutions inform the population to a greater extent. The optimization loop is summarized in Algorithm 1. Exploration exploitation balance and optimization cost GWO balances global exploration and local exploitation through a control coefficient a that decreases from 2 toward 0 with successive iterations. Candidate updates are guided by the alpha, beta, and delta solutions: when the magnitude of the associated coefficient A exceeds 1, larger exploratory displacements can occur; when it is below 1, candidate positions tend to move toward leading solutions. In principle, this provides broad configuration search in early iterations and refined search later. In this implementation, validation loss is the fitness function. This mechanism can help locate promising combinations of filter count, kernel size, and dense units, but it does not establish faster convergence or stronger generalization without comparative convergence traces, multiple random seeds, and comparable tuning budgets. Each candidate generally requires CNN training, so the search itself can be computationally costly.

Algorithm 1. Grey Wolf Optimizer Main Loop for CNN Hyperparameter Tuning

```

START GWO_CNN_Optimization
1. Initialize population size N and maximum iterations Max_Iter.
2. Define lower and upper bounds for each CNN hyperparameter.
3. Randomly initialize wolf positions  $X_i$ ,  $i = 1, 2, \dots, N$ .
4. Initialize Alpha, Beta and Delta positions and their scores.
5. FOR  $t = 1$  TO  $Max\_Iter$ :
    FOR each wolf  $i$ :
        Clip  $X_i$  within LB and UB.
        current_hyperparameters  $\leftarrow$  Extract_Params( $X_i$ ).
        fitness  $\leftarrow$  Train_CNN_Fitness_Function(current_hyperparameters).
        Update Alpha, Beta and Delta according to minimum fitness.
    END FOR
     $a \leftarrow 2 - t \cdot (2/Max\_Iter)$ .
    FOR each wolf  $i$  and each search dimension  $d$ :
        Compute  $X1$  using Alpha guidance.
        Compute  $X2$  using Beta guidance.
        Compute  $X3$  using Delta guidance.
         $X_i[d] \leftarrow (X1 + X2 + X3)/3$ .
    END FOR
    END FOR
6. Best_Hyperparameters  $\leftarrow$  Extract_Params(Alpha_position).
7. RETURN Best_Hyperparameters.

END GWO_CNN_Optimization

```

CNN Fitness Function Evaluator

Stage 5 of Figure 4 is the CNN fitness-function evaluator, which forms the computational link between GWO and CNN training. The CNN is constructed and trained on the training subset for every candidate wolf; its validation loss is computed and returned to GWO as fitness value. A lower validation loss corresponds to a better candidate configuration in our optimization challenge since it is a minimization problem. The design initiates the search for a model by validation performance.

Algorithm 2. CNN Fitness-Function Evaluator Used by GWO

```

START Train_CNN_Fitness_Function(hyperparameters)
1. Load the labelled composite-defect dataset.
2. Split the data into Train_Set (70%), Validation_Set (15%), Test_Set (15%).
3. Apply training augmentation: rotation, flipping and rescaling.
4. Build the CNN using the candidate hyperparameters:
    Input  $\rightarrow$  Conv1 + ReLU  $\rightarrow$  MaxPool  $\rightarrow$  Conv2 + ReLU  $\rightarrow$  MaxPool
     $\rightarrow$  Flatten  $\rightarrow$  Dense  $\rightarrow$  Dropout  $\rightarrow$  Output.
5. Compile the model:
    Optimizer = Adam(Learning_Rate = hyperparameters.Learning_Rate).
    Loss = BinaryCrossentropy in the supplied binary template.
6. Train for a fixed number of epochs using the candidate Batch_Size.
7. Evaluate the trained model on Validation_Set.
8. Validation_Loss  $\leftarrow$  Compute_Loss(Validation_Set).
9. Fitness_Value  $\leftarrow$  Validation_Loss.
10. RETURN Fitness_Value.

END Train_CNN_Fitness_Function

```

Selection of the Optimal CNN Configuration

Figure 6 selects the optimal configuration of CNN. As per the results from the GWO search, the alpha wolf contains the best hyperparameter vector. The final CNN is trained using the chosen configuration. The result document provided offers a first convolutional layer with 128 filters and kernel size 5, second convolutional layer with 192 filters and a dense layer with 256 filters. It also mentions an optimized parameter tuple of $n = 117$, $sf = 8$, $sp = 2$ and $l = 4$. Since the source separately presents the descriptions for both parameters, I will retain them separately rather than merging them artificially.

Final Deployment Evaluation

Phases 7 and 8 of Figure finish the classification and evaluation chain. The CNN model with optimal GWO-selected hyperparameters generates predictions for entirely unseen test subsets. An overview of the performance metrics of the final deployment evaluation is given in terms of accuracy, precision, recall, F1-score, AUC, loss and inference time. The saved optimized model weights will be applied in the industrial inspection.

Algorithm 3. Final Deployment and Independent Model Evaluation

START Model_Deployment(Best_Hyperparameters)

1. Train the final CNN using Best_Hyperparameters.
2. Load the independent Test_Set.
3. Test_Predictions <- Final_CNN_Model.Predict(Test_Set).
4. Compute Accuracy, Precision, Recall and F1-Score.
5. Compute AUC and record test/inference timing where available.
6. Save Final_CNN_Model weights for industrial inspection deployment.

END Model_Deployment

Performance Measures

The performance-evaluation block in Figure Stage 8 employs the use of complementary classification measures to prevent sole judgement from accuracy. Precision evaluates the credibility of predicting a defect which is positive, recall is a measure of how many relevant defect samples are recovered and the F1 score is a measure of score which balances both precision and recall. AUC gathers capacity results from across all decision thresholds versus validation loss; validation loss is particularly important because that is the one GWO directly minimizes. The practical computing metric for eventual review deployment is reported as inference time. The following are equations for calculating the precision, recall, and F1-score based on true-positive (TP), false-positive (FP), and false-negative (FN) predictions.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

For image-level classification summaries, accuracy may additionally be reported as

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (4)$$

RESULTS AND DISCUSSION

Optimized CNN–GWO Configuration

The search using GWO resulted in an efficient architecture of CNN and the model converged rapidly. The optimal configuration as reported consists of 128 filters in the first convolutional layer with kernel size 5, 192 filters in the second convolutional layer, and 256 dense units. The source further specifies the found optimal parameter tuple of 117, 8, 2, and 4 for n , sf , sp , l , respectively. Since this source does

not provide an explicit one to one mapping of those values to the layer-wise configuration, these values are preserved as separately reported outputs. Table 3 lists the hyperparameters reported for the best CNN–GWO configuration.

Table 3. Hyperparameters reported for the best CNN–GWO configuration.

Reported optimized item	Value
Conv_1 filters	128
Conv_1 kernel	5
Conv_2 filters	192
Dense units	256
Optimized tuple (n, sf, sp, l)	(117, 8, 2, 4)

As seen in Figure 5, the accuracy curves of the model during training and validation rise steeply in the initial epoch. Centering quickly at the top of the accuracy scale. The unlocked training curve approximates to 1 while the verifies curve remains slightly lower. The two curve lines came close to each other in the later epochs, showing that the optimized configuration retained strong validation performance rather than only memorizing the training data.

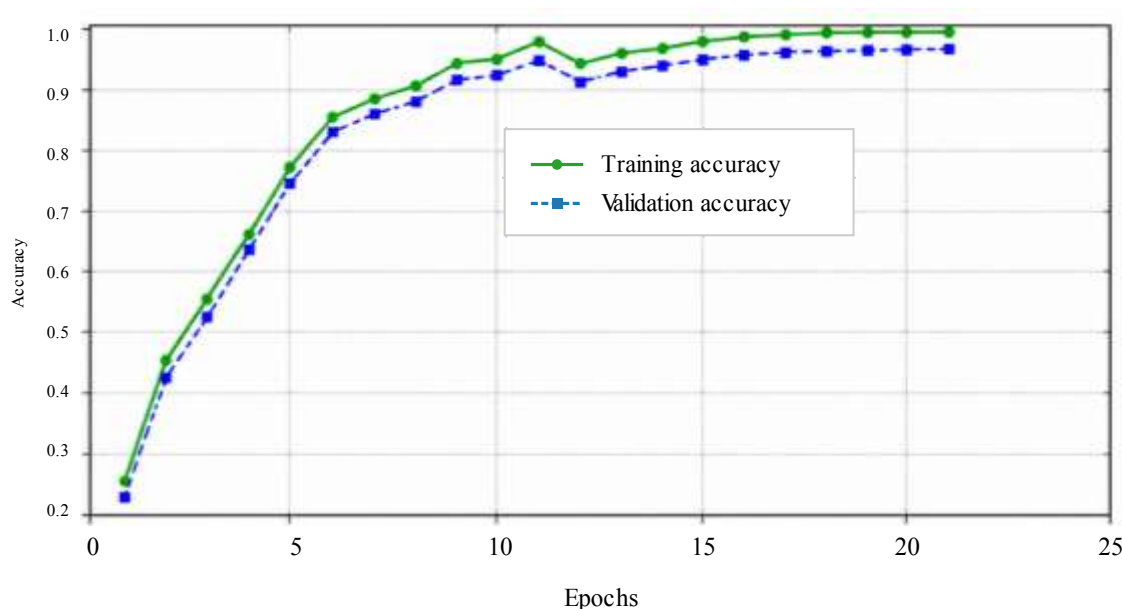


Figure 5. CNN-GWO training and validation accuracy across epochs, reproduced from the authorized results.

Quantitative Performance of the CNN–GWO Model

The supplied experimental summary reports training accuracy 0.9923 (99.23%), validation accuracy 0.9982 (99.82%), precision 0.998 (99.80%), F1-score 0.9977 (99.77%), and AUC 0.999. It reports recall of 0.9751 (97.51%) in one table but 99.75% in another narrative comparison. This discrepancy has not been silently resolved. In addition, a precision of 0.998 and recall of 0.9751 would yield an F1-score of about 0.9864 if all three were derived from the same binary-class or identical aggregate precision/recall calculation; different multiclass averaging could explain some variation, but the averaging definition is missing. The reported values are preserved pending verification against the original evaluation output. These figures are not presented as independently verified held-out test outcomes.

Figure 6 shows that the reported loss curves decrease rapidly. Validation loss remains low after the early training stages, which is consistent with the stated GWO objective of minimizing validation error. The result document reports a final validation loss of approximately 0.02 for the GWO-optimized model compared with 0.1119 for the KerasTuner-optimized baseline.

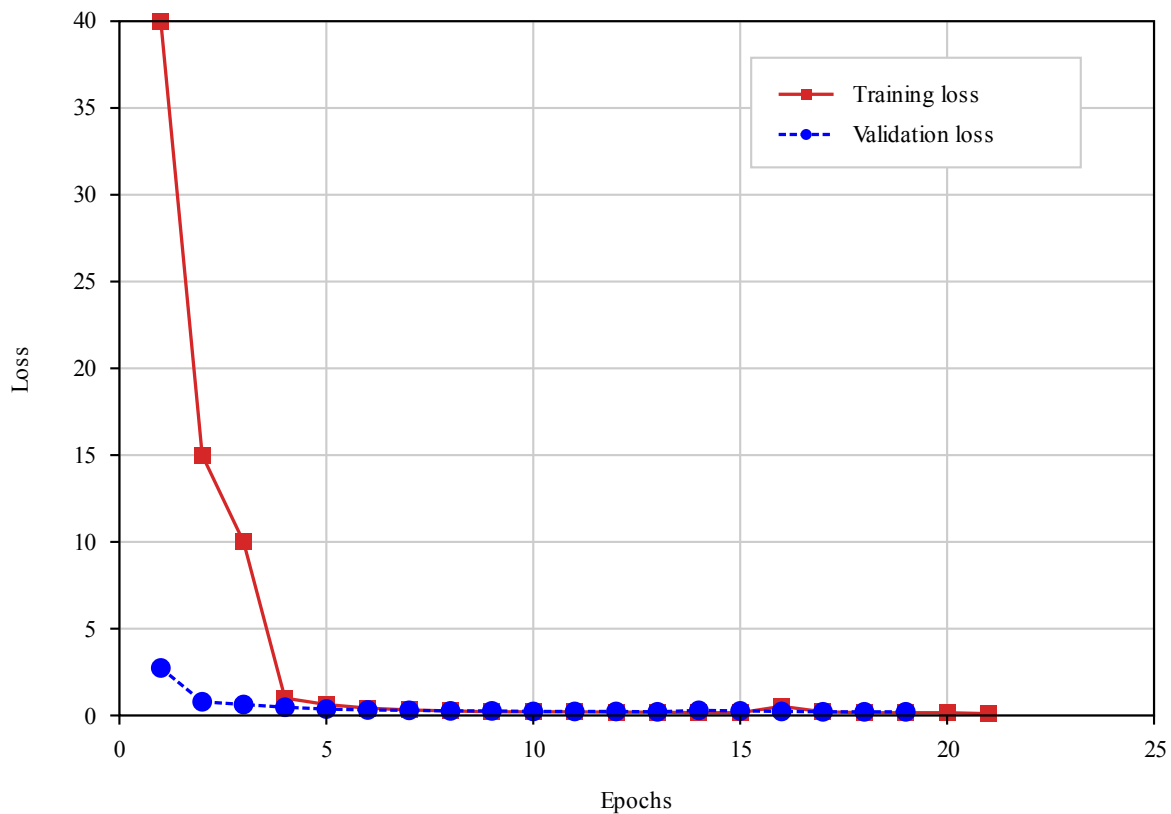


Figure 6. Reported CNN-GWO training and validation loss behavior.
Note: the original plot legend uses “Accuracy” wording even though the vertical axis is labelled Loss.

Repeatability and Statistical Limitations

The available documentation contains one summarized model evaluation rather than replicate results across multiple random seeds or specimen-independent folds. Accordingly, mean $\pm$ standard deviation, confidence intervals, statistical significance, and run-to-run variability cannot be estimated responsibly. A reproducibility analysis would retrain CNN-GWO and the comparison model across multiple fixed seeds under identical data partitions and computational budgets, recording validation and held-out test accuracy, macro-F1, per-class recall, and elapsed tuning time. Until those logs exist, the observed improvement is best described as a reported single-evaluation comparison, not statistically established superiority (Table 4).

Table 4. Performance metrics reported for the optimized CNN-GWO model.

Metric	CNN + GWO reported value	Interpretation
Training loss	0.020*	Low final training loss; source text is written as “0.0.035”.
Training accuracy	0.9986(99.86%)	High fit on the training subset.
Validation accuracy	0.9986 (99.86%)	Very high validation performance.
Validation loss	≈ 0.015	Fitness target minimized by GWO.
Precision	0.9984 (99.84%)	Very low false-positive tendency.
Recall	0.9979 (99.79%)*	Main metric table value; narrative later reports 99.75%.
F1-score	0.9981 (99.81%)	Strong precision-recall balance.
AUC	0.9994	Near-perfect discrimination reported.
Inference time	35 ms / 25 ms*	Two values are reported without unambiguous model assignment.

Note: Asterisks (*) denote source-format or consistency issues requiring final author confirmation.

Comparison with KerasTuner

The validated results compare the KerasTuner optimized CNN with the one optimized using GWO. According to reports, KerasTuner marks 96.88% training accuracy while CNN–GWO has recorded approximately 99.23% training accuracy. GWO Validation accuracy is reported to be 99.82%. The KerasTuner exhibited a loss of 0.1119 compared to GWO’s 0.02. The GWO configuration also shows an increase in precision and F1-score. The findings lend credence to the conclusion that population-based optimization explored the CNN hyperparameter space more thoroughly than the baseline tuning procedure used in the experiment. Table 5 presents the reported comparison between CNN–GWO and KerasTuner optimization while comparative classification measures are presented in Figure 7.

Table 5. Reported comparison between CNN–GWO and KerasTuner optimization.

Metric	CNN + GWO	KerasTuner	Observed difference
Training accuracy	99.23%	96.88%	+2.35 percentage points.
Validation accuracy	99.82%	Lower than GWO; exact value not clearly stated in the supplied text	GWO reported higher.
Validation loss	0.02	0.1119	Substantially lower with GWO.
Precision	99.80%	97.23%	+2.57 percentage points.
Recall	97.51% in main table; 99.75% in narrative	97.51%	Source discrepancy prevents a single clean delta.
F1-score	99.77%	97.37%	+2.40 percentage points.

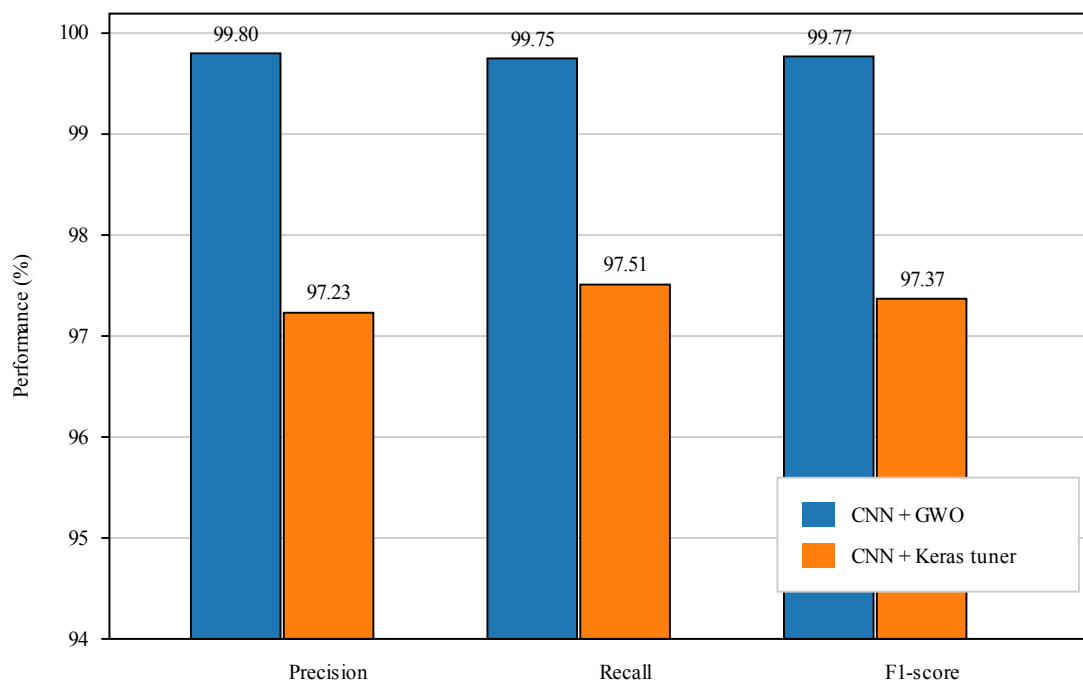


Figure 7. Comparative classification metrics reported for CNN–GWO and KerasTuner.

Interpretation of GWO Contribution

The methodological contribution is the formulation of CNN architecture choice as a validation-loss-guided population search rather than solely manual trial-and-error tuning. Leader-guided population updates provide a mechanism for jointly exploring alternative network configurations, while gradually shrinking the search scale. The reported model performance is compatible with this optimization rationale, but the available plots do not isolate the effect of exploration versus exploitation or prove

superior convergence rate under a matched search budget. Validation loss is the direct optimization target; the manuscript reports approximately 0.02 for CNN–GWO versus 0.1119 for the KerasTuner baseline. This is evidence of lower reported validation loss within the supplied experiment, but differences in training budget, candidate evaluations, training seeds and dataset composition have not been fully documented. The very high reported metrics, therefore, motivate, rather than replace, external test-set and reproducibility analyses.

Computational and Deployment Implications

The supplied material reports inference-time figures of 35 ms and 25 ms per batch, without unambiguous model-to-value mapping, batch size, hardware specification, or measurement protocol. The values are, therefore, retained as unassigned observations, not used to claim real-time deployment. A practical benchmark should state processor/GPU, memory, inference framework, input resolution, batch size, warm-up, number of repetitions and mean or median latency per image. Offline GWO search cost should be distinguished from runtime inference cost: the deployed classifier uses optimized CNN weights without rerunning GWO for each image.

Resource Requirements and Cross-Domain Transfer

Deployment feasibility also depends on model size, peak RAM/VRAM, throughput, acquisition-to-decision latency and the cost of retraining when imaging conditions change. Changes in reinforcement type, fiber orientation, microscope magnification, image contrast and defect morphology can cause distribution shift. A credible transferability study would evaluate images from different specimens, acquisition devices or composite families without mixing near-duplicate views across train/test splits; fine-tuning or domain adaptation may be required. These are proposed engineering validation steps, not experiments already completed in the present work. For implementation, trained CNN weights may be exported to an inspection workstation and connected to a microscopy or other image-acquisition pipeline. This is a prospective application; no industrial end-to-end demonstration or resource profile is documented in the supplied results. The present evidence supports promising image-level classification under the reported evaluation conditions, while timing and model portability remain open engineering questions.

In comparison to KerasTuner, the GWO setting is said to yield greater accuracy and superior precision/F1 performance alongside reduced validation loss. These observations reinforce the application of GWO as a hyperparameter-selection mechanism in the proposed technology. The concluding experimental storyline can be given as: the inspection data is pre-processed and partitioned; a CNN learns defect-sensitive spatial representations; GWO optimizes the CNN hyperparameters by minimizing validation loss; the best configuration α -wolf is retained as best model; and the optimized CNN is evaluated on unseen test data using accuracy, precision, recall, F1-score, AUC, loss, and inference time. This structure mirrors the experimental implementation provided and ensures consistent alignment of the proposed approach to the reported results.

A confusion matrix is used to identify class-specific error patterns, particularly between visually similar conditions such as porosity and debonding or fiber clustering and fiber misalignment. For multiclass classification, define TP, FP, and FN separately for each class using a one-versus-rest interpretation, calculate each class's precision, recall and F1, then specify whether reported aggregate metrics are micro-, macro-, or support-weighted averages. The F1-score calculated from separately averaged precision and recall is not necessarily identical to an average of class-wise F1-scores. ROC–AUC also requires explicit clarification of binary versus multiclass averaging. In the present manuscript, class-wise confusion-matrix values, averaging convention and per-class supports are not supplied and, therefore, cannot be reconstructed reliably.

The confusion matrix of CNN-PSO is shown in Figure 8(a) which establishes strong classification capability across all defect categories, with only a few misclassifications observed between Fiber Pull-Out, Voids, and Microcracks. Most samples are concentrated along the main diagonal, indicating high prediction accuracy and reliable defect recognition. The results confirm that PSO effectively optimizes CNN parameters and improves classification performance. On the other hand (Figure 8b), the confusion

matrix of CNN–GWO exhibits near-perfect classification, with almost all samples correctly assigned to their respective classes and only a single misclassification observed. The strong diagonal dominance reflects the superior discriminative ability of the GWO-optimized CNN, resulting in an overall accuracy of 99.86%. These findings demonstrate the effectiveness of GWO in enhancing feature learning and reducing classification errors compared to CNN-PSO.

True labels	Fiber-matrix debonding	158	0	0	0	0	True labels
	Fiber pull-out	0	152	0	3	1	
	Fiber fracture	0	1	138	0	0	
	Voids	1	0	2	137	1	
	Microcracks	0	2	0	1	137	
	Predicted labels						
	Fiber-matrix debonding	Fiber pull-out	Fiber fracture	Voids	Microcracks		

(a)

True labels	Fiber-matrix debonding	158	0	0	0	0
	Fiber pull-out	0	156	0	0	0
	Fiber fracture	0	0	139	0	0
	Voids	0	0	1	140	0
	Microcracks	0	0	0	0	140
		Fiber-matrix debonding	Fiber pull-out	Fiber fracture	Voids	Microcracks
		Predicted labels				

(b)

Figure 8. Confusion matrix for (a) CNN with PSO and (b) CNN with GWO.

CONCLUSION

- *CNN–GWO Framework:* A hybrid CNN–GWO framework was developed for automated defect classification in fiber-reinforced composite materials by combining CNN-based feature learning with GWO-based hyperparameter optimization.
- *Optimized CNN Configuration:* GWO identified an effective CNN configuration with 128 filters and kernel size 5 in the first convolutional layer, 192 filters in the second layer, and 256 dense units.
- *Model Performance:* The supplied report lists 99.23% training accuracy, 99.82% validation accuracy, 99.80% precision, 97.51% recall, 99.77% F1-score and 0.999 AUC. Recall and F1-score require checking against the original class-wise evaluation and averaging conventions before these metrics are treated as reconciled.
- *Comparison with KerasTuner:* CNN–GWO achieved better reported performance than KerasTuner, with validation loss reduced from 0.1119 to approximately 0.02, along with improvements in accuracy, precision, and F1-score.
- *Dataset and Evaluation:* The dataset was divided into 70% training, 15% validation, and 15% testing, with augmentation applied only to training data and the test set kept independent for final evaluation.
- *Practical Significance:* The results demonstrate the potential of CNN–GWO for automated and consistent composite-defect inspection, with scope for integration into future industrial inspection systems.
- *Future Scope:* Further work should document class-wise image counts, specimen-independent testing, repeated training runs, confusion matrices, inference hardware and memory usage, and transfer evaluation across materials and imaging conditions. These measurements are needed to establish statistical robustness and industrial deployment readiness.

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Approval

Not Applicable (based on the stated image-analysis study; subject to author confirmation).

Abbreviations and Symbols

Abbreviation	Full Form/Symbols with Units
AI	Artificial Intelligence.
ANN	Artificial Neural Network.
AUC	Area Under the Curve.
CNN	Convolutional Neural Network.
CNN–GWO	Convolutional Neural Network–Grey Wolf Optimization.
CT	Computed Tomography.
DCNN	Deep Convolutional Neural Network.
DL	Deep Learning.
F1	F1-Score.
FEM	Finite Element Method.
FID	Fréchet Inception Distance.
FRP	Fiber-Reinforced Polymer.
GWO	Grey Wolf Optimization.

IoU	Intersection over Union.
KNN	K-Nearest Neighbors.
mAP	Mean Average Precision.
ML	Machine Learning.
NDE	Non-Destructive Evaluation.
PSO	Particle Swarm Optimization.
RF	Random Forest.
SEM	Scanning Electron Microscopy.
SFRC	Steel Fiber-Reinforced Concrete.
SVR	Support Vector Regression.
TN	True Negative.
TP	True Positive.
FP	False Positive.
FN	False Negative.
YOLO	You Only Look Once.

REFERENCES

1. Medeghini F, Tiberti G, Guhathakurta J, Simon S, Plizzari GA, Mark P. Fiber orientation and orientation factors in steel fiber-reinforced concrete beams with hybrid fibers: A critical review. *Struct Concr.* 2025 Feb;26(1):481–500.
2. Fraile DH, Faccin E, Minelli F, Plizzari G, Muttoni A. Fibre orientation in SFRC slabs and consequences for punching shear and flexural resistance. *Eng Struct.* 2024 Mar 1;302:117364.
3. Mujalli MA, Dirar S, Mushtaha E, Hussien A, Maksoud A. Evaluation of the tensile characteristics and bond behaviour of steel fibre-reinforced concrete: An overview. *Fibers.* 2022 Dec 2;10(12):104.
4. Deng Y, Zhang Z, Shi C, Wu Z, Zhang C. Steel fiber–matrix interfacial bond in ultra-high performance concrete: A review. *Engineering.* 2023 Mar 1;22:215–232.
5. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2014. p. 580–587.
6. Girshick R. Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2015. p. 1440–1448.
7. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst.* 2015;28.
8. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2017. p. 2961–2969.
9. Lu H, Cantero-Chinchilla S, Yang X, Gryllias K, Chronopoulos D. Deep learning uncertainty quantification for ultrasonic damage identification in composite structures. *Compos Struct.* 2024 Jun 15;338:118087.
10. Huang H, Liu Y, Zhan J, Jia B, Dong Z, Zhao M, et al. Recent progress of AI in CFRP: From design and manufacturing to structural health monitoring. *Thin-Walled Struct.* 2026 Apr 3:114936.
11. Cheng J, Tan W, Yuan Y, Zhao Z, Cheng Y. Research on defect detection method for composite materials based on deep learning networks. *Appl Sci.* 2024 May 14;14(10):4161.
12. Han J, Wang Z, Liu W, Song B, Wang C, Liu X. Data-centric deep learning for defect detection in fiber composite materials: A review of algorithms, challenges, and emerging trends. *Knowl Based Syst.* 2026 Aug 4:116784.
13. Vasilache MM, Tayong RB, Velisavljevic V. Comprehensive review on the use of machine learning techniques applied to the ultrasound data for the characterisation of porosity across carbon fibre reinforced polymer layers. *Appl Compos Mater.* 2025 Aug;32(4):1315–1339.

14. Beldar P. Optimized convolutional neural networks for defect classification in fiber-reinforced composites. *Sigma*. 2026 Jun;44(3):2207–2218.
15. Peng S, Addepalli S, Farsi M. Machine learning in thermography non-destructive testing: A systematic review. *Appl Sci*. 2025 Sep 1;15(17):9624.
16. Yazdani MH, Azad MM, Khalid S, Kim HS. A hybrid deep transfer learning framework for delamination identification in composite laminates. *Sensors*. 2025 Jan 30;25(3):826.
17. Azad MM, Shah AU, Prabhakar MN, Kim HS. Deep learning-based microscopic damage assessment of fiber-reinforced polymer composites. *Materials*. 2024 Oct 29;17(21):5265.
18. Silenzi A, Castorani V, Tomassini S, Falcionelli N, Contardo P, Bonci A, et al. Quality control of carbon look components via surface defect classification with deep neural networks. *Sensors*. 2023 Sep 1;23(17):7607.
19. Daghigh V, Daghigh H, Lacy TE Jr, Naraghi M. Review of machine learning applications for defect detection in composite materials. *Mach Learn Appl*. 2024 Dec 1;18:100600.
20. Liu C, Chen Y, Xu X. Structural digital Twin for damage detection of CFRP composites using meta transfer Learning-based approach. *Expert Syst Appl*. 2025 Feb 1;261:125458.
21. Deng K, Liu H, Cao J, Yang L, Du W, Xu Y, et al. Attention mechanism enhanced spatiotemporal-based deep learning approach for classifying barely visible impact damages in CFRP materials. *Compos Struct*. 2024 Jun 1;337:118030.
22. Liu J, Long X, Jiang C, Liao W. Multi-feature vision transformer for automatic defect detection and quantification in composites using thermography. *NDT E Int*. 2024 Apr 1;143:103033.
23. Li Z, Li Y, Lu J, Zhu H, Zheng Y, Xue J, et al. Baseline-free assisted lamb wave-based damage detection in CFRP composites using graph convolutional networks and Transformer models. *Measurement*. 2025 Jan 1;242:116159.
24. Wen R, Tao C, Ji H, Qiu J. Classification, localization and quantization of eddy current detection defects in CFRP based on EDC-YOLO. *Sensors*. 2024 Oct 21;24(20):6753.
25. Wang G, Zhang Z, Yin W, Zhou G, He C, Chen H, et al. Real-Time defect detection in CFRP with enhanced YOLO via differencing and attention for line laser thermography. *Infrared Phys Technol*. 2026 May 25:106673.
26. Liu W, Liu J, Zhu J, Han K, Zhuo L, Liu Z. CM-YOLO: An improved YOLOv8 for defect detection and classification of CFRP composites using infrared thermography. *Nondestruct Test Eval*. 2026 Aug 19:1–21.
27. Fan L, Li Z, Dong F, Gao M, Sun Y, Tang Z, et al. Infrared image defect detection for carbon fiber reinforced polymer composites via infrared preprocessing and improved YOLOv8. *Infrared Phys Technol*. 2026 Sep 1:106886.
28. Kandiyil AP, Atkinson GA, Damghani M. Localisation of defects on carbon fibre surfaces using deep learning. *NDT E Int*. 2025 Aug 7:103498.
29. Liang C, Hu Q, Gao X, Wu J, Mei H, Qi F, et al. Accurate segmentation and quantitative evaluation of Cf/SiC fiber fracture defects using an enhanced deep learning method. *Mater Charact*. 2025 Feb 1;220:114712.
30. Fan W. Composite material surface microscopic defect detection and classification combining diffusion models and zero-shot learning. *Sci Rep*. 2025 Dec 19;16(1):388.
31. Al-Shamasneh AA, Mahmoodzadeh A, Karim FK, Saidani T, Alghamdi A, Alnahas J, et al. Application of machine learning techniques to predict the compressive strength of steel fiber reinforced concrete. *Sci Rep*. 2025 Aug 21;15(1):30674.
32. Khan A, Ahmad MS, Siddiqui AA. Machine learning methods for prediction of compressive strength of steel fibre reinforced concrete at elevated temperatures: A comparative analysis. *Discov Civ Eng*. 2026 Feb 2;3(1):30.
33. Beldar P, Chandramore R, Jagtap S, Bendale D. Enhanced convolutional neural networks for defect detection in fiber-reinforced composites: A hyperparameter optimization approach. *J Eng Appl Sci*. 2025 Dec;72(1):125.
34. Huang M, Xu Z, Cai C, Hu C, Wei X, Yin W, et al. Multi-structure transfer damage detection for composite materials based on singular value decomposition and 2D-CNN network. *Structures*. 2025 Apr 1;74:108656.

35. Jiang L, Xiao S, Wu J, Liao Z, Zhu G, Zhang B. CT and image post-processing for fiber composites: Defect analysis, deep learning, digital volume correlation, and FE simulation—a review. *Mater Des*. 2026 Feb 22;115698.
36. Azad MM, Kumar P, Kim HS. Delamination detection in CFRP laminates using deep transfer learning with limited experimental data. *J Mater Res Technol*. 2024 Mar 1;29:3024–3035.
37. Kojima Y, Hirayama K, Endo K, Harada Y, Muramatsu M. Transfer-learning-aided defect prediction in simply shaped CFRP specimens based on stress distribution obtained from finite element analysis and infrared stress measurement. *Compos Part B Eng*. 2025 Feb 15;291:111958.
38. Buzpinar MA. Intelligent defect detection in composite materials: The role of deep learning and machine learning fusion. *Nondestruct Test Eval*. 2026 Aug 9:1–38.
39. Abid H, Hassan D, Nauman S. Comparative study on the implementation of Custom Convolutional Neural Networks (CNNs), MobileNet V3 and ResNet-50 Architecture for Defect Detection in Carbon Fiber Reinforced Plastics (CFRPs). *Next Res*. 2026 Aug 1:102260.
40. Li M, Leonetti D, Zappalá D, Snijder HB. Infrared-based damage detection in thick composites via transfer learning on simulated and experimental data. *Compos Part B Eng*. 2025 Oct 20;113112.
41. Ding Y, Zhou SX, Yuan HQ, Pan Y, Dong JL, Wang ZP, et al. Crack identification method of steel fiber reinforced concrete based on deep learning: A comparative study and shared crack database. *Adv Mater Sci Eng*. 2021;2021(1):9934250.
42. Jung C, Redenbach C, Schladitz K. VoroCrack3d: An annotated semi-synthetic 3D image data set of cracked concrete. *Data Brief*. 2024 Jun 1;54:110474.
43. Hampe M. Micro-CT dataset of a steel fibre reinforced concrete sample. [Dataset]. Available from: <https://doi.org/10.5281/zenodo.10785684>.
44. Souza JP, Silva AC, Congro M, Roehl D, Paiva AC, Pereira S, et al. Automated crack detection in micro-CT scanning for fiber-reinforced concrete using super-resolution and deep learning. *Electronics*. 2025 Oct 28;14(21):4208.
45. Mathiyazhagan GR, Paramasivam SK. Effect on AI-driven Ant Lion Optimization framework using fibre reinforced concrete with dual-stage building crack for structural applications. *Matéria (Rio J)*. 2026 Feb 16;31.
46. Kawkabi KW, Nziko ASNT, Ngangura PM, Long X. Morphology-aware CNN with statistical descriptor regression and conditional image synthesis for predicting crack patterns in steel fiber-reinforced concrete. *Comput Mater Contin*. 2026. doi: [10.32604/cmc.2026.086951](https://doi.org/10.32604/cmc.2026.086951).
47. Wang P, Yun X, Hong S, Fang G, Dong B, Zhang Y, et al. Identification and characteristic analysis of internal corrosion-expansion cracks in reinforced concrete based on deep learning. *J Build Eng*. 2026 Sep 5:117287.
48. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Adv Eng Softw*. 2014 Mar 1;69:46–61.