

Enhanced Deep Learning Techniques for Defect Detection in Steel-Fiber-Reinforced Composites: A Comparative Study of KerasTuner, CNN–PSO, and CNN–GWO

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

Abstract

The identification of microstructural defects in steel-fibre-reinforced composites is useful for the material integrity, mechanical reliability and longevity. Classic material inspection techniques face difficulties in recognizing the complex and visually similar defects because of heterogeneous material morphology, changing imaging conditions, and limits of manual feature extraction. To tackle those problems, this study presents a comparative deep-learning framework that integrates Convolutional Neural Networks (CNNs) with three hyperparameter optimization techniques namely KerasTuner, Particle Swarm Optimization (PSO), and Grey Wolf Optimization (GWO). The framework focuses on automated recognition of composite defects through image pre-processing, hierarchical extraction of relevant features, and optimization-supported configuration. Three methods are evaluated in a common comparative framework to study learning behaviour, convergence characteristic, and classification performance. The experiment assesses both training and validation accuracy as well as their associated losses over 25 epochs. The graphical comparison shows that CNN–GWO achieved the highest training and validation accuracy followed by CNN–PSO and CNN–KerasTuner. Values derived from the given performance curves suggest that CNN–GWO achieves a validation accuracy of 99.5%, followed by CNN–PSO at 97.0% and CNN–KerasTuner at 94.5%. Moreover, CNN–GWO achieves high initial loss reduction and demonstrates good convergence properties, indicating its potential effectiveness in choosing discriminative CNN configurations for composite defect recognition. While these results demonstrate the advantages of CNN–GWO in the reported experimental setting, the statistical robustness and generalization of the results need to be evaluated with independent test sets and repeated experimental trials. The swarm-intelligence-assisted deep learning framework has the potential applications in automated material inspection. Moreover, it offers a comparative background for developing reliable defect classification systems for steel-fibre-reinforced composites. Steel fiber reinforced composites (SFRCs) are widely used in the construction industry to improve strength and durability. However, the detection of defect in SFRCs is performed manually which is time consuming and does not provide accurate results. There are various automated deep learning approaches which can detect the defect in SFRCs quickly and provide reliable result. This study aims to carry out the hyperparameter optimization of convolutional neural network (CNN) using KerasTuner and with the help of advanced metaheuristic algorithms such as particle swarm optimization (PSO) and grey wolf optimization (GWO). The authors will develop a model which will provide better defect detection accuracy than the existing model and will make the process more automated. Results will be presented in the latter sections to validate the effectiveness of the model.

*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, Department of Mechanical Engineering, KNIT, Sultanpur, Uttar Pradesh, India

Received Date: October 15, 2026

Accepted Date: October 21, 2026

Published Date: October 23, 2026

Citation: Dibya Tripathi, Kheelraj Pandey, Akhilesh Kumar Chauhan, Ashwani Sharma. Enhanced Deep Learning Techniques for Defect Detection in Steel-Fiber-Reinforced Composites: A Comparative Study of KerasTuner, CNN–PSO, and CNN–GWO. Journal of Polymer & Composites. 2026; 14(5): 377–403p.

Keywords: Steel-fiber-reinforced composites, defect detection, convolutional neural network, KerasTuner, particle swarm optimization, grey wolf optimization, hyperparameter optimization, deep learning

INTRODUCTION

The increase in demand for engineering materials that have better mechanical strength, durability, fracture resistance, and structural reliability has speeded up the development and usage of fibre-reinforced composites through civil, mechanical, automotive, aerospace, and industrial engineering. Steel-fiber-reinforced composites are among these materials that have been subjected to considerable attention because the addition of discrete steel fibers can improve the crack resistance, energy absorption, load transfer after cracking and toughness of the material.

The performance of these materials, however, largely depends on how the reinforcing fibers are distributed, oriented, geometrically shaped and bonded to each other. Manufacturing inconsistencies, environmental exposure, and mechanical loading can introduce microstructural imperfections that may slowly endanger structures and shorten their life. The identification and classification of defects have thus become essential requirements for materials characterization and quality assessment.

Zhang et al. [1] indicated that the steel-fiber content and distribution affect the fracture behaviour of steel-fiber-reinforced concrete. These characteristics exhibit correlations with mechanical performances. The results of the inspection reinforce the importance of reliable inspections that can detect defects before they cause serious damage to the material of these bridge components.

The defect inspection usually takes place through visual inspection, optical microscopy, scanning electron microscopy, ultrasonic inspection, infrared thermography, and X-ray computed tomography. Whilst these methods provide informative data on morphology/materials and the inner damage of the part, they can be operator experience dependent, specimen preparation dependent, imaging condition dependent, equipment requirement dependent, and complexity of interpretation after inspection is data volume dependent.

The consistently growing number of defects has raised question on the accuracy of human inspectors but lately it has raised significant interest in their replacement to lessen dependence on subjective interpretation and improve consistency of a defect recognition through automated image analysis and associated deep learning. Specifically, they can automatically learn hierarchical representations directly from inspection images, and can differentiate complex visual features related to cracks, fiber discontinuities, voids, and interfacial damage.

The performance of a CNN classification depends on the quality of the existing images and the configuration of the network. To this end an optimization-assisted deep-learning frameworks was developed that select effective architectures and training hyperparameters. The study examines the efficacy of KerasTuner, PSO, and GWO optimization strategies to improve the classification of defects in steel fiber reinforced composite (SFRC) using CNN model.

Background and Significance of Steel-Fiber-Reinforced Composite Materials

Steel-fiber-reinforced composites are those engineered materials into which discreet steel fibers are added to the surrounding matrix to improve mechanical performances and damage propagation behaviour. Steel-fiber-reinforced concrete (SFRC) contains steel fibers that help bridge cracks and transfer stress across cracks, allowing for improved post-cracking behaviour and fracture toughness ductility energy-dissipating capacity.

The various reinforcing mechanisms depend on several interacting material properties including fiber geometry, aspect ratio, volume fraction, spatial orientation, distribution uniformity and fiber–matrix bonding

strength. According to Zhang et al. [1], the fracture properties of SFRC have been reviewed and it is found that steel-fiber content and distribution play a vital role on the fracture resistance. Based on their findings, the study of fracture mechanisms will greatly enhance the evaluation of structure performance.

Steel-fiber-reinforced composites present several advantages; however, defects arising during manufacturing and service can threaten their mechanical reliability. During fabrication, incomplete mixing, insufficient compaction, nonuniform fiber distribution and variations in the matrix composition gives rise to local areas presenting with voids, fiber agglomeration and poor interfacial bonding.

Under mechanical loading, the cracking due to different damage mechanisms is caused due to the initiation of crack, interfacial debonding, crack growth, pull out of fiber, and fracture of fiber. Such mechanisms impact to stress transfer in the material and may reduce effectiveness of reinforcing fiber in resistance to progressive structural damage. Identifying the variations is, therefore, necessary to understand the degradation of materials and improve the quality-control systems. Nevertheless, the complex and heterogeneous characteristics of steel-fiber-reinforced composites do not favour precise recognition of defects, mainly when damage descriptors are cropped or morphologically similar.

The main defect categories will be fiber–matrix debonding, fiber pull-out, fiber fracture, void formation, and microcracking as per the confirmation of the available pivot dataset labels. The fiber-matrix debonding is the deterioration/separation at the place of reinforcing fibers and the matrix which decreases the efficient transfer of stress. Pull-out of the fiber occurs when the embedded fibers become partially or fully dislodged from the matrix while fracture of fiber refers to the discontinuity of the reinforcing element itself.

Void formation relates to localized regions of materials which are free from materials. Microcracking refers to fine structural discontinuities which may develop due to shrinkage, thermal effects or mechanical stresses applied. These defects have distinct physical properties but can look similar under microscopic examination. Intensity patterns resembling those produced by fiber pull-out cavities and voids can also be produced by fine cracks and interfacial debonding. As a result, the defect classification must rely on image-analysis techniques capable of learning subtle differences in texture, shape, boundary continuity and local spatial organization.

Interfacial separation, fiber displacement, fiber rupture, material voids and fine matrix cracks are shown in Figure 1. The overlapping textures and defect boundaries create challenges when performing image-level classification. The diagram is for illustrative purposes only; the final classes in the datasets used will be the labels from the verified dataset.

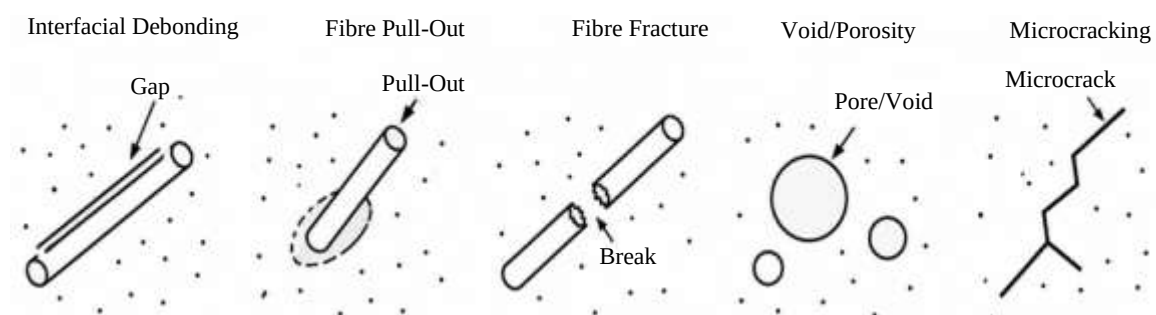


Figure 1. Schematic representation of five candidate defect mechanisms in fiber-reinforced composites.

The increasing accessibility of digital imaging methods has opened doors for the automation of assessing such flaws. The information of surface morphology and fracture characterisation can be achieved from optical microscopy and scanning electron microscopy, but internal structure arrangement and localized defect could be obtained from X-ray micro-computed tomography.

Nevertheless, conventional approaches to feature extraction can be impacted by fluctuations in lighting, spatial resolution, image contrast, specimen preparation, and acquisition methodologies. These challenges emphasize the growing significance of computational techniques that learn robust representations from intricate material images. Recent efforts have, therefore, focused on automated composite inspection based on deep-learning architectures, especially convolutional neural networks (CNNs), which allow for hierarchical extraction of defect-sensitive visual features. In other words, it is not necessary to specify all relevant descriptors manually.

Challenges in Conventional Defect Detection and the Emergence of Deep Learning

Detection of defects in steel-fiber-reinforced composites is mostly done using destructive tests, manual visual inspection and non-destructive testing. Through visual inspection, preliminary examination, and microscopic examination, it is possible to observe the surface irregularities and fracture-related features of cracks. The dependability of these results, however, will depend on the choice of operators, quality of images, visibility of the defects and interpretation of the intricate microstructural features.

With ultrasonic inspection, it is possible to detect the presence of a change in the continuity of a material through the interaction of acoustic waves. Likewise, infrared thermography provides information regarding the distribution of temperature in the presence of certain internal irregularities. X-ray tomography offers detailed representations of internal material structures and provides a means of access to spatial fiber distribution and damage morphology.

Although diagnostic tools, these inspection techniques often necessitate sophisticated equipment, controlled acquisition conditions, and extensive data interpretation. Additionally, traditional image processing methods – such as thresholding, edge detection, morphological operations, and handcrafted texture descriptors – may have limited adaptability when the defect appearance varies across samples and imaging conditions.

A learning framework that directly learns discriminative representations from image data is through deep learning techniques. Conventional machine-learning methods often necessitate the manual selection of descriptors, in contrast, CNN architectures integrate feature extraction and classification into a single computational architecture. Convolutional layers pick up local spatial patterns, such as edge, fiber boundary, crack trajectory, and texture discontinuities, while further layers combine these elementary features into more abstract ones.

These hierarchical feature-learning capabilities are relevant to composite inspection since local defects are irregular, low contrast, and involve complex material backgrounds. Ding et al. [2] explored deep-learning-based crack identification in steel-fiber-reinforced concrete using YOLOv4, Mask R-CNN, and an enhanced Mask R-CNN framework. The study indicates how useful automatic inspection and architecture selection for crack identification can be, but it is a study that considers a different experimental task than the present comparison of three-way optimization.

Deep learning-based inspection offers many advantages; however, these techniques suffer from various constraints. The presence of limited annotated datasets, differences in defect frequencies, class imbalance, image noise, and variations in microscopic acquisition conditions may yield models with low generalization. In particular, visually analogous defects may cause category confusion, while complex network architectures can ramp up computational costs and risk overfitting.

CNN performance also depends on several architectural parameters like the number of convolutional filters, kernel dimensions, pooling arrangements, arrangement of the dense-layer, and regulation techniques. Increasing complexity of a network does not ensure better defect identification. To make CNN a powerful model, one should select a suitable choice of representational ability, optimization

stability, computational efficiency, and generalization performance. Yang et al. [3] reviewed hyperparameter optimization techniques for machine learning, comparing their performance, applications, available tools, and challenges in selecting optimal configurations. Figure 2 depicts the image acquisition and controlled pre-processing follow by defect classification and CNN feature learning. The aim is to demonstrate why hyperparameters should be systematically selected, not to take credit for a validated inspection workflow or measured gain.

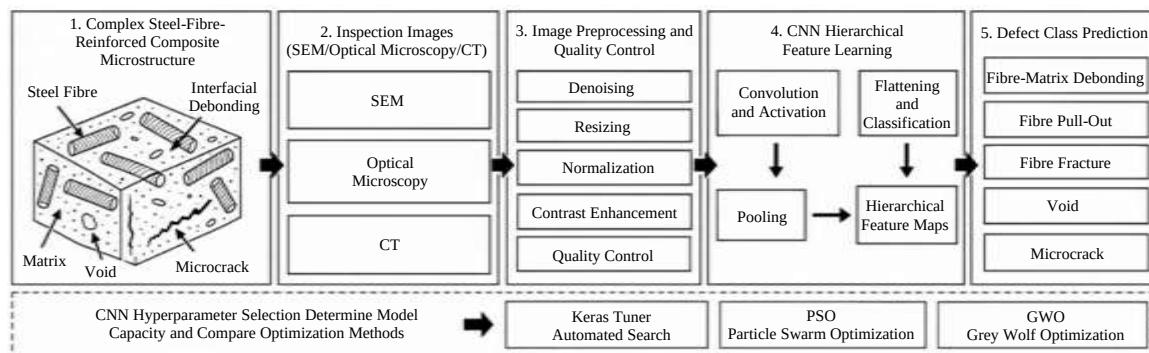


Figure 2. A conceptual change from composite inspection images to the classification of defects using CNN illustration as per deep learning inspection principles [3] without experimental outputs.

As a result of these limitations, they have encouraged the investigators to investigate systematic hyperparameter selection instead of relying only on manually configured CNN architectures. Yang and Shami stressed hyperparameter optimization contribute much to machine-learning performance. Further, the different optimization methods have some advantages and disadvantages based on the application used.

According to their findings, one should consider the characteristics of the search space, computational requirements, and interrelationships between model parameters when selecting an appropriate optimization strategy. Thus, automated hyperparameter optimization integrated with CNN-based composite inspection offers a useful avenue for enhancing classification performance while diminishing the arbitrary choice of architecture.

Motivation for Optimization-Assisted CNN-Based Defect Classification

The effective use of CNN architectures for recognizing material defects relies not just on their ability to learn hierarchical representations of images but also on the configuration of architectural and training parameters. A simple modification in the number of convolutional filters, kernel sizes, learning rates, batch sizes or regularization parameters can have vastly different convergence and generalization behaviour in a CNN.

Adjusting these manually can be expensive and time-consuming since as the dimensionality increases, the number of configurations increases rapidly. Moreover, optimizing individual parameters independently may ignore the interaction of architectural and training parameters. These difficulties lead to systematic search methods which assess configurations, finding effective combinations of components that allow achieving the desired performance.

KerasTuner [4] creates an automated framework for searching CNN hyperparameter configurations through various strategies such as Random Search, Bayesian Optimization, and Hyperband. By being part of the Keras family, it makes building, training, and evaluating models easy using your own search space and objectives.

The post Hyperband and Successive Halving appeared first on Towards Data Science. Nevertheless, its effectiveness is contingent on the search algorithm utilized, computational budget of the experiment and hyperparameter boundaries.

In contrast, population-based optimization techniques – like PSO and GWO – utilize iterative search methods where several candidate solutions may help identify more promising network configurations. Kennedy and Eberhart [5] provided the personal and collective experience of the particles to guide movement in the space of solution using PSO. Mirjalili et al. [6] proposed GWO algorithm which were hierarchy in nature. The GWO uses the alpha, beta and delta candidate solutions to update the population. The diverse search mechanisms uniquely position us to assess the relative effectiveness of alternative CNN hyperparameter optimization strategies.

Recent research has shown the optimization-assisted CNN models can classify fiber-reinforced composite defects. Beldar [7] used KerasTuner and Particle Swarm Optimization for tuning convolutional neural networks to identify microscopic defect classes in fiber-reinforced composite materials. The work in reference is a related methodological reference for applying optimization to assist in image classification, but the experiments are not the results of this work, and they did not perform a direct three-way evaluation with Grey Wolf Optimization. Accordingly, it ascertained that a more in-depth investigation into the comparison of KerasTuner, CNN-PSO, and CNN-GWO, based on consistent data partitions, search objective, and evaluation criteria, is essential.

As shown in Figure 3, GWO was selected because of its underlying adaptive population-search mechanism, which may be a good fit for searching nonlinear hyperparameter spaces. The GWO algorithm updates the position of candidates with coordinated updates based on the 3 leading candidates as opposed to using velocity and best-known solutions like PSO. This method alters the balance between population exploration and local refinement in effective convolutional neural networks (CNNs) search.

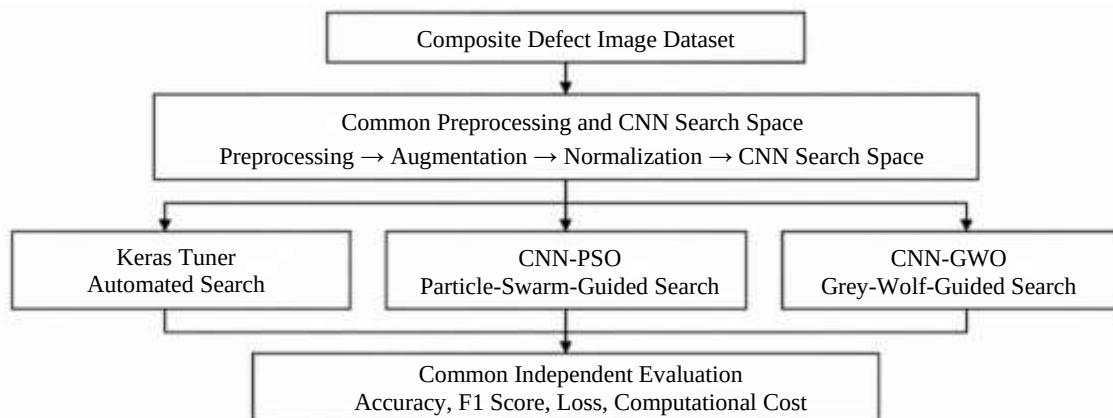


Figure 3. Conceptual comparison of the three CNN hyperparameter-optimization pathways proposed for the present study.

However, the algo which is 34tter cannot be judged based on their search mechanism or results from other datasets which are not related to each other. A controlled experimental comparison needs to be done to compare the effectiveness of these optimization strategies for defect classification in SFRP composites. The present research is motivated by this requirement and studies whether CNN–GWO can enable improved functioning over CNN–PSO and KerasTuner with respect to computational efficiency and classification reliability.

LITERATURE REVIEW

The increasing deployment of fiber-reinforced composite materials in structural and industrial engineering has raised demand for effective, trustworthy and automatic defect detection techniques. Particular steel-fiber-reinforced composites display complex microstructure defects. Any defect in its context – such as fiber–matrix debonding, fiber pull-out, fiber fracture, void formation, and microcracking – can affect the mechanical performance and structural integrity as well as longevity.

Traditional inspection techniques – like microscopy, UTC, infrared thermography, and X-ray CT – provide valuable information on material degradation; however, they may be restricted in effectiveness due to complexity of inspection, imaging conditions, operator dependency and interpretation of large inspection data.

Advancements in artificial intelligence, especially convolutional neural networks (CNNs), are creating new avenues towards defect recognition and classification of complex microstructures via automated feature extraction. However, the classification performance of the CNN-based system will be substantially dependent on the choice of architectural and training hyperparameters or factors such as a convolutional filter size, kernel sizes, learning rate, batch sizes, depth of the network, and regularisation factors.

Hyperparameter tuning is an important area of research that improves convergence properties, generalization ability, and computational cost. Methods – like KerasTuner, PSO, and GWO – examine sophisticated hyperparameter search space and help to come up with an efficient CNN configuration.

Existing research has studied the classification of composite materials through deep learning and optimization techniques. However, owing to the use of different datasets, and evaluation protocols and optimization techniques, which make direct performance comparison impossible.

This literature review systematically investigates the evolution of composite defect inspection, traditional and deep-learning-based detection methods, hyper-parameter optimization of CNN, and applicability of KerasTuner, CNN-PSO, and CNN-GWO. Research priorities are defined identifying methodological weaknesses and research opportunities to motivate a unified comparative framework for defect classification in steel-fiber-reinforced composites.

Defect Characteristics and Inspection Challenges in Steel-Fiber-Reinforced Composites

Steel-fiber-reinforced composites have become very popular in the cement and concrete industries with high tensile resistance and many mechanical engineering and structural engineering applications owing to their enhanced fracture toughness, ductility, energy absorption capacity, and improved resistance to crack propagation.

Steel fiber reinforced concrete (SFRC) includes steel fibers either at random or deliberate orientations in the cementitious medium and helps in the stress transfer across the developing cracks and also improves the mechanical behaviour under different loading conditions. Nonetheless, the usefulness of those reinforcing mechanisms depends strongly on fiber orientation, spatial distribution and geometrical characteristics, interfacial bonding and uniformity of the surrounding matrix.

Flaws in the preparation of materials, mixing, casting, compacting, curing, and subsequent mechanical loading might introduce local defects which may adversely affect the structural performance and service life of the composite. Thus, reliable identification and characterization of such defects are important for evaluating the quality of materials and structures, and their long-term properties [8, 9]. Figure 4 presents some possible microstructural defect morphologies in steel-fiber-reinforced composites which comprise, for example, interfacial debonding, pull-out, fiber fracture, void and microcracking, conceptual illustration author-developed based on [8, 9].

A further challenge is the presence of large, balanced, accurately annotated datasets of the various defects of materials. Getting an experimental image by microscopy is sometimes only possible with the assistance of specific instrumentation, in the labs, and after a considerable expenditure of time to process.

As a result, for certain defect categories, the available datasets may have limited observations, which would raise the class imbalance issue causing biased predictions. In addition, images harvested from varied regions of the same specimen exhibit strong correlations, which may generate over-optimistic classification results when closely related samples are distributed in the training and testing subsets.

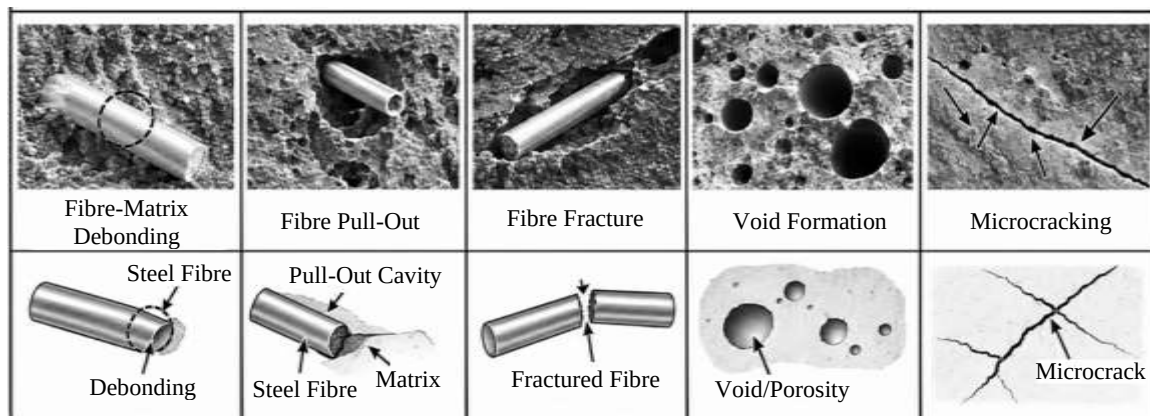


Figure 4. Illustrative microstructural defect morphologies in steel-fiber-reinforced composites: Interfacial debonding, pull-out, fiber fracture, void formation, and microcracking.

For the creation of robust automated classification frameworks, the dataset partitioning, defect-preserving, and image augmentation or independent performance assessment are very important. These considerations have led to the adoption of deep-learning approaches, which extract discriminative microstructural representations directly from inspection images and enable accurate identification of intricate defect patterns [10].

Conventional and Non-Destructive Techniques for Composite Defect Detection

Testing for defects in fiber-reinforced composite materials is generally carried out through visual inspection, destructive material characterization, and non-destructive testing (NDT). The conventional visual inspection enables the identification of surface discontinuities, exposed fibers, cracking and visible structural failure, but this will depend on the experience of the inspector, surface access, lighting and the defect size.

Destructive methods, such as mechanical testing and microscopic examination of fractured specimens, provide valuable insight into fiber–matrix interaction, fracture mechanism, and material degradation. Still, such methods may need extensive sample prep and may change or destroy the specimen being examined. As a result, several alternatives have been identified to evaluate the structural integrity of components without significantly impairing their functioning [11, 12].

Figure 5 indicates that optical and electron microscopy mainly offer information about accessible surface morphology, while ultrasonic, infrared and X-ray-based approaches provide supplementary information about internal or subsurface morphology. In terms of penetration, resolution, preparing the items and sensitivity to material’s heterogeneity, the techniques differ. No single method will independently resolve every defect type.

Visual Examination	Optical Microscopy	Ultrasonic Testing	Infrared Thermography	X-ray Computed Tomography (CT)
<ul style="list-style-type: none"> • Surface defects • Simple and low cost • Operator dependent • Limited to surface information 	<ul style="list-style-type: none"> • High-resolution surface/microstructure • Useful for fine defects • Requires specimen preparation • Limited field of view 	<ul style="list-style-type: none"> • Internal discontinuities • Non-destructive • Real-time inspection • Sensitive to material heterogeneity 	<ul style="list-style-type: none"> • Thermal anomalies • Large-area inspection • Fast and non-contact • Influenced by heating conditions and depth 	<ul style="list-style-type: none"> • 3D internal structure • Porosity and cracks • High spatial resolution • High acquisition and processing cost

Increasing Information Depth and Inspection Complexity

Figure 5. Pictorial comparison of conventional inspection modalities used to characterize fiber-reinforced composites.

The conventional inspection and handcrafted image-analysis approaches have certain limitations, which have led to the advent of automated machine learning and deep learning in recent years. NDT after collage of images is a good example of the last one.

Instead of relying on hand-crafted feature descriptors, deep-learning models can automatically develop layered representations of material morphology and learn complicated relationships between defect classes. Due to their ability to extract patterns, edges, textures, and higher-level structure from images, convolutional neural networks show potential for image-based material inspection.

The performance of a Convolutional neural network is still sensitive to network topology, optimisation parameters, experimental setup and quality of the data. Through its reliance on the systematic element, the limitation of traditional inspection tools directly motivates the need for their systematic deep-learning optimization, which further motivates the use of KerasTuner, Particle Swarm Optimization, and Grey Wolf Optimization in a single composite-defect classification. Table 1 presents a conventional inspection technique and their relevance to composite defect assessment.

Table 1. Conventional inspection techniques and their relevance to composite defect assessment.

Technique	Typical information	Practical limitation
Visual/optical microscopy	Surface cracks, exposed fibers	Cannot reliably visualize deeper internal defects.
SEM	Fine surface morphology, pull-out and fracture features	Sample preparation; restricted field of view.
Ultrasonic inspection	Internal discontinuities from wave response	Scattering in heterogeneous materials.
Infrared thermography	Thermal signatures of certain subsurface anomalies	Sensitivity to heating conditions and depth.
X-ray micro-CT	3D fiber geometry, porosity and internal cracks	High acquisition and processing requirements.

Figure 6 shows comparative hyperparameter optimization strategies for CNN-based composite-defect classification without any experimental outcome.

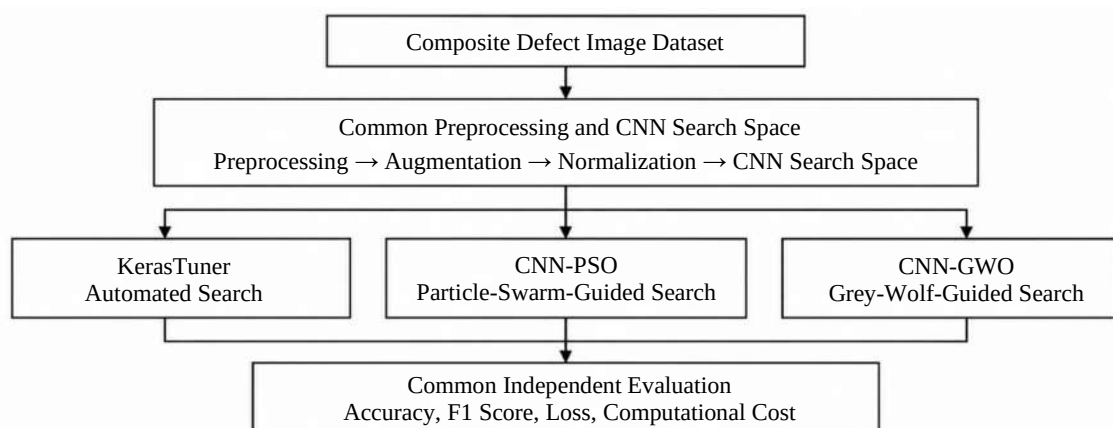


Figure 6. Comparative hyperparameter optimization strategies for CNN-based composite-defect classification.

The illustration displays how the common input dataset, pre-processing strategy and candidate CNN search space were established in the implementations. In contrast to independent hyperparameter attempts, the PSO paradigm adopts both a best particle solution and a best swarm solution. Meanwhile, GW is based on three wolves: the alpha, beta and delta leader.

An impartial evaluation by an independent team is necessary to establish that any variations in model performance are due to the optimizer changes.

The growing need for dependable structural inspection and manufacturing quality assurance has spurred the gradual adoption of machine learning (ML) and deep learning (DL) techniques in automated material defect recognition systems. Traditional image processing techniques rely mostly on hand-crafted descriptors, threshold-based segmentation, edge detection and statistical texture analysis for identifying any surface and microstructural irregularities.

Nonetheless, such methods often show limited robustness on complex composite materials having disordered fiber patterns, non-uniform lighting, cluttered backgrounds, and faint defects. The machine-learning algorithms – such as Support Vector Machines (SVM), Random Forests (RF), k-Nearest Neighbours (KNN), Decision Trees, and Artificial Neural Networks (ANN) – can offer an alternative framework for discovering links between extracted material features and defect categories.

Liu et al. [13] reviewed steel-fiber-reinforced concrete (SFRC) mechanical properties and durability, synthesizing material-performance findings without image-based CNN experiments or optimization. Daghigh et al. [14] reviewed the applications of machine learning in composite material defect detection and damage assessment study. It shows the potential of data-driven models for characterization of structural defects and response evaluation of material. Still, traditional machine-learning techniques tend to rely on the quality and representativeness of manually chosen input features, thus limiting their ability to capture complex morphological variability across defect categories.

CNN-Based Feature Extraction and Defect Classification

CNNs are an emerging deep-learning architecture that allows images to be automatically interpreted owing to their ability to learn representations hierarchically and considering the spatial structure of the images directly from the images themselves. A traditional CNN includes convolutional layers, nonlinear activation functions, pooling, and fully connected classification layers.

At the time of feature extraction, the learnable convolutional kernels can identify localized patterns present in images. Some of them can be edges, corners, changes in intensities, textures, structural discontinuities, etc. When information passes through several layers, these simple representations are combined to represent visually more complex structures.

In fiber-reinforced composite inspection, such representations are particularly useful in distinguishing defect specific morphological characteristics like irregular crack routes, fiber discontinuities, pull-out pits, interfacial separation, and localized voids. As a result, the ability of CNNs to learn spatial features give them an important advantage over traditional machine-learning methods with manually defined descriptors. LeCun et al. [15] proposed the original introduction of hierarchical representation learning.

Hyperparameter Optimization in Deep Learning Architectures

The choice of suitable hyperparameters that determine the networks' structure, learning, regularization and computational complexity hugely influences the performance of deep-learning architectures. In contrast to trainable parameters which are automatically updated during backpropagation, hyperparameters are typically set prior to model training or are selected via some external optimization procedure.

In Convolutional Neural Networks (CNNs), the architectural hyperparameters include: the number of convolutional layers, filter dimensions, kernel sizes, pooling, fully-connected neurons, etc. Training-related hyperparameters include learning rate, batch size, dropout probability, optimizers parameters, training time, etc. When these parameters interact with one another, they create a complex search space with ineffective configurations that either result in the insufficient extraction of features, unstable convergence, excessive use of computational resources and overfitting. Hyperparameter optimization

can be classified into theoretical and practical ones. Specifically, the model performance is not only dependent on the optimization algorithm but also the nature of the underlying search space. In addition, their analysis highlighted the significance of selecting appropriate optimization techniques, depending on the computational requirements, interaction between parameters and model complexity.

Manual experimentation, grid search, and random search are standard methods for hyperparameter selection. Tuning parameters manually is very much dependent on the experience of the researcher and becomes very inefficient when network parameters are to be tuned simultaneously. The use of grid search evaluates a predetermined set of values for each possible hyperparameter combination. This guarantees that every combination is explored to be able to derive the best result from a discrete space. However, as the number of parameters increase, this becomes a more expensive computational exercise. The random search may lead to a more efficient exploration of the hyperparameter search space than grid search, since in many applications the performance of a predictor can depend on a relatively small subset of the hyperparameters.

The results of their research showed how random search serves as an important baseline for automated model configuration. Furthermore, this research encourages the search for more sophisticated search strategies. These strategies have the potential to target computational resources more effectively. Nonetheless, random sampling does not take advantage of any relationships between previously evaluated candidate configurations and so motivates the use of adaptive as well as population-based optimization techniques. Table 2 depicts the evolution of computational approaches for automated composite defect recognition.

Table 2. Evolution of computational approaches for automated composite defect recognition.

Approach	Principal capability	Major limitation
Traditional ML	Classification using extracted material features	Depends on handcrafted descriptors.
Conventional CNN	Automatic hierarchical feature learning	Sensitive to hyperparameter selection.
Region-based CNN	Defect localization and segmentation	Higher computational and annotation requirements.
Transfer-learning CNN	Reuses pretrained visual features	Domain mismatch may reduce performance.
Optimization-assisted CNN	Systematic architecture and training parameter selection	Additional optimization cost.

KerasTuner-Based CNN Optimization

Because deep-learning architectures are getting so complex, automated hyperparameter optimization frameworks have been proposed to alleviate reliance on manual experimentation and simplify the model choice. KerasTuner is an extensible framework for the selection of hyperparameters in Keras neural networks.

The framework incorporates search algorithms namely Random Search, Bayesian Optimization, Hyperband for researchers to explore alternative methods using the unmodified search space and objective function. KerasTuner can be used for the optimization of number of convolutional filters, kernel sizes, number of hidden units, dropout rates, learning rates, and other training-related parameters in CNN based defect classification.

A candidate configuration is built and trained according to a selected search strategy and evaluated on an objective such as validation accuracy or validation loss. Through this optimization procedure, a reproducible method of configuring the network is created, provided that the algorithm that will be used to search for the optimum, the bounds for the search, the computational budget and choice of random initializations are documented clearly.

The efficacy of KerasTuner relies on the search algorithm utilized and the location of candidate configurations within a pre-defined hyperparameter space. Random Search explores combinations independently and can provide useful baseline performance when the most influential parameters are not known beforehand. Bayesian Optimization uses past evaluation information to steer future trials towards promising areas within the search space.

Hyperband uses a strategy to allocate resources to candidate configurations. Certain poorly performing candidates are stopped while other promising configurations are allocated further training resources. Li et al. [16] invented the Hyperband optimization framework and demonstrated the benefits of adaptive resource allocation for fast tracking hyperparameter search in machine learning applications.

These strategies demonstrate the flexibility of KerasTuner, but they also indicate that KerasTuner refers to an optimization framework, not one specific search algorithm. Thus, it is essential that comparative studies clearly state which KerasTuner search strategy is being applied to interpret performance accurately.

Particle Swarm Optimization for CNN-Based Defect Classification

Particle Swarm Optimisation (PSO) is a population-based metaheuristic optimization technique modelled on the behaviour of biological systems examples such as bird flocks and fish schools coordinating movement.

The PSO approach defines a population of candidate solutions or particles that move through a multi-dimensional search space. Each particle representing a possible solution to the problem updates its position based on information based on its previously recognised best position and the best solution identified by the population. The individual experiences and the group intelligence of the particles in PSO achieves a balance between local fine-tuning and global diversification without gradient information of the optimization objective.

Due to its comparatively simpler mathematical structure, adaptability, and capacity to deal with nonlinear optimization problems, PSO has been widely studied for model configuration and hyperparameter selection in artificial neural networks.

Combining PSO with deep networks allows the systematic procedure to determine the configuration of a network instead of trying out many. Ribalta Lorenzo et al. [17] investigated the use of PSO to select hyperparameters in deep neural networks and showed that it is suitable for various configuration types. Utilizing the network performance to influence the update of the particle population during the optimization process.

The investigation revealed that PSO can be a substitute for traditional parameter-search techniques with a potential for parallel execution opportunities. Ye [18] developed a PSO-based automatic parameter-selection framework that integrated global population search with gradient neural network training.

The framework encoded candidate network configurations as multidimensional particles and assessed them with trained neural-network models. The experiments presented showed better generalization than the random-search method analyzed in that study, constituting yet more evidence that population-based optimization is useful for configuring neural networks.

Grey Wolf Optimization for Enhanced CNN Performance

Grey Wolf Optimization (GWO) has become a population-based metaheuristic to solve complex nonlinear optimization problems particularly with high-dimensional search space and decision variables' interaction. GWO creates computational optimization framework based on leadership hierarchy and coordinated hunting behaviour of grey wolves.

The GWO framework divides the candidate solutions into four levels which are alpha (α), beta (β), delta (δ) and omega (ω). The three best wolves will lead the position update of the remaining population. The optimization process mathematically describes a strategy for searching, encircling, and attacking prey, using adaptive position-update equations.

Mirjalili et al. used numerical benchmark functions and engineering optimization problems. The results revealed that the performance of the proposed algorithm was competitive compared to the results of several other techniques. These include PSO, DE, and GSA. The results suggest that GWO might be a good candidate algorithm for optimization problems with complex search spaces and nonlinear objective functions.

Kumar et al. [19] examined an improvement in the Grey Wolf Optimization approach for the hyperparameter selection of CNN in kidney image classification. The problem of selecting an appropriate quantity and size of convolutional filters has been addressed. The standard GWO algorithm has insufficient global exploration and local exploitation due to which the GWO algorithm has been suggested to be adapted.

The Enhanced Grey Wolf Optimization (E-GWO) approach makes modifications which improve search behavior, which is tested against the conventional CNN, Galaxy Algorithm (GA), PSO, and standard GWO approaches. Within the experimental setting of the authors, the proposed search method showed to be superior compared to alternative optimizers.

The investigations highlighted in the periodic application of the GWO showed in the work which focused on medical images rather than composite materials that the GWO based optimizations are applicable in convolution architecture selections while also demonstrating that various changes in the working mechanism can impact classification performance. Table 3 shows the comparative assessment of optimization-assisted CNN approaches relevant to automated defect recognition.

Table 3. Comparative assessment of optimization-assisted CNN approaches relevant to automated defect recognition.

Reference	Optimization approach	Application
O'Malley et al. [4]	KerasTuner	General deep-learning model configuration.
Mirjalili et al. [6]	GWO	General numerical and engineering optimization.
Beldar [7]	KerasTuner and CNN-PSO	Fiber-reinforced composite defects.
Ribalta Lorenzo et al. [17]	PSO	Deep neural-network hyperparameter selection.
Kumar et al. [19]	E-GWO-CNN	Kidney-image classification.
Bergstra and Bengio [20]	Random Search	General hyperparameter optimization.
Yu et al. [21]	MGWO-CNN	Crop disease and pest classification.

Evidence suggests that optimization-assisted deep learning can potentially lead to improved automated defect classification via hyperparameter selection. The optimization mechanisms, evaluation conditions, and target applications differ in published studies which limit the generalisability of individual performance findings. The studies reviewed here do not establish sufficiently the direct comparison of KerasTuner, CNN-PSO, and CNN-GWO for the same fiber-reinforced composite defect classification task. This can be translated into a controlled comparative framework which allows examining the techniques beyond numerical accuracy values in isolation.

The current research contribution is not the new CNN architecture nor a new metaheuristic algorithm. It is a systematic and comparative evaluation of existing optimization techniques on a specific problem of composite-defect classification that was developed. The study will assess the classification

performance of CNN–GWO relative to CNN–PSO and KerasTuner and the trade-offs and generalization performance involved in it.

The proposed framework aims to give a reproducible methodology for intelligent inspection of materials through the linking of optimization problems identified in the literature with the morphological complexity of steel-fiber-reinforced composite defects. A better-informed selection of optimization-assisted CNN architectures for automated quality assessment and further industrial inspection applications can be expected from the resulting analysis. Table 4 shows the research gaps addressed by the proposed comparative framework while Table 5 depicts the comprehensive comparative analysis of existing studies on fiber-reinforced composite defect detection and deep-learning optimization

Table 4. Research gaps addressed by the proposed comparative framework.

Identified research gap	Proposed research response
Limited controlled comparison of three optimization strategies	Compare KerasTuner, CNN–PSO, and CNN–GWO.
CNN sensitivity to architecture and training hyperparameters	Apply systematic optimization within a common search space.
Difficulty distinguishing morphologically similar defects	Evaluate optimized CNN representations for multiple defect classes.
Variation in datasets and experimental conditions	Standardize preprocessing, partitioning and evaluation.
Limited assessment of computational efficiency	Compare convergence, optimization time and inference cost.
Inconsistent generalization assessment	Use independent test data and specimen-level separation where possible.

Table 5. Comprehensive comparative analysis of existing studies on fiber-reinforced composite defect detection and deep-learning optimization.

Ref.	Year	Author/study	Inspection data/application	Methodology	Principal contribution	Reported performance/findings	Identified limitation
[1]	2023	Zhang et al.	Steel-fiber-reinforced concrete	Review of fracture models and experimental methods	Analysed fracture behaviour and effects of steel-fiber distribution	Established major factors influencing fracture resistance	No direct comparison of DL optimization algorithms.
[2]	2021	Ding et al.	Steel-fiber-reinforced concrete crack images	YOLOv4, Mask R-CNN, improved Mask R-CNN	Developed and compared automated crack-identification methods	YOLOv4: 82.60%; Mask R-CNN: 90.44%; improved Mask R-CNN: 96.09% average prediction accuracy	Primarily crack identification; limited multi-defect coverage.
[3]	2020	Yang and Shami	Machine learning models across diverse applications	Comparative review of hyperparameter optimization techniques and benchmark experiments	Reviewed optimization methods, libraries, frameworks, and research challenges	Compared optimization methods to identify effective hyperparameter configurations	Method effectiveness may vary across problems; universal optimal configurations remain challenging.
[4]	2019	O'Malley et al.	Deep-learning model configuration	KerasTuner software	Provided an automated tuning framework supporting multiple search algorithms	Practical hyperparameter-search framework	Performance depends on selected tuner, search space and trial budget.

[5]	1995	Kennedy and Eberhart	General optimization benchmarks	Particle Swarm Optimization	Introduced particle-based collective search	Established the PSO optimization principle	Original work does not address CNN hyperparameters.
[8]	2025	Medeghini et al.	Steel-fiber-reinforced concrete	Fiber orientation analysis; imaging and micro-CT	Examined relationships between fiber orientation and post-cracking behaviour	Demonstrated the significance of fiber orientation	Focuses on material characterization rather than automated defect classification.
[9]	2022	Amin et al.	Steel-fiber-reinforced concrete	Systematic literature review	Summarized mechanical behavior, material properties and research trends	Identified key developments in SFRC	Does not experimentally evaluate image-based defect classification.
[10]	2024	Hampe	SFRC micro-CT specimen	Three-dimensional X-ray micro-CT acquisition	Provided a volumetric image dataset for studying composite structure	Public microstructural imaging resource	Limited specimen diversity; not a ready-made multi-class classification benchmark.
[11]	2022	Chen et al.	Fiber-reinforced polymer composites	Review of NDT and defect-evaluation techniques	Examined techniques for composite-defect characterization	Established capabilities of different inspection modalities	Does not provide a controlled CNN optimization comparison.
[12]	2024	Huang	Fiber-reinforced polymer composites	Review of non-destructive inspection methods	Discussed inspection capabilities and limitations	Comparative technical review	Focuses on inspection techniques rather than hyperparameter optimization.
[13]	2026	Liu et al.	Steel-fiber-reinforced concrete	Review of mechanical properties and durability	Examined SFRC behavior and durability	Material-performance synthesis	No image-based CNN optimization experiment.
[14]	2024	Daghigh et al.	Composite-material defects and damage	Review of ML and DL approaches	Investigated ML methods for composite damage identification and analysis	Identified opportunities for data-driven composite assessment	Review study; no unified three-optimizer experiment.
[15]	2015	LeCun et al.	General image and representation-learning applications	Deep neural networks	Established principles of hierarchical representation learning	Foundational DL review	Does not address composite defect inspection specifically.
[16]	2018	Li et al.	General hyperparameter optimization	Hyperband	Introduced adaptive resource allocation among candidate configurations	Improved resource efficiency in studied optimization tasks	Not specifically developed for composite CNN classification.
[22]	2024	Cheng et al.	Composite-material defect images	Deep-learning-based defect detection	Developed a method to improve composite-defect recognition in challenging imaging environments	Demonstrated application-specific detection improvements	Focus on detection architecture rather than three-way hyperparameter comparison.

[23]	2025	Bai et al.	Industrial material defects	Survey of ML-based defect detection	Synthesized ML methods for material inspection	Broad comparison of defect recognition approaches	Review study; no specific SFRC three-method experiment.
[24]	2024	Khanam et al.	Industrial surface-defect images	Review of CNN-based inspection	Discussed CNN classification, detection and deployment	Identified high-performing industrial inspection applications	Dataset heterogeneity and model-complexity challenges.
[20]	2012	Bergstra and Bengio	General machine-learning models	Random Search	Established the effectiveness of random hyperparameter sampling	Demonstrated advantages over grid search in studied settings	Does not adapt particles or population positions through collective learning.
[25]	2024	Yu et al.	Large-scale optimization and real-world engineering problems	Velocity integration, Inverse Multi-quadratic Function (IMF), and adaptive population updates	Enhanced exploration, convergence speed, and optimization capability	Outperformed comparison algorithms by 85.4–97.4% across CEC benchmarks; solved 19 engineering problems	Original GWO faces slow convergence and reduced problem-solving capability on large-scale problems; no image-based CNN optimization reported.
[26]	2025	Manickaraj et al. – Hybrid jute/BSL composites	Natural-fiber composites for engineering applications	Mechanical testing, FTIR, SEM	Evaluated effects of jute, banana stem leaves, and tamarind shell powder on composite performance.	20J10BSL: tensile 40.4 MPa; flexural 86.0 MPa; impact 3.56 J. Improvements: 24.6%, 27.9%, and 39.1%, respectively.	Limited composition range; long-term durability and broader application validation were not reported in the supplied summary.
[27]	2025	Palanisamy et al. – Machine learning for natural-fiber composites	Material selection, property prediction, and composite design optimization	Review of ML algorithms, evolutionary computing, databases, and product-development applications	Synthesized ML approaches, benefits, limitations, and emerging research directions for NFCs.	ML supports prediction and optimization across material selection and end-use development.	Model interpretability, data availability, and materials-science implementation challenges remain.
[28]	Not specified	Sumesh et al. – Natural/glass-fiber hybrid composites	Mechanical performance, friction, and wear applications	Tensile, flexural, impact, wear testing, and SEM	Investigated hybrid-fiber loading and alkali treatment effects on epoxy composites.	20 wt% AF + 20 wt% GF, with 5% alkali treatment, showed reported tensile improvement of 62.6% against the untreated reference.	Full experimental details and comparison baseline require verification; long-term performance was not specified.
[29]	2025	Bachtiar et al. – Sugar palm fiber/TPU composites	Mechanical, thermal, and morphological applications	NaOH treatment, extrusion, compression molding, XRD, FTIR, SEM, EDX, TGA	Established the influence of alkaline treatment on SPF–TPU compatibility and performance.	2% NaOH optimized flexural strength; 6% maximized impact strength and thermal stability.	Higher NaOH concentrations caused surface cracks; optimal treatment varied by target property.
[30]	2022	Palanisamy et al. – Acacia caesia bark fibers	Potential reinforcement for polymer composites	NaOH treatment, ASTM D3822 tensile testing, optical microscopy, SEM	Evaluated treatment-dependent fiber strength and fracture morphology.	10% NaOH: average tensile strength 16.2 ± 5.63 N; elongation $4.6 \pm 1.0\%$.	Single-fiber results do not establish composite-level performance; broader application testing was not reported.

PROPOSED METHODOLOGY

The proposed research incorporates Convolutional Neural networks (CNNs) with 3 optimization strategies to automate the defect classification of steel-fiber-reinforced composites. The three optimization strategies involve KerasTuner, Particle Swarm Optimization (PSO), and Grey Wolf optimization (GWO). The method will be used to examine the implications of various optimization mechanisms on the chosen architecture of the CNN, feature-learning ability, convergence behaviours, classification performance and computational cost.

In contrast to methods that rely solely on one optimisation algorithm, the present study investigates three parallel optimisation pipelines that operate on an identical image dataset under similar experimental conditions. The automated hyperparameter-search benchmark is KerasTuner, the CNN-PSO is based on collective particle optimization, and the CNN-GWO is based on leadership-guided population search. The aim of the comparative methodology was to investigate the relative suitability of the methods for the recognition of fiber-matrix debonding, fiber pull-out, fiber fracture, voids, and microcracking on the condition of available and verified labels of the datasets.

The proposed framework consists of eight processes that are interconnected, which include:

- Acquisition and verification of composite defect images.
- Pre-processing and partitioning of the dataset.
- Development of a common CNN architecture and hyperparameters search space.
- Independent optimization KerasTuner, PSO, and GWO.
- Evaluation of candidate CNNs using a common fitness criterion.
- Selection and training of the best CNN configuration obtained from each method.
- Prediction of defect categories using the independently evaluated models.
- Comparative assessment using classification and computational performance measures.

This arrangement allows the different optimization approaches to be evaluated under controlled conditions while retaining their individual search characteristics. The methodology aims to allow for reproducible analysis and to isolate improvements due to hyperparameter optimization from variations arising from inconsistent datasets, pre-processing, or evaluations.

Overall Architecture of the Proposed Comparative Framework

The framework combines learning convolutional features and hyperparameter optimization in a single computational framework. The input consists of labelled microscopic or other suitable inspection images representing selected defect categories in steel-fiber-reinforced composites. In the processing stage of pre-processing, the acquired images are resized, intensity normalized and enhanced at a controlled level to have similar input characteristics.

When necessary, operations to reduce noise and enhance contrast are applied without damaging the fine boundaries of cracks, fibers, and interfacial discontinuities. The processed images are then divided into training, validation, and test sets. To diversify the training sample without including augmented versions, image augmentation is applied only on the training subset but not on validation or test set.

The image fusion techniques – like stacking, averaging, and principal component analysis – are explored. Low-level features – like edges and local intensity variation – are captured by the convolutional layers, while complex texture representations and morphological structures are captured by deeper ones. Instead of adjusting all network hyperparameters manually, the design is paired with three separate optimizers. Different search methods, in their own ways, explore candidate CNN configurations, satisfying performance while using a predefined search space. The final model training and independent evaluation are based on the best configuration identified by each optimizer.

The three methods mainly differ in their hyperparameter-search mechanisms. KerasTuner employs a chosen automated search algorithm to evaluate candidate configurations. PSO updates candidate configurations by altering particle velocity and position. GWO modifies the position of the candidates using alpha-, beta-, and delta-guided search. Despite these differences, the three methods will ultimately optimize the same CNN design with comparable objectives and budget. The resulting models are evaluated in terms of standard classification metrics, permitting a direct check of the optimization. Figure 7 shows the conceptual workflow for the hyperparameter optimization of the three-way CNN. The workflow stresses the importance of using the same data partitions and evaluation plans for the 3 optimization approaches. These controls are important as variations in dataset composition, pre-processing and conditions of training may affect classification performance regardless of optimizer. As a result, the proposed framework aims to isolate the contribution of hyperparameter-search strategies while allowing flexibility to investigate differences in their convergence and computational requirements.

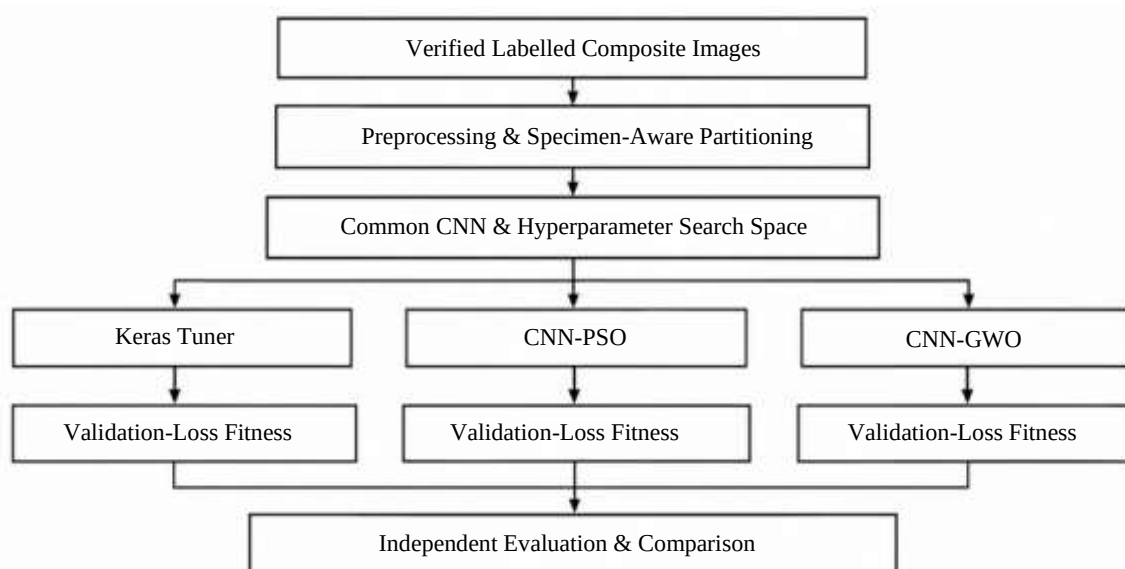


Figure 7. Conceptual workflow for three-way CNN hyperparameter optimization.

Dataset Acquisition, Pre-Processing, and Data Partitioning

In the first step of the procedure, suitable images of the microstructural features of steel-fiber-reinforced composites are collected. Depending on the incident dataset and the type of defect to be analyzed, the imaging methods could be scanning electron microscopy, optical microscopy, or X-ray micro-computed tomography. Images used for training must carry verified classification labels representing the respective material conditions.

The fiber–matrix debonding, fiber pull-out, fracture, voids and microcracking are the defect categories. The categories materialize from damage mechanisms found commonly in the composite inspection literature; however, the final number of classification classes and their definition must directly relate to the actual annotations available in the experimental dataset.

There may be variations in the resolution, lighting, background texture, noise level, and contrast of the acquired images. Inconsistency may impact convolution filter response and reduce reliability of the learning representations. As such, a standard pre-processing pipeline is designed, which has identical input characteristics and retains defect information.

Resizing images standardizes the sizes of the datasets compatible with the chosen CNN architecture, while intensity normalization standardizes the pixel-value scales. Controlled Gaussian filtering can be considered to reduce acquisition noise. Contrast enhancement or morphological operations can also be

considered if they increase the visibility of relevant structures. The application of these operations must be cautious because aggressive smoothing or thresholding may filter out small cracks and the interface boundary essential for accurate classification.

After pre-processing, the dataset is split into training, validation, and test set in the ratio of 70:15:15 as per the original paper. The CNN weights are estimated using the training subset and performance criterion for selecting hyperparameter configuration are used with validation subset. The testing subset that is independent is not used for the selection of hyperparameters, only after optimization and final training.

When many images come from a single specimen or micro-CT volume, specimen-level grouping should be used before any partitioning to limit information leakage. The training set only is subjected to augmentation, with appropriate rotations, flips and other defect-preserving transformations. Table 6 illustrates the proposed image preprocessing and dataset preparation framework.

Table 6. Proposed image preprocessing and dataset preparation framework.

Stage	Proposed operation	Purpose
Image acquisition	Labelled microscopy / inspection images	Create valid defect dataset.
Quality review	Verify labels, specimens, duplicates	Avoid ambiguity and leakage.
Preprocessing	Resize; normalize; controlled noise/contrast handling	Retain defect-sensitive boundaries.
Partitioning	70:15:15 proposed split, grouped by specimen where possible	Separate fitting, selection and testing.
Augmentation	Training-only suitable rotation and flips	Increase diversity without leakage.
Image acquisition	Labelled microscopy / inspection images	Create valid defect dataset.

The aim of the pre-processing process is to avoid physically creating defects with enhanced design rather than artificially generating patterns that do not represent the material. To ensure consistency in experiments, the three CNN optimization approaches will use the same pre-processing rules and dataset partitions.

Although the dataset includes five defect types, further testing is needed to check how well the model works with different specimens and imaging condition

Comparative Evaluation and Final Model Selection

The final model selection will be done based on comparative assessment. Upon the completion of the optimization, the KerasTuner, PSO, and GWO hyperparameter configurations will be used in the construction of the optimized CNN models. A common final training procedure will be carried out on each selected configuration, which will then be evaluated on the held-out test set. Classification measures complementary to the predictive effectiveness of three models will be applied to them for evaluation purpose.

Accuracy is the ratio of samples that were classified correctly, and precision evaluates the reliability of positive prediction. Recall is the capability to identify sample points of a class. The F1-score summarizes the precision-recall balance. Confusion matrices will also be referred to identify the errors specific to each class, especially misclassification of defects.

To account for differences in class distribution, the macro-averaged and weighted-averaged metrics may be reported for multi-class evaluation. The main measurement tools are stated as follows.

$$\text{Accuracy} = \text{Correct predictions} / \text{Total predictions} \quad (1)$$

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

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

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

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively, and P and R denote precision and recall.

In this situation, TP, FP and FN are the true positives, false positives and false negatives, respectively, and P and R are precision and recall.

Besides classification performance, the comparative study will also assess optimization convergence, computational cost and model-inference requirements. The graph will show the best validation loss at every iteration, while the computational evaluation will measure the total optimization time and inference latency when applicable. It is important to put additional measures in place since a model which achieves minimally improved accuracy on classification may require much more computational power. Table 7 presents a comparative evaluation framework for CNN hyperparameter optimization using KerasTuner, PSO, and GWO.

Table 7. Comparative evaluation framework for CNN hyperparameter optimization using KerasTuner, PSO, and GWO.

Measure	KerasTuner	CNN-PSO	CNN-GWO
Dataset and split	Identical	Identical	Identical.
Validation objective	Cross-entropy	Cross-entropy	Cross-entropy.
Search budget	To fix	To fix	To fix.
Independent test accuracy	To measure	To measure	To measure.
Macro precision / recall / F1	To measure	To measure	To measure.
Best validation loss	To measure	To measure	To measure.
Optimization time / inference	To measure	To measure	To measure.

The ultimate choice of model will depend mainly on classification performance on independent test data, aided by class specific tests and cost analysis. The proposed comparative study focuses on CNN-GWO as the priority strategy but does not pre-suppose its superiority in the experimental design. Any observed improvement must be proven to be reproducible, under the same evaluation conditions.

In Equations (8)–(11) “ $\odot$ ” denotes element-wise multiplication, X_p a guiding wolf position, A , C coefficient vectors, and X_1 – X_3 positions that are updated independently using α , β and δ wolves. According to Equations (15)–(18), TP_k , FP_k and FN_k are counted for class k , and macro-averages can be reported across the K defect classes. Figure 8 shows the training accuracy versus epoch for CNN-GWO, CNN-PSO and CNN-KerasTuner.

The reconstruction of the supplied plot shows the training accuracy versus epoch for CNN-GWO, CNN-PSO and CNN-KerasTuner.

The accuracies of all three configurations increase quickly due to the early epochs and near plateau towards the end of training. In the reconstructed learning curves, the model exhibiting the maximum late-epoch training accuracy is CNN-GWO, followed by CNN-PSO and then CNN-KerasTuner. The given ordering reflects the provided experiment and is not meant to indicate statistically significant superiority on the algorithm level. Figure 9 shows the validation accuracy versus epoch for the three

optimized CNN configurations.

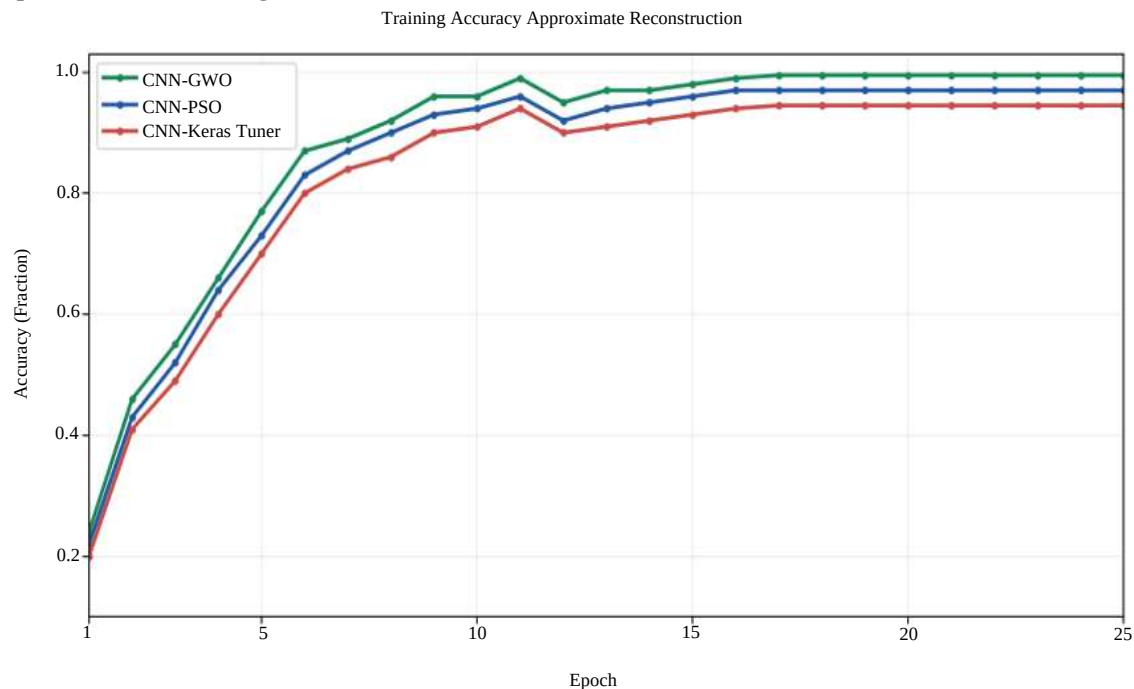


Figure 8. Training accuracy versus epoch for CNN–GWO, CNN–PSO, and CNN–KerasTuner.

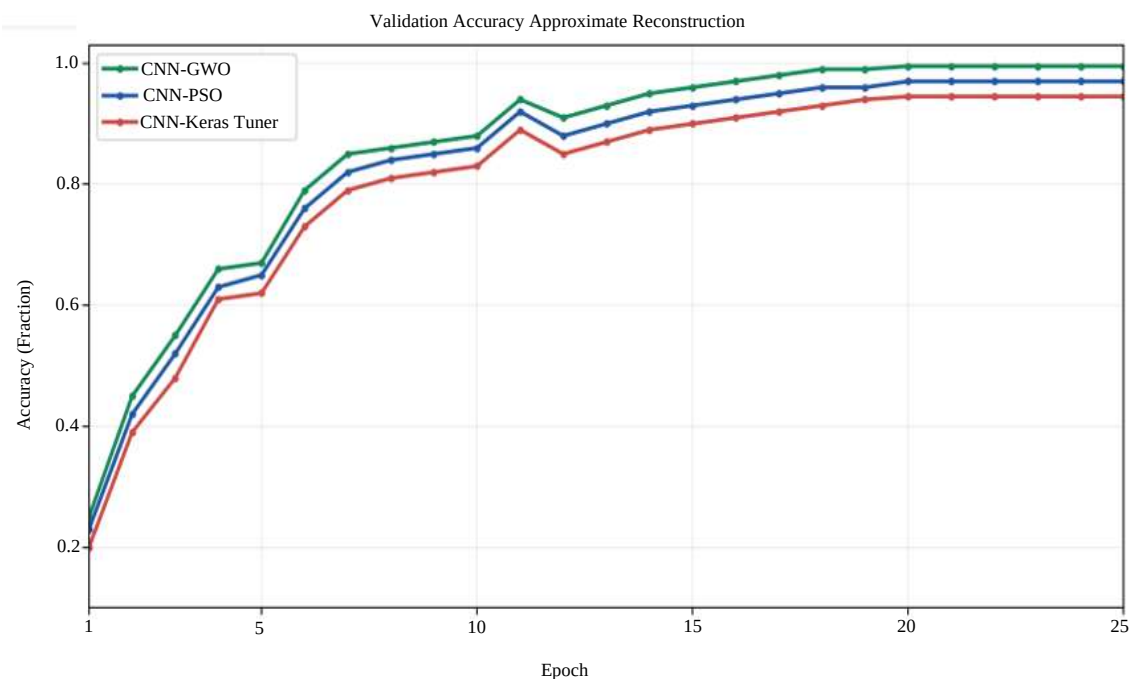


Figure 9. Validation accuracy versus epoch for the three optimized CNN configurations.

Validation accuracy against epoch for three optimized CNN configurations, approximate reconstruction. The curves of validation accuracy show a similar late-epoch ordering. Temporary fluctuations during training signifies the existence of an informative non-transient plateau; not individual epochs though the plotted CNN–GWO curve continues to be higher than at the end of the displayed run, it will still need to be confirmed using repeated runs. Figure 10 shows the training loss across 25 epochs for CNN–

GWO, CNN-PSO, and CNN-KerasTuner.

Training Loss Approximate Reconstruction

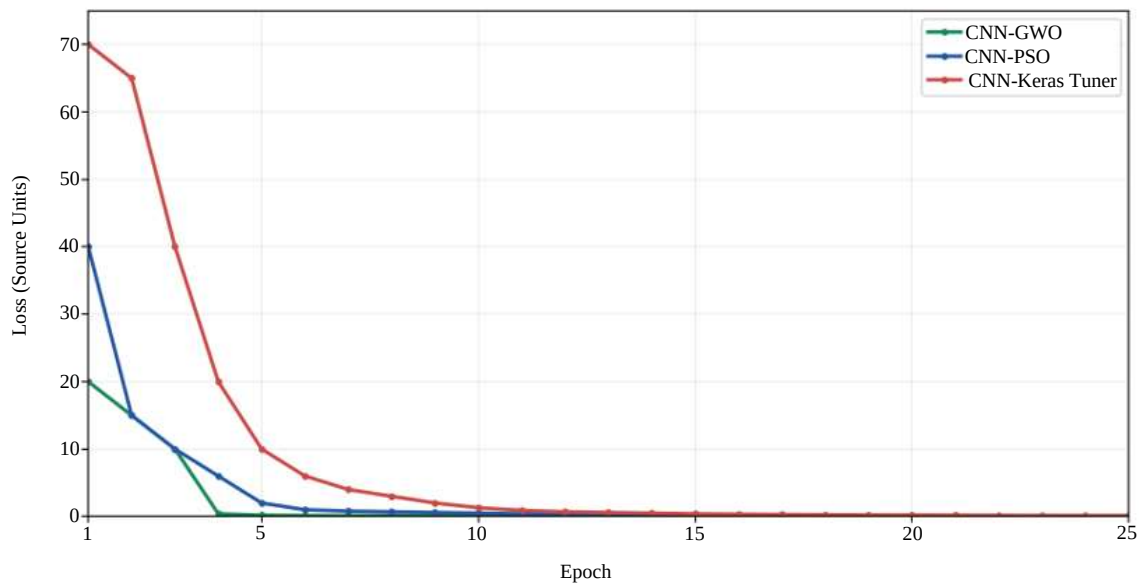


Figure 10. Training loss across 25 epochs for CNN-GWO, CNN-PSO, and CNN-KerasTuner.

Training loss over 25 epochs: CNN-GWO, CNN-PSO, and CNN-KerasTuner, approximate reconstruction. The retraining loss reduces significantly in the early epochs. CNN-GWO is shown to reach a near-zero region earlier in the depicted trajectories; however, the original axes span large early losses, compressing differences visually in late epochs. The definition of raw loss, reduction mechanism, and training logs would be required for a strict magnitude comparison. Figure 11 shows the approximate reconstruction of validation loss versus epoch for the three configurations

Validation Loss Approximate Reconstruction

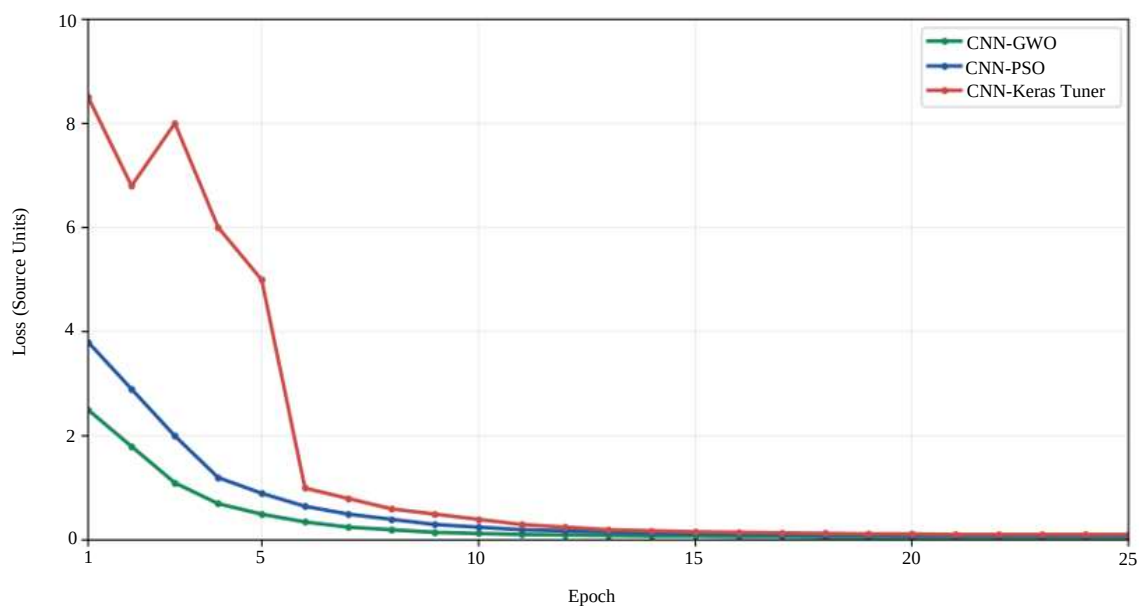


Figure 11. Validation loss versus epoch for the three configurations Graph of validation loss against epochs for different configurations in FLINT. The strength of all methods is confirmed by the validation-loss plots appearing low. As can be observed from the displayed run, GWO may converge rather favorably with respect to CNN. However, differences near zero cannot be accurately established at the original figure resolution. The curves cannot establish whether the methods use the same search

budget or training conditions.

Table 8 shows the approximated values visually interpreted from the reconstructed curves at epoch 25. All accuracies are presented in percentage for readability. These values come from images, not original experimental files. The information presented in Table 8 is not based on raw experiments but is comparative read-offs.

Table 8. Approximated values visually interpreted from the reconstructed curves at epoch 25.

Metric (epoch 25)	CNN-GWO	CNN-PSO	CNN-KerasTuner
Training accuracy	≈99.5%	≈97.0%	≈94.5%
Validation accuracy	≈99.5%	≈97.0%	≈94.5%
Training loss	≈0.030	≈0.100	≈0.140
Validation loss	≈0.060	≈0.090	≈0.110

The final epoch reconstructed validation accuracy for CNN-GWO is about 2.5 percentage point higher than that of CNN-PSO and about 5.0 percentage point higher than that of CNN-KerasTuner, see Table 8. The differences are arithmetic comparisons between visually estimated points, not confidence-bounded measured improvements. Figure 12 shows the grouped comparison of approximate epoch-25 validation accuracy and validation loss for the three optimized CNN models.

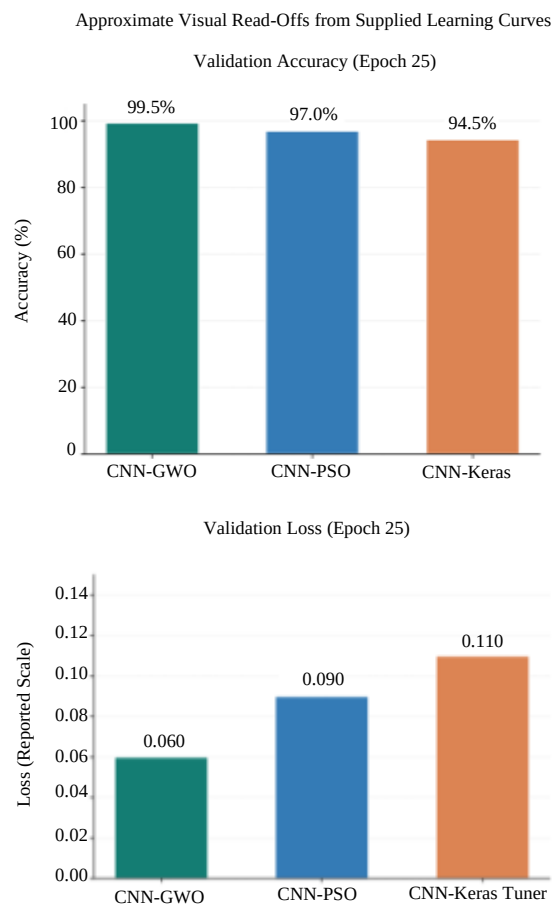


Figure 12. Grouped comparison of approximate epoch-25 validation accuracy and validation loss for the three optimized CNN models.

The estimated values of image reconstructions instead of source logs for the three optimized CNN models are compared in terms of the approximate validation accuracy and validation loss. A cross-model comparison comparable to the bar-chart presentation in the referenced paper is shown in Figure 12 but implemented using the three algorithms investigated in this work. The first graph depicts the estimated validation accuracy while the second one showcases the estimated validation loss on its scale.

Different panels prevent misinterpretation between units.

The greater accuracy trajectory of CNN-GWO coincides with GWO finding a beneficial CNN configuration in this particular experiment. The KerasTuner can apply varied search strategies depending on the selected tuner. In contrast, PSO uses population-based exploration. However, the provided plots are not sufficient to determine if the observed differences are caused through search heuristics, search-budget allocation, initialization, CNN architecture, or dataset sampling. Algorithmic explanation needs to be construed as a hypothesis than a verified causal finding.

It has been observed that the training and validation curves have converged in this run, although they do not predict final performance on an independent test set. The provided source document did not include class-wise confusion matrices, precision, recall, F1-score, repeated-seed statistics, optimization runtime, parameter counts, and search-budget details. As the plots are manually derived from low-resolution images, the numbers in Table 8 and in Figure 12 are not very precise.

Figure 13(a–c) shows the confusion matrix. Confusion matrix of CNN-PSO demonstrates 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. CNN-GWO: 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. The classification performance of the CNN model optimized using KerasTuner for identifying five defect categories in fiber-reinforced steel composites: Fiber–Matrix Debonding, Fiber Pull-Out, Fiber Fracture, Voids, and Microcracks. The model correctly classified 702 out of 734 samples, achieving an overall accuracy of 95.64%. The diagonal elements indicate correct predictions, with 148 samples each for Fiber–Matrix Debonding and Fiber Pull-Out, 136 each for Fiber Fracture and Voids, and 134 for Microcracks. Minor misclassifications were observed across the defect categories, particularly between visually similar defect patterns. These results demonstrate the effectiveness of KerasTuner in optimizing CNN hyperparameters for multiclass defect identification.

Confusion Matrix for CNN with GWO

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				

(a)

Confusion Matrix for CNN with PSO

True Labels	Fiber-Matrix Debonding	158	0	0	0	0
	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
		Fiber-Matrix Debonding	Fiber Pull-Out	Fiber Fracture	Voids	Microcracks
		Predicted Labels				

(b)

Confusion Matrix for CNN with KerasTuner

True Labels	Fiber-Matrix Debonding	148	3	2	1	4
	Fiber Pull-Out	3	148	0	4	1
	Fiber Fracture	0	2	136	0	1
	Voids	1	2	0	136	2
	Microcracks	1	1	4	0	134
		Fiber-Matrix Debonding	Fiber Pull-Out	Fiber Fracture	Voids	Microcracks
		Predicted Labels				

(c)

Figure 13 Confusion matrix (a) CNN with GWO, (b) CNN with PSO, and (c) CNN with KerasTuner.

CONCLUSION

1. The present work analyses three different types of hyperparameter optimization strategies combined with Convolutional Neural Networks for the purpose of defect classification on fiber-reinforced composite image dataset. The first one was KerasTuner, followed by Particle Swarm Optimization and Grey Wolf Optimization. Throughout the 25-epoch training and validation accuracy curves shown, CNN-GWO achieved the highest visible late-epoch accuracy, while CNN-PSO performs better than KerasTuner-optimized CNN whose performance remains below the two swarm-based approaches.
2. The curves corresponding to the training loss indicate a notably fast decrease for CNN-GWO, while the validation loss trajectories converge towards low-loss areas for all three models. The findings demonstrate that CNN-GWO has a qualitative edge over the other two methods in the supplied experimental run.
3. The coloured evidence created is in line with the idea that population-based optimization can enhance the selection of CNN hyperparameters for complex material images where fine structural details and heterogeneous textures complicate classification. The GWO seems to predict a setup that follows a desirable trajectory of accuracy and loss in the assessed plots. Nonetheless, presently available evidence does not indicate that GWO is always superior to other strategies with respect to different datasets, independent test samples, random starting points, or equivalent computational budgets.
4. The comparative chart's approximate endpoint values were visually estimated from existing figures and were not derived from original epoch logs. Thus, they should not be interpreted as specific experimental results or proof of statistical significance.
5. Even so, the results provide a useful basis for the next phase of comparative evaluation. A full assessment in the future should test on a fresh, preferably specimen-independent test set, use per-class precision and recall, macro-F1, confusion matrices, variability across repeated runs, model size, counts of search-evaluation, and actual tuning/ inference times.
6. This would indicate if the CNN-GWO behaviour extends to more realistic curves and also sheds light on the trade-off between performance and cost. Provided these checks are satisfied, the three-method framework can assist towards a more reproducible and efficient deep-learning-assisted inspection of defects in fiber-reinforced composites.

REFERENCES

1. Zhang P, Wang C, Gao Z, Wang F. A review on fracture properties of steel fiber reinforced concrete. *J Build Eng*. 2023;67:105975.
2. 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;2021:9934250.
3. Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*. 2020;415:295–316.
4. O'Malley T, Bursztein E, Long J, Chollet F, Jin H, Invernizzi L, et al. KerasTuner [Internet]. 2019. Available from: Github.com
5. Kennedy J, Eberhart R. Particle swarm optimization. In: *Proceedings of ICNN'95—International Conference on Neural Networks*. 1995;4:1942–1948.
6. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Adv Eng Softw*. 2014;69:46–61.
7. Beldar P. Optimized convolutional neural networks for defect classification in fiber-reinforced composites. *Sigma*. 2026;44(3):2207–2218.
8. 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;26(1):481–500.
9. Amin MN, Ahmad W, Khan K, Ahmad A. Steel fiber-reinforced concrete: A systematic review of the research progress and knowledge mapping. *Materials (Basel)*. 2022;15(17):6155.
10. Hampe M. Micro-CT dataset of a steel fibre reinforced concrete sample. doi: [10.5281/zenodo.10785684](https://doi.org/10.5281/zenodo.10785684).
11. Chen J, Yu Z, Jin H. Nondestructive testing and evaluation techniques of defects in fiber-reinforced

- polymer composites: A review. *Front Mater.* 2022;9:986645.
12. Huang L. Review of nondestructive testing methods for fiber-reinforced polymer composites. *Acta Aeronaut Astronaut Sin.* 2024;45(5):529697.
13. Zheng Y, Lv X, Hu S, Zhuo J, Wan C, Liu J. Mechanical properties and durability of steel fiber reinforced concrete: A review. *J Build Eng.* 2024;82:108025.
14. Daghigh V, Daghigh H, Lacy TE Jr, Naraghi M. Review of machine learning applications for defect detection in composite materials. *Mach Learn Appl.* 2024;18:100600.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436–444.
16. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J Mach Learn Res.* 2018;18(185):1–52.
17. Lorenzo PR, Nalepa J, Kawulok M, Ramos LS, Pastor JR. Particle swarm optimization for hyperparameter selection in deep neural networks. In: *Proceedings of the Genetic and Evolutionary Computation Conference.* 2017. p. 481–488.
18. Ye F. Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data. *PLoS One.* 2017;12(12).
19. Kumar D. Grey Wolf Optimization based hyper-parameter optimized convolution neural network for kidney image classification. *Int J Intell Eng Syst.* 2022;15(3):319.
20. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res.* 2012;13(2).
21. Yu X, Zhang H, Duan Y, Wang X. An optimized convolutional neural network based on multi-strategy grey wolf optimizer to identify crop diseases and pests. *Displays.* 2026:103341.
22. 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;14(10):4161.
23. Bai J, Wu D, Shelley T, Schubel P, Twine D, Russell J, et al. A comprehensive survey on machine learning driven material defect detection. *ACM Comput Surv.* 2025;57(11):1–36.
24. Khanam R, Hussain M, Hill R, Allen P. A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access.* 2024;12:94250–94295.
25. Yu M, Xu J, Liang W, Qiu Y, Bao S, Tang L. Improved multi-strategy adaptive Grey Wolf Optimization for practical engineering applications and high-dimensional problem solving. *Artif Intell Rev.* 2024;57(10):277.
26. Manickaraj K, Karthik A, Palanisamy S, Jayam M, Ali SK, Sankar SL, et al. Improving mechanical performance of hybrid polymer composites: Incorporating banana stem leaf and jute fibers with tamarind shell powder. *BioResources.* 2025;20(1):1998.
27. Palanisamy S, Ayrilmis N, Sureshkumar K, Santulli C, Khan T, Junaedi H, et al. Machine learning approaches to natural fiber composites: A review of methodologies and applications. *BioResources.* 2025;20(1).
28. Sumesh KR, Palanisamy S, Khan T, Ajithram A, Ahmed OS. Mechanical, morphological and wear resistance of natural fiber/glass fiber-based polymer composites. *BioResources.* 2024;19(2).
29. Bachtiar D, Mohammed AA, Palanisamy S, Imran AI, Siregar JP, bin Mat Rejab MR, et al. Effect of alkaline treatment on the thermal and mechanical properties of sugar palm fibre reinforced thermoplastic polyurethane composites. *Sci Rep.* 2025;15(1):14085.
30. Palanisamy S, Mayandi K, Dharmalingam S, Rajini N, Santulli C, Mohammad F, et al. Tensile properties and fracture morphology of Acacia caesia bark fibers treated with different alkali concentrations. *J Nat Fibers.* 2022;19(15):11258–11269.