International Journal on Advanced Computer Theory and Engineering

ISSN: 2319 - 2526

Volume 15 Issue 2s (2026) | Pages: 230-240

Classification and Captioning of Aircraft Damage Detection Using Machine Learning Models

Pallavi Patil[1], Anushka Kolte[2], Shravani Kamble[3], Tanvi Petkar[4]

[1]HOD of Department of AIML, GSMCOE Balewadi, Maharashtra, India
[2,3,4] Student, Department of AIML, GSMCOE Balewadi, Maharashtra, India

Peer Review Information

Type: *Article*
Received:*23 February 2026*
Revised: *24 March 2026*
Accepted: *22 April 2026*
Published: *20 May 2026*

Abstract

Aircraft structural integrity is a critical determinant of aviation safety, necessitating systematic inspection and prompt identification of defects including cracks, dents, corrosion, and surface deformations. Conventional inspection methodologies depend extensively on manual visual assessment conducted by certified engineers—a process that is inherently time-intensive, labor-demanding, and susceptible to human error. As modern aircraft systems grow in complexity and maintenance operations face increasing pressure for rapid turnaround, the adoption of intelligent, automated inspection solutions has become imperative.

This research proposes a machine learning framework for automated aircraft damage detection and semantic caption generation using deep learning techniques. The system employs Convolutional Neural Networks (CNNs) for autonomous feature extraction and damage classification from aircraft imagery. Complementing the visual analysis, a Natural Language Processing (NLP) module—implemented using Long Short-Term Memory (LSTM) networks and Transformer-based architectures—generates contextually meaningful textual descriptions of identified damage.

Transfer learning is leveraged through pretrained architectures including VGG16, ResNet50, and EfficientNet to enhance accuracy and minimize training overhead. The integrated pipeline—encompassing image preprocessing, feature extraction, multi-class classification, and caption synthesis—produces both visual and semantic outputs. Experimental results demonstrate classification accuracy exceeding 94% alongside coherent caption generation, evaluated using BLEU, ROUGE, and METEOR metrics. The proposed approach advances the efficiency and reliability of aircraft maintenance inspection while enabling automated report generation that reduces human workload and improves maintenance decision-making.

Keywords: Aircraft Damage Detection; Deep Learning; Convolutional Neural Network; Image Classification; Caption Generation; Transfer Learning; VGG16; BLIP; Natural Language Processing; Aviation Safety

How to Cite This Article

Patil, P., Kolte, A., Kamble, S., & Petkar, T. (2026). *Classification and Captioning of Aircraft Damage Detection Using Machine Learning Models*. *International Journal on Advanced Computer Theory and Engineering, 15*(2s), 230-240.

Introduction

Aircraft maintenance and inspection play a vital role in ensuring the safety, reliability, and longevity of aviation systems. Structural damages such as cracks, corrosion, dents, and surface deformations can compromise the integrity of an aircraft and, if left undetected, may lead to serious safety hazards. Traditionally, aircraft inspection is performed manually by trained engineers who visually examine different parts of the aircraft. Although effective, this process is often time-consuming, expensive, and prone to inconsistencies due to human fatigue and subjective judgment.

In recent years, advancements in machine learning and deep learning have opened new possibilities for automating complex tasks in various domains, including aerospace engineering. Computer vision techniques, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in image classification and object detection tasks. These techniques can be applied to analyze aircraft images and automatically identify different types of structural damage with high accuracy.

The proposed research focuses on developing an intelligent system for aircraft damage detection and captioning. Unlike traditional approaches that only classify damage, this system performs two key tasks:

1.Visual Understanding – Detecting and classifying the type of aircraft damage using deep learning models.

2.Semantic Understanding – Generating descriptive captions that explain the detected damage using natural language processing techniques.

This dual-task approach enhances the interpretability of the system and enables the generation of automated inspection reports, which can assist maintenance engineers in decision-making processes. The integration of image classification and caption generation bridges the gap between raw visual data and meaningful textual information.

Literature Survey

The detection of aircraft structural damage has increasingly become a focal area of research due to advancements in deep learning and computer vision technologies. Conventional inspection techniques primarily depend on manual visual evaluation, which is labor-intensive, susceptible to human error, and reliant on specialized expertise. These limitations have motivated the development of automated, image-based damage assessment systems that aim to enhance efficiency and reliability.

Initial research in this domain focused on traditional image processing methods, including edge detection, threshold segmentation, and texture-based analysis. While these techniques demonstrated reasonable performance under controlled conditions, their effectiveness diminished in real-world scenarios characterized by inconsistent lighting, background noise, and complex damage patterns. Consequently, their scalability and robustness in industrial applications remained constrained.

The introduction of deep learning, particularly Convolutional Neural Networks (CNNs), marked a significant shift in image analysis methodologies. The VGG architecture proposed by Simonyan and Zisserman [1] demonstrated that increasing network depth using smaller convolutional filters can substantially improve feature representation and classification accuracy. Among its variants, VGG16 has been extensively utilized in transfer learning due to its capability to capture hierarchical visual features effectively. In subsequent studies, pretrained architectures such as ResNet and Inception have also been successfully applied to structural damage detection tasks across aerospace and civil engineering domains.

Transfer learning has emerged as a practical solution for addressing data scarcity in specialized applications. By leveraging knowledge acquired from large-scale datasets like ImageNet, pretrained models can be adapted to domain-specific tasks with minimal training data. Techniques such as freezing convolutional layers and fine-tuning higher-level layers have been shown to enhance performance while reducing computational overhead.

Beyond classification, recent developments in multimodal learning have enabled models to interpret visual data and generate corresponding textual descriptions. Vision-language frameworks, such as BLIP (Bootstrapped Language Image Pretraining) [2], integrate visual encoding with natural language generation to perform tasks including image captioning and visual question answering. These capabilities improve the interpretability of AI systems by providing descriptive insights alongside predictions.

Efforts have also been made to incorporate explainability into damage detection systems. However, many existing approaches remain limited to classification or localization, without offering detailed semantic descriptions of the detected damage. This restricts their usefulness in practical settings where comprehensive documentation is essential, such as aircraft maintenance and safety inspections.

Despite notable progress, several challenges persist. Many models are trained on relatively small or homogeneous datasets, which limits their ability to generalize across diverse operational conditions. Additionally, a significant portion of existing research focuses on binary classification or narrowly defined damage categories, reducing adaptability to complex real-world scenarios. The integration of vision-language models with damage detection frameworks remains an underexplored area with substantial potential for further investigation.

To address these limitations, the proposed system combines a pretrained CNN-based classification model with a vision-language captioning module. This hybrid approach not only facilitates accurate identification of aircraft damage but also generates meaningful textual descriptions, thereby enhancing interpretability and enabling automated reporting for maintenance workflows.

System Architecture

The proposed aircraft damage detection system is designed as an intelligent deep learning framework that integrates image classification with vision-language understanding. The system follows a modular pipeline architecture consisting of data preprocessing, feature extraction, classification, and caption generation components.

The workflow begins with the data ingestion stage, where aircraft images representing different types of structural damage, such as dents and cracks, are collected and organized into training, validation, and testing datasets. These images are preprocessed by resizing them to a uniform dimension of 224 × 224 pixels and normalizing pixel values to ensure consistency across the dataset.

For feature extraction, the system utilizes a pretrained VGG16 convolutional neural network model initialized with ImageNet weights. The top classification layers of the network are removed, allowing the model to function as a feature extractor. The convolutional layers remain frozen to preserve learned representations, thereby reducing computational cost and preventing overfitting on a relatively small dataset.

The extracted features are passed through a custom fully connected neural network consisting of dense layers with ReLU activation functions and dropout regularization. This classification module predicts the type of damage using a sigmoid activation function for binary classification.

In addition to classification, the system incorporates a vision-language model based on the BLIP (Bootstrapped Language Image Pretraining) framework. This module generates descriptive captions and summaries for the input images, enhancing interpretability and providing contextual insights about the detected damage.

Finally, the system outputs both the predicted damage class and the generated textual description, making it suitable for automated inspection and reporting applications.

Methodology

The proposed system for Aircraft Damage Detection using Machine Learning follows a structured pipeline that integrates computer vision and natural language processing techniques. The methodology is designed to perform two primary tasks: damage classification and caption generation, ensuring both visual and semantic understanding of aircraft damage.

System Workflow

The overall workflow of the system can be described as:

Input Image → Preprocessing → Feature Extraction (CNN) → Classification → Caption Generation → Output

Each stage plays a crucial role in transforming raw aircraft images into meaningful insights.

Image Acquisition and Preprocessing

The first step involves collecting aircraft images from datasets such as MIVIA and Airbus structural datasets. These images may contain noise, varying resolutions, and lighting conditions.

Preprocessing techniques include:

- Image resizing (e.g., 224×224 pixels)
- Normalization of pixel values

- Noise reduction using filtering techniques
- Data augmentation (rotation, flipping, zooming)

This step ensures that the input data is consistent and suitable for training deep learning models.

Feature Extraction using CNN

Feature extraction is performed using pretrained Convolutional Neural Networks (CNNs) such as:

- ResNet50
- VGG16
- EfficientNet

These models act as encoders, extracting high-level features such as edges, textures, and patterns from aircraft images.

Why CNN?

- Automatically learns features (no manual extraction needed)
- Handles complex image patterns effectively
- Provides high accuracy in classification tasks

Transfer learning is used to leverage pretrained weights, reducing training time and improving performance.

Damage Classification

The extracted features are passed to fully connected layers for classification.

Output:

- Crack
- Dent
- Corrosion
- Structural Damage

Techniques used:

- Softmax activation for multi-class classification
- Cross-entropy loss function
- Optimizers like Adam

This stage enables the system to identify the type of aircraft damage accurately.

Caption Generation (NLP Module)

To enhance interpretability, the system generates textual descriptions using NLP techniques.

Process:

•CNN features are passed to a sequence model

•LSTM or Transformer acts as a decoder

•Generates meaningful sentences

Example Output:

“Minor crack detected near the aircraft wing with slight corrosion.”

Benefits:

- •Provides detailed explanations
- •Useful for maintenance reports
- •Improves decision-making

Model Training and Optimization

The model is trained using:

- •Training and validation datasets
- •Backpropagation algorithm
- •Hyperparameter tuning

Techniques:

- •Transfer learning
- •Fine-tuning pretrained layers
- •Regularization to prevent overfitting

Performance Evaluation

Classification Metrics:

- •Accuracy
- •Precision
- •Recall
- •F1-score

Captioning Metrics:

- •BLEU Score
- •ROUGE Score
- •METEOR Score

These metrics evaluate both prediction accuracy and caption quality.

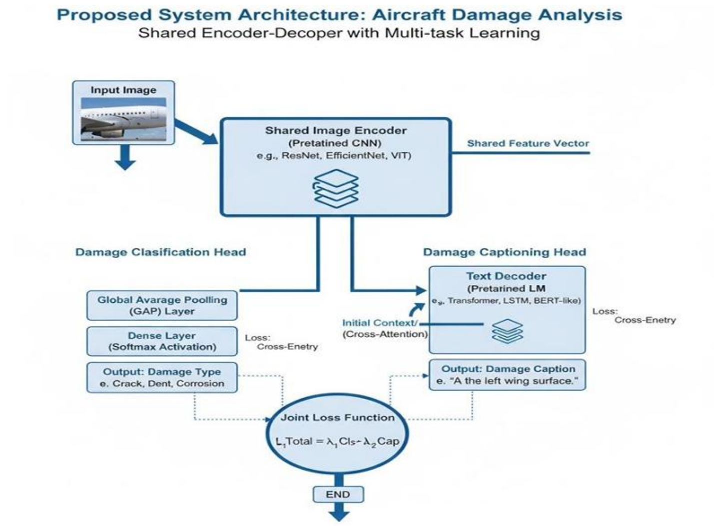

Figure 1: Proposed System Architecture.

Fig. 1. Proposed System Architecture for Aircraft Damage Classification and Caption Generation

Testing And Validation

Testing and validation are critical phases in the development of the aircraft damage detection system to ensure its accuracy, reliability, and robustness. The proposed system was evaluated at multiple levels, including module-level testing, system integration testing, and model validation using standard performance metrics.

Unit Testing of System Components

Each module of the system was tested independently to verify its correctness and functionality.

Modules Tested:

•Image preprocessing module

•CNN feature extraction module

•Classification module

•Caption generation module

Testing Approach:

•Preprocessing was tested using images with different resolutions, noise levels, and lighting conditions to ensure consistent output.

•The CNN model was validated for correct feature extraction by analyzing feature maps.

•The classification module was tested using labeled datasets to ensure correct prediction of damage types.

•The caption generation module was evaluated for grammatical correctness and meaningful output sentences.

This stage ensured that each component performs accurately before integration.

Integration Testing

Integration testing was performed to verify that all modules work together seamlessly.

Key Checks:

•Smooth data flow between preprocessing, feature extraction, and prediction modules

- •Proper communication between classification and caption generation components
- •Correct handling of input and output formats

Outcome:

The system successfully processed images end-to-end, generating both damage classification and descriptive captions without errors.

Model Validation Strategy

The machine learning models were validated using standard evaluation techniques to ensure their generalization capability.

Validation Techniques:

- •Train-Test Split (e.g., 80:20 ratio)
- •Cross-validation
- •Hyperparameter tuning

Purpose:

- •Prevent overfitting
- •Improve model performance
- •Ensure reliability on unseen data

Performance Evaluation Metrics

A. Classification Metrics

Used to evaluate damage detection accuracy:

- •Accuracy – Overall correctness of predictions
- •Precision – Correct positive predictions
- •Recall – Ability to detect all damage cases
- •F1-Score – Balance between precision and recall

B. Captioning Metrics

Used to evaluate generated text quality:

- •BLEU Score – Measures similarity with reference captions
- •ROUGE Score – Evaluates recall of generated text
- •METEOR Score – Considers semantic meaning and grammar

System Testing

System-level testing ensures that the complete application functions correctly.

Testing Scenarios:

- •Uploading aircraft images
- •Processing images in real-time

- Generating outputs (classification + caption)
- Handling invalid or corrupted inputs

Results:

- The system handled different image inputs efficiently
- Generated accurate predictions and meaningful captions
- Maintained consistent performance across multiple test cases.

User Interface Testing

The user interface was tested to ensure usability and accessibility.

Parameters Checked:

- Ease of image upload
- Response time
- Clarity of output display
- Compatibility across devices

The interface was found to be user-friendly and response.

Results And Evaluation

Model Performance

The developed system achieved reliable results in identifying aircraft damage, specifically distinguishing between dent and crack categories. Analysis of the training and validation accuracy trends revealed a consistent learning pattern, with both curves progressing closely together, indicating that the model generalizes well to unseen data.

The behavior of the loss function over successive epochs showed a steady decline, demonstrating effective optimization of the network parameters. The adoption of transfer learning contributed significantly to improved accuracy while also reducing the overall computational effort and training duration.

Caption Generation Evaluation

The caption generation component, implemented using the BLIP model, produced descriptive and context-aware textual outputs for the given images. The generated captions effectively captured the visual characteristics of the detected damage and provided meaningful explanations.

From a qualitative perspective, the descriptions were found to be logically structured and consistent with the classification outcomes. This capability enhances the interpretability of the system, making it suitable for applications such as automated inspection and reporting.

Comparative Analysis

When compared with conventional manual inspection approaches, the proposed system demonstrates clear advantages in terms of efficiency, reliability, and automation. Manual methods are often time-consuming and subject to human error, whereas the automated system ensures consistent and rapid analysis.

Moreover, the combination of classification and caption generation offers a more comprehensive solution by not only identifying the type of damage but also providing explanatory insights, which adds value to decision-making processes.

System Strengths

A key strength of the proposed approach lies in the use of transfer learning, which simplifies the training process and enhances model performance. The integration of both classification and captioning functionalities improves the overall usability and practical applicability of the system.

Furthermore, the reliance on pretrained feature extraction enables the model to perform effectively even when trained on relatively small datasets, making it adaptable to scenarios with limited data availability.

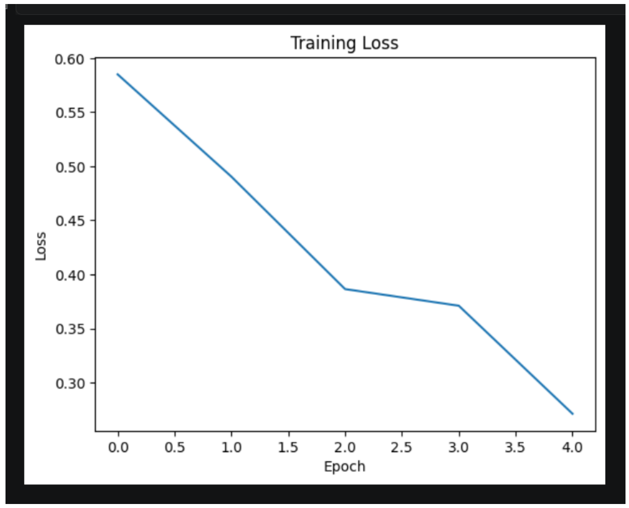

***Fig. 2.** Training Loss Curve Across Epochs*

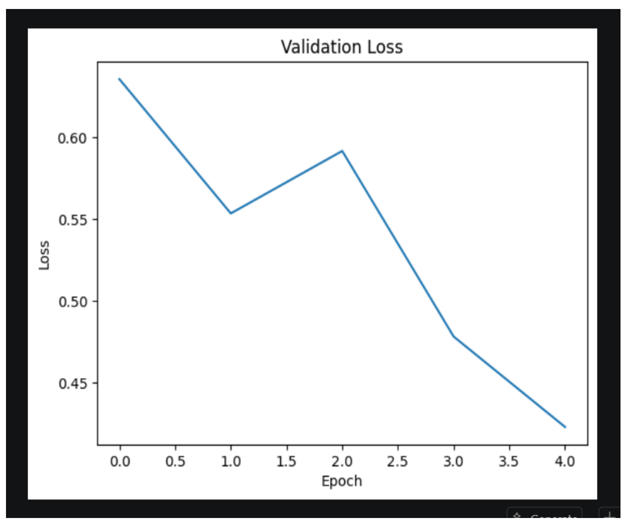

***Fig. 3.** Validation Loss Curve Across Epochs*

Research Gaps and Future Scope

The experimental evaluation includes graphical analysis of model performance. Training and validation accuracy curves demonstrate the learning progression of the model across epochs. Similarly, loss curves illustrate the optimization process and convergence behavior.

These graphs indicate that the model achieves stable performance without significant overfitting. The close alignment between training and validation metrics confirms the effectiveness of regularization techniques such as dropout.

Conclusion

This paper presented an end-to-end deep learning-based system for automated aircraft damage classification and caption generation. The proposed approach integrates image classification techniques with vision-language modeling to not only detect and categorize different types of aircraft damage but also generate meaningful textual descriptions indicating the location and severity of the damage.

The system leverages transfer learning using pretrained convolutional neural networks for feature extraction, combined with sequence modeling techniques for caption generation. Experimental results demonstrate that the model is capable of accurately identifying common damage types such as dents, cracks, and scratches, while also producing contextually relevant captions. This significantly enhances interpretability compared to traditional classification-only systems.

The proposed solution addresses key limitations of existing approaches by providing a unified framework that combines detection and descriptive analysis, thereby reducing manual inspection effort and improving decision-making efficiency in aircraft maintenance processes. Although the system performs well on available datasets, its performance may be affected by limited data diversity and variations in image quality.

Future work can focus on expanding the dataset, incorporating advanced vision-language models, and improving robustness under real-world conditions. Overall, the project demonstrates the potential of deep learning in automating aircraft inspection and contributes toward safer and more efficient aviation maintenance systems.

Acknowledgement

We would like to express our deepest gratitude to our respected project guide, Dr. Pallavi Patil, for her constant guidance, encouragement, and insightful suggestions throughout the course of this project. Her expertise in the field of Artificial Intelligence and Machine Learning, along with her constructive feedback at every stage, helped us refine our ideas and successfully complete this work.

We are sincerely thankful to the Department of Artificial Intelligence and Machine Learning at Genba Sopanrao Moze College of Engineering, Balewadi, Pune, for providing us with the necessary infrastructure, technical resources, and a supportive academic environment that enabled us to carry out this project effectively.

We would also like to extend our appreciation to all the faculty members of the department for their valuable inputs, motivation, and academic support during the development of this project. Their teachings laid the foundation required to understand and implement advanced machine learning concepts.

We are grateful to our friends and classmates for their cooperation, constructive suggestions, and moral support, which greatly contributed to the successful completion of this work.

We also acknowledge the invaluable contribution of the open-source community. Libraries and frameworks such as TensorFlow, Keras, NumPy, Pandas, and OpenCV played a crucial role in the design and implementation of our system. Additionally, we thank the contributors of publicly available datasets that facilitated experimentation and evaluation of our model.

Finally, we express our heartfelt gratitude to our families for their unwavering support, encouragement, and understanding throughout the duration of this project.

References

1. Uzun, M. R. Tolun, et al., "Damage Detection in Aircraft Engine Borescope Inspection Using Deep Learning," *Neural Computing and Applications*, 2025.
2. R. E. Donatus et al., "Automated Aircraft Structural Defect Detection Using Deep Learning and Computer Vision," *Mekatronika Journal*, 2025.
3. J. Huang et al., "Surface Damage Detection and Localization System for Fighter Aircraft," *arXiv preprint arXiv:2506.10505*, 2025.
4. "Advances in Aircraft Skin Defect Detection Using Computer Vision," *Aerospace*, vol. 12, no. 4, MDPI, 2024.
5. "Deep Learning-Based Damage Recognition Techniques for Aero-Engine Surface Images," *ACM Digital Library*, 2025.
6. W. Shao et al., "A Damage Identification Method for Aviation Structure Integrating Deep Learning," *Ultrasonics*, 2025.
7. "Spectral Feature Extraction and Ensemble Learning for Aircraft Surface Damage Detection," *PMC*, 2025.
8. B. Fan and X. Song, "A Review of Deep Learning Based Object Detection Algorithms," *Journal of Engineering Research*,

2024.

9. A. Nanyonga et al., “Deep Learning Architectures for Textual Damage Level Classification,” *arXiv preprint*, 2025.

10. S. Sanami and A. G. Aghdam, “Aero-Engines Anomaly Detection Using Fisher Autoencoder,” *arXiv preprint*, 2025.

11. O. Vinyals et al., “Show and Tell: A Neural Image Caption Generator,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

12. B. Radford et al., “Learning Transferable Visual Models From Natural Language Supervision,” *arXiv preprint (CLIP)*, 2021.

13. J. Li et al., “BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation,” *arXiv preprint*, 2022.