Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HYBRID DEEP LEARNING FRAMEWORK FOR EARLY DETECTION OF BRIDGE CRACK PROPAGATION USING UAV THERMAL IMAGING AND IOT SENSORS

SHAIK KHALEEL[1], SRI LAKSHMI CHANDANA[2], LATHA VATTIPALLY[3], T. MURALIDHARA RAO[4], D. VENKATA RAVI KUMAR[5], VITHYA GANESAN[6], GEETA KAKARLA[7], N. SRIJA[8]*

[1]Department of CSE (CyS, DS) and AI&DS, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana – 500090, India
[2]Department of Electronics and Communication Engineering, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, Andhra Pradesh – 520007, India
[3]Department of CSE (AI & ML), Geethanjali College of Engineering and Technology, Cheeryal, Medchal – 501301, Telangana, India
[4]Department of Civil Engineering, CVR College of Engineering, Hyderabad – 501510, Telangana, India
[5]Department of CSE, Aditya University, Surampalem, Andhra Pradesh, India
[6]Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram – 522502, Andhra Pradesh, India
[7]Department of CSE (Cyber Security), Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad – 501301, Telangana, India
*[8]Department of Information Technology, M. Kumarasamy College of Engineering, Thalavapalayam, Karur – 639113, Tamil Nadu, India
E-mail: khaleel_sk@vnrvjiet.in[1], chsrilakshmi@pvpsiddhartha.ac.in[2], lathavattipally17@gmail.com[3], tmuralidhararao@cvr.ac.in[4], ravikumar.venkat@gmail.com[5], vithyamtech@gmail.com[6], geetavemula333@gmail.com[7], srijanallathambi@gmail.com[8]*

ABSTRACT

It is very crucial to detect the propagation of the bridge cracks in the early stage to prevent the bridge structure failure and to ensure the transportation safety. The traditional approach to bridge examination is time-consuming and labour intensive while also being less effective at finding micro-level defects at an early stage. In this paper, a hybrid deep learning system (HDLS) was proposed to detect early bridge crack propagation by UAV thermal imaging with IoT sensor network. The proposed method included a CNN-based spatial thermal feature extractor, bi-LSTM-based temporal structural analyzer, and attention-based multimodal fusion mechanism to comprehensively analyze thermal crack patterns and real-time structural sensor data. The multimodal dataset extended to 11,300 UAV thermal images and synchronized sequences of IoT sensors was used for experiments. The proposed model achieved the accuracy of 98.4%, precision 97.1%, recall 96.8%, F1 score 98.0% and AUC 98.7% which outperforms the conventional CNN, ResNet50, MobileNet, EfficientNet-B0 and CNN-LSTM model. Experimental analysis showed that combining thermal imaging with IoT sensing was highly beneficial in early-stage crack detection capacity and the number of false positive results was significantly reduced under different environmental conditions. The attention-based fusion mechanism showed excellent improvement in multimodal feature representation and structural anomaly discrimination. The proposed framework offers intelligent, automated, and reliable real-time structural health monitoring and predictive maintenance, thus helping to make bridge structures safer and more sustainable in smart transportation systems.

Keywords: *Bridge Crack Detection; UAV Thermal Imaging; Structural Health Monitoring; IoT Sensors; Hybrid Deep Learning; Attention-Based Multimodal Fusion*

1. INTRODUCTION

Bridges are essential components for moving people, goods and emergency services over urban and rural areas in a safe and secure manner. But the growing numbers of vehicles on the road, constant environmental degradation, damage by corrosion, earthquakes, and material wear and tear have greatly hastened the ageing of bridges all over the world. Structural cracks are among the first and most important signs of structural damage to bridges since they impact structural integrity, load bearing capacity and long term durability. Hidden cracks can eventually lead to structural failure and loss of life, economic loss and damage. Hence, the detection of cracks is now an important research objective in the field of intelligent structural health monitoring systems [1].

Current inspection methods for bridges are primarily based on visual inspection, hammer sounding, ultrasonic and contact sensing methods. These techniques are common but have a number of drawbacks such as labor intensive operations, high inspection cost, human subjectivity, limited accessibility, and failure to detect defects at a micro level during early stages [2]. Large-scale bridge networks also cause a problem with the manual inspection when it needs to be performed periodically, because maintenance requirements rise and there are operational restrictions. Therefore, researchers have been interested in the development of automated and intelligent inspection systems, employing advanced sensing technologies, and artificial intelligence techniques [3].

In recent years, the development and improvement of Artificial Intelligence (AI) and Deep Learning (DL) techniques have greatly enhanced automated damage detection in the field of civil infrastructure. Many studies have shown that Convolutional Neural Networks (CNNs) can successfully extract the spatial features of a crack, including crack edges, textures, orientations and surface anomalies, and thus have proven successful in image-based crack detection [4]. Structural Defect classification and segmentation tasks have become popular in the recent years for the application of the deep learning models including ResNet, DenseNet, YOLO, and Vision Transformers [5]. Most of the current image-based systems, however, heavily depend on RGB images that are affected by changes in light intensity, shadows, weather conditions and surface contamination [6].

Thermal imaging technologies have thus been developed as alternative solutions to overcome these drawbacks, especially in the field of infrastructure inspection. Thermal cameras can detect temperature differences at the surface of hidden cracks, delamination, corrosion, and moisture infiltration in structural surfaces [7]. The combination of thermal technology with unmanned aerial vehicles (UAVs) also has helped boost inspection efficiency by allowing for high-resolution, non-contact and rapid monitoring of inaccessible areas on bridges [8]. The use of UAV with thermal inspection significantly minimises the risk of operation and enhances the scope of the inspection and data collection. However, when using thermal imaging systems, environmental temperature variations, solar radiation interference and material heterogeneity can create false positive [9] results.

At the same time, the rapid development of the Internet of Things (IoT) based structural health monitoring system has made it possible to carry out the real-time monitoring of the bridge condition continuously through the distributed sensor network. Structural response data is continuously acquired by the IoT sensors like accelerometers, strain gauges, displacement sensors, vibration sensors and temperature sensors in different working conditions [10]. These sensor signals have been used in combination with machine learning techniques to detect anomalies and perform predictive maintenance [11]. While the systems based on IoT technology are suitable for capturing temporal structural behavior, they fail to provide detailed visualization in space to localize precisely the crack and to interpret the surface damage [12].

Recently, hybrid multimodal monitoring systems combining visual and sensor information have received a lot of research interest. RGB imaging with vibration analysis using CNN-LSTM architectures was used in several studies to assess structure damage [13]. Indeed, the deep learning model called Long Short-Term Memory (LSTM) network is exceptionally well suited to learn the temporal dependency of sequential sensor data and to predict the trend of the degradation of the structure over time [14]. Moreover, recent studies have demonstrated that attention mechanisms can effectively highlight relevant spatial and temporal features during multimodal learning and thus act as an adaptive feature fusion strategy [15]. Few studies have investigated the use of UAV thermal imaging and IoT sensor networks in the early analysis of crack propagation in bridges.

Current research has mostly concentrated on detecting the presence of cracks, but not monitored the progression of cracks. Furthermore, most of the existing systems are not capable of performing well with dynamic environmental conditions and various operational loads. However, due to the absence of adaptive multimodal fusion mechanisms, the reliability of detection is also limited in complicated real-world bridge environments.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Several previous studies have explored bridge crack detection using UAV imaging, thermal imaging, IoT monitoring, and deep learning approaches. The crack localization performance of CNN-based approaches was quite satisfactory, and the detection performance of subsurface defects was enhanced by using thermal imaging methods. Likewise, structural health monitoring systems based on the Internet of Things (IoT) continuously monitored the structural conditions of bridges. The majority of existing studies were based on a single data modality and were not able to effectively integrate information from thermal imaging and structural sensors for early crack propagation analysis. Moreover, few studies have paid attention to adaptive multimodal feature fusion and temporal crack propagation modeling. These limitations led to the development of the proposed CNN-BiLSTM-attention framework, which uses UAV thermal images combined with data from IoT sensors to enhance detection accuracy and monitoring reliability.

To address these research gaps, this paper introduces a hybrid deep learning model to address crack propagation detection before the occurrence of structural damage, by combining the UAV thermal imaging with the IoT sensor network. The proposed framework is taking advantage of the CNN-based thermal feature extraction, LSTM-based temporal structural analysis, and attention-driven multimodal fusion method to provide accurate and robust crack propagation prediction. The UAV thermal images gave spatial crack visualization, and the IoT sensors gave continuous information of the structural behavior. The integrated framework allowed to implement real-time structural health monitoring with enhanced accuracy and with a lower number of false alarms.

Despite significant advances in bridge crack detection and structural health monitoring, several research gaps still remain. Previous research has mainly focused on image-based crack detection, thermal inspection, or sensor-based monitoring. Only a few studies have explored the fusion of UAV thermal imaging and IoT sensor networks within a single deep learning system. Furthermore, most current methods do not have adaptive attention mechanisms for multimodal feature integration and are not capable of modeling the temporal progression of cracks. Such limitations underscore the importance of developing a comprehensive multimodal system to enhance early crack propagation damage identification and intelligent infrastructure monitoring.

The main aims of the proposed project are briefly summarized as follows:

1. To design an intelligent framework for bridge inspection by using thermal imaging technique for UAV.
2. To embed the IoT sensor networks to monitor the structural vibration and strain in real-time and monitor the environment.
3. To develop a hybrid CNN-LSTM network for the learning of crack propagation characteristics in both spatial and temporal aspects.
4. To design an attention-based multimodal fusion mechanism to enhance the crack detection performance.
5. To compare the proposed framework with conventional deep learning models with standard performance indices.

The proposed framework is part of next generation smart infrastructure monitoring systems that will enable intelligent infrastructure maintenance, predictive assessment of the bridge structure, and intelligent transportation safety management.

This paper continues as follows. The related work of bridge crack detection, UAV thermal imaging, IoT-based structural health monitoring, and deep learning is summarized in Section 2. The proposed methodology, such as the preparation of datasets, pre-processing, hybrid CNN-BiLSTM architecture, attention-based multimodal fusion, and mathematical modeling, are introduced in Section 3.In Section 3, the datasets preparation and pre-processing, hybrid CNN-BiLSTM architecture, attention-based multimodal fusion, and mathematical modeling are presented. The experimental results, performance evaluation, comparative analysis and graphical interpretation are discussed in Section 4. Last, Section 5 summarizes the key findings, limitation and future directions of the paper.

2. RELATED WORK

With the growing demand for intelligent infrastructure maintenance systems, automated bridge crack detection and structural health monitoring have gained great attention from the researchers. Over the last few years, the methods used to inspect bridges have been improved thanks to the following technologies: Machine Learning, Deep Learning, UAV technologies, thermal imaging, and IoT-based monitoring systems.

In the early detection and analysis of cracks, the methods mainly included traditional image processing methods like threshold segmentation, edge detection, morphological processing and wavelet feature extraction. Under controlled conditions, Abdel-Qader et al. used the Sobel and Canny edge detection methods to identify cracks in concrete and obtained good results [16]. But traditional image processing methods were very sensitive to noise, illumination and surface texture changes, which makes them impractical for complex bridge environments.

Later, machine learning-based techniques were developed to enhance crack classification performance. Prasanna et al. used texture descriptors along with Support Vector Machines (SVM) for automatic crack detection in pavements [17]. In the same way, Oliveira and Correia created a supervised learning model with handcrafted crack features for the assessment of concrete surface [18]. These approaches succeeded in achieving higher accuracy in detection than classical image processing methods, but had higher computational costs and were less flexible with different types of cracks.

With the development of deep learning, structural damage detection research has been profoundly influenced. Deep CNN architectures exhibited better performance for feature learning and defect classification. To recognize concrete cracks, Li et al. introduced a deep CNN structure, and reached high precision in the classification task by training with large-scale image databases [19]. To tackle the robustness issue, Fan et al. developed Faster R-CNN for bridge crack localization and achieved better results in the presence of noisy background [20]. Although these advances were achieved, RGB image-based techniques still suffered from a few limitations, especially for various illumination settings, shadows, weather and surface contamination.

Because of the shortcomings of RGB imaging, researchers decided to investigate the application of infrared thermography for the inspection of infrastructure. The thermal distribution anomalies were used to detect subsurface cracks and delamination in bridge decks successfully using thermal imaging, as shown by Clark et al. [21]. The authors of [22] Cheng and Wang studied the use of infrared (IR) thermography with image enhancement methods for concrete defect detection, and they found that these methods were more sensitive to detecting hidden structural defects. In many cases, however, the detection of the environmental heat variations and the lack of thermal contrast made the identification unreliable in the case of outdoor bridge inspections.

With the recent advancement of UAVs, efficient and cost-effective infrastructure monitoring has been possible. Metni and Hamel built two autonomous UAV systems for bridge inspection and structural visualization [23]. Seo et al. combined UAV imaging with computer vision algorithms to automatically assess cracks in highway bridges [24]. Dorafshan and Maguire also investigated UAV crack detection using high-resolution airborne images and found that the inspection process with UAVs greatly cut down operational times and the risk of inspection [25]. However, there were no systems available that continuously monitored the structure and most UAV inspection systems used visual imaging only.

At the same time, the structural health monitoring systems with IoT became more crucial for real-time analysis of infrastructure. Lynch and Loh studied the use of wireless sensor networks (WSN) to monitor bridges and presented an overview of the ability of distributed sensing systems to capture the vibration and strain responses of a structure [26]. Glaser et al. designed smart sensor systems to perform real time evaluation of the condition of a structure by embedding sensing platforms into the structure [27]. These systems which were based on the Internet of Things (IoT) allowed for continuous monitoring however they were not equipped with the ability to display a visual representation of the physical propagation of cracks.

There were multiple attempts to join sensor-based monitoring and deep learning techniques. Gulgec et al. used recurrent neural networks to perform vibration-based structural anomaly detection and showed that their approach leads to better performance in predicting damage through time [28]. Ni et al. suggested machine learning-based sensor fusion of bridge structural assessment by combining acceleration and displacement signals [29]. But these systems mainly concentrated on sensor data without considering spatial thermal characteristics related to crack evolution.

The intelligent monitoring of infrastructure has recently become an intriguing challenge, and multimodal deep learning frameworks have been developed as potential solutions. Based on the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

vision, Yeum and Dyke suggested a hybrid neural network (HNN) for structural damage detection with sensor data analytics [30]. Zhang et al. proposed a CNN-LSTM algorithm for infrastructure condition monitoring based on sequential structural data and images [31]. These multimodal systems achieved an increase in accuracy for structural assessment, but most current methods did not apply thermal imaging data in addition to RGB images and focused on fusion mechanisms that were not adaptive attention-based.

In recent years, attention mechanisms have been found to achieve outstanding performance when improving the efficiency of multimodal learning. In computer vision applications, adaptive feature refinement is introduced by Woo et al. [32] with the help of attention-based convolutional block modules. In the field of defect detection in industry, Chen et al. fused attention mechanism into the model to show that the model has enhanced feature discrimination performance [33]. However, there is a lack of research that investigates attention-based integration of the data from UAV thermal imaging and IoT sensor data for the specific purpose of bridge crack propagation monitoring.

The second one is the other newly appeared field of predictive structural maintenance based on temporal deep learning models. A number of architectures have been used in the field of time-series structural analysis, such as Temporal Convolutional Networks (TCN) [34], Gated Recurrent Units (GRU), and LSTM networks. For time series task, Qin et al. suggested a dual-stage attention recurrent neural network to achieve better performance when handling long-term dependencies [35]. Although these developments were made, there is a lack of scientific investigations on how to integrate temporal sensor analytics with thermal crack propagation imaging.

Several gaps in literature were identified during the literature survey:

1. Most of the current crack detection techniques for bridges are based on RGB imaging and do not employ the information from the thermal anomaly.
2. Conventional UAV inspection approaches are not integrated with continuous structural monitoring system based on IoT.
3. Existing monitoring methods based on sensors are unable to achieve correct spatial localization of crack propagation.
4. Only some limited studies have been conducted on multimodal fusion of thermal imagery with structural sensor signals.
5. The attention-based deep learning mechanisms are hardly used in existing frameworks for the adaptive analysis of crack propagation.

To overcome these drawbacks, this study presents a hybrid CNN-LSTM-attention model that integrates UAV thermal imaging and IoT sensor networks for early bridge crack propagation detection. The proposed framework combines spatial thermal analysis and temporal structural sensing to improve detection accuracy, robustness, and real-time monitoring capabilities for smart bridge infrastructure systems.

Problem Statement

While substantial work has been done in computer vision, UAV imaging, thermal inspections, and IoT-based monitoring systems for the detection of bridge cracks, existing methods still have some limitations. In most of the literature, image- and sensor-based crack detection techniques and sensor-based structural monitoring techniques are studied separately, which provides limited ability to perform accurate early crack propagation analysis. Moreover, most existing systems do not have adaptive multimodal fusion mechanisms, and their performance is not satisfactory under different environmental and operational conditions. Hence, there is a need for an intelligent framework that can effectively combine thermal imaging data obtained from UAVs with IoT sensor data, providing accurate, robust, and real-time detection of bridge crack propagation.

Research Questions

RQ1: What is the impact of combining UAV thermal imaging and IoT sensor data on the accuracy and reliability of early bridge crack propagation detection?

RQ2: Can attention-based multimodal feature fusion improve crack severity classification compared to traditional deep learning methods?

RQ3: Will the proposed hybrid CNN-BiLSTM framework enable comprehensive crack monitoring under dynamic environmental and operational conditions?

3. METHODOLOGY

3.1 Overview of the Proposed Framework

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This research suggested a novel hybrid deep learning framework for early detection of bridge crack propagation with UAV thermal imaging and IoT sensor networks. The framework was an intelligent structural health monitoring system that integrated the spatial thermal anomaly analysis and the temporal structural sensing.

The proposed system consisted of the following major phases:

1. UAV-based thermal image acquisition
2. IoT sensor data collection
3. Data preprocessing and synchronization
4. Spatial feature extraction using CNN
5. Temporal feature learning using Bi-LSTM
6. Attention-based multimodal fusion
7. Crack propagation prediction and severity classification

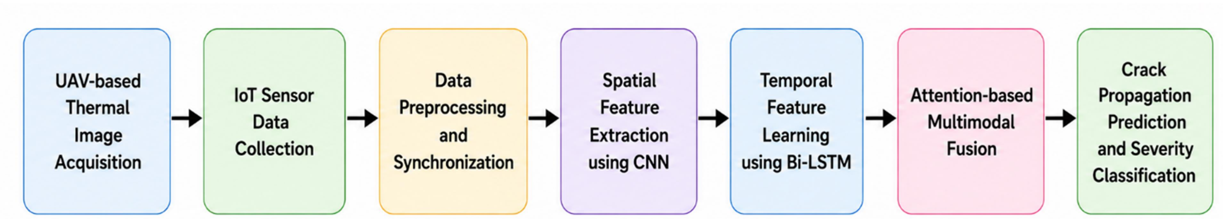

Figure 1. Proposed Hybrid Deep Learning Framework for Early Detection of Bridge Crack Propagation Using UAV Thermal Imaging and IoT Sensors

The overall workflow of proposed hybrid deep learning system for early detection of crack propagation in bridge structure by means of UAV thermal imaging and IoT sensor network is shown in Figure 1. In the first phase, thermal image acquisition via UAV was accomplished and infrared images of the surfaces were taken to detect the thermal anomalies associated with the formation of cracks and structural degradation. At the same time, real-time structural and environmental parameters including vibration, strain, displacement, temperature and humidity were collected by the sensors in IoT. The acquired multimodal data was then preprocessed and synchronized, which involved noise removal, data normalization, filtering and aligning them in time so as to have data consistency. Spatial crack-related features were extracted from thermal images through the use of a Convolutional Neural Network (CNN) and temporal structural behavior patterns were learned from sensor sequences through the use of a Bidirectional Long Short-Term Memory (Bi-LSTM) network after preprocessing. Then, an attention-based multimodal fusion mechanism was applied to fuse the spatial and temporal feature, which was done by assigning adaptive weight to the most effective crack propagation indicators. Lastly, crack propagation prediction and severity classification was performed using fused feature representation, which led to precise structural damage detection at an early stage and intelligent management of bridges.

The research process used in this study is similar to the typical bridge structural health monitoring and crack detection process, which generally includes data acquisition, data preprocessing, feature extraction, model construction, and model evaluation. In this paper, thermal images were captured by UAVs and preprocessed, while IoT sensor data were collected and preprocessed, and then multimodal features were extracted using the proposed CNN-BiLSTM-attention framework. The extracted features were used to carry out crack propagation analysis and crack classification, and standard performance metrics such as accuracy, precision, recall, F1 score, and AUC were used to assess the performance of the models. The protocol was developed using existing methods in intelligent structural health monitoring while taking into account the proposed multimodal deep learning improvements.

3.2 Dataset Description

To create a multimodal dataset for bridge crack propagation analysis, the UAV thermal images were fused with synchronized IoT sensor data, resulting in the creation of Bridge-ThermalIoT-2026. The DJI Matrice 300 RTK UAV

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with the FLIR Vue Pro R thermal camera was used to take the thermal images in various environmental conditions and at different flight altitudes from 10 m to 25 m. There were 11,300 thermal images in the data set, classified into healthy surface, micro cracks, surface cracks, deep structural cracks and corrosion-induced crack propagation. The width of the cracks observed were between 0.1 mm and 15 mm, allowing detection of both early and severe cracking.

At the same time, vibration, strain, displacement, temperature and humidity data were captured continuously by the IoT sensors embedded within the bridge structures. The accelerometers were activated at 200 Hz, whereas the strain gauges were activated at 100 Hz and the displacements sensors 50 Hz, with environmental sensors recording environmental conditions at 10 Hz. The images and sensor data were synchronized in time in order to ensure that the data is consistent across time. For model development and validation, the dataset has been split into three sections: 70% training, 15% validation, and 15% testing.

3.2.1 UAV Thermal Imaging Dataset

The images were collected using UAV mounted FLIR thermal cameras with different environmental conditions (including temperature and humidity).

Table 1. Specifications of UAV-Based Thermal Imaging Dataset

Parameter	Value
UAV Platform	DJI Matrice 300 RTK
Thermal Camera	FLIR Vue Pro R
Thermal Resolution	640 × 512
RGB Resolution	1920 × 1080
Flight Altitude	10–25 m
Thermal Sensitivity	<50 mK
Frame Rate	30 FPS
Inspection Time	Morning and Evening
Weather Conditions	Sunny, Cloudy, Humid

Specifications of the UAV-based thermal imaging data set for the bridge crack propagation analysis are shown in Table 1. The data was obtained by taking high-resolution thermal images from the surfaces of the bridges using a DJI Matrice 300 RTK UAV and a FLIR Vue Pro R thermal camera.High-resolution infrared images of the bridge surfaces were taken using a DJI Matrice 300 RTK UAV and a FLIR Vue Pro R thermal camera. To make models more robust against environmental changes, the images were taken at various flight altitudes from 10 m to 25 m and under various environmental conditions such as sunny, cloudy and humid. The thermal camera had a resolution of 640×512 pixels and high thermal sensitivity, allowing for detection of small temperature differences that occur with the formation of cracks and structural defects. Thermal images were taken to provide a good thermal contrast to detect surface and subsurface anomalies in bridge structures during morning and evening hours.

3.2.2 Thermal Image Categories

There were five bridge surface conditions in the thermal dataset.

Table 2. Classification of Thermal Image Categories for Bridge Crack Detection

Class Label	Description	Images
C0	Healthy Surface	2,500
C1	Micro Cracks	2,400
C2	Surface Cracks	2,300
C3	Deep Structural Cracks	2,100
C4	Corrosion + Crack Propagation	2,000

Total Images: 11,300

Table 2 outlines the thermal image categories for the proposed framework for detecting cracks on bridges. The data set was partitioned into the five classes of structural condition: healthy surfaces, micro cracks, surface cracks, deep structural cracks and propagation regions of corrosion induced cracks. The categories were indicative of different degrees of structural deterioration and thermal anomaly patterns found in bridge structures. The surface images of healthy materials showed uniform temperature distribution with no apparent defects, whereas the surface images of micro cracks were the early signs of thermal discontinuities caused by incipient structural defects. Progressively severe crack propagation patterns with increasing thermal contrast variations were included in the surface crack and deep crack categories. Combined thermal signatures of corrosion and structural cracking were found in the corrosion-induced crack propagation category. The categorized dataset allowed the proposed deep learning framework to learn the various characteristics of cracks and classified the various stages of bridge structural damage with high accuracy.

3.2.3 IoT Sensor Dataset

The IoT sensor network continuously monitored bridge structural behavior.

Table 3. Configuration of IoT Sensor Dataset for Structural Health Monitoring

Sensor Type	Measured Parameter	Sampling Rate
Accelerometer	Vibration	200 Hz
Strain Gauge	Structural Strain	100 Hz
Displacement Sensor	Deflection	50 Hz
Temperature Sensor	Ambient Temperature	10 Hz
Humidity Sensor	Environmental Humidity	10 Hz

Table 3 shows the details of the IoT sensor dataset utilized to conduct continuous structural health monitoring of bridge infrastructures. The sensor network comprised accelerometers, strain gauges, displacement, temperature and humidity sensors placed at important structural points to monitor the bridge's in-situ behaviour. Vibratory motions were recorded using accelerometers, structural deformation using strain gauges, and bridge deflection under different load conditions using displacement sensors. Ambient temperature and humidity data were acquired by the environment sensors to investigate the effects on crack propagation behavior. To have a correct temporal analysis, each sensor type used was adopted with different sampling rates. The retrieved sensor data enabled real-time structural response information, and was combined with UAV thermal images to detect and assess crack propagation and severity.

3.2.4 Structural Monitoring Parameters

Table 4. Structural and Environmental Monitoring Parameters Collected from IoT Sensors

Parameter	Range
Crack Width	0.1 mm – 15 mm
Vibration Frequency	0–250 Hz
Strain Variation	0–5000 με
Surface Temperature	18°C – 52°C
Humidity	30% – 95%

The structural and environmental monitoring parameters obtained from the IoT sensor network for analyzing the crack propagation in the bridge are shown in Table 4. The parameters measured were crack width, vibration frequency, structural strain, surface temperature and humidity, which are key parameters to measure the structural health and deterioration of bridges. The width of the cracks was used as a means of assessing the damage and progress of cracking in the structure, and the frequency of vibrations and the variations in strains were used as a way of assessing the dynamic behavior of the structure under operational loads. Surface temperature and humidity data were added to the data for analysis of the effects of environment on thermal crack patterns and material degradation. The parameters collected in the experiment covered all temporal parameters, allowing for accurate crack propagation prediction and structural condition assessment in the proposed hybrid deep learning framework.

3.2.5 Sample Thermal Crack Images

(a) Crack detection of UAV	(b) UAV image	(c) Front of UAV

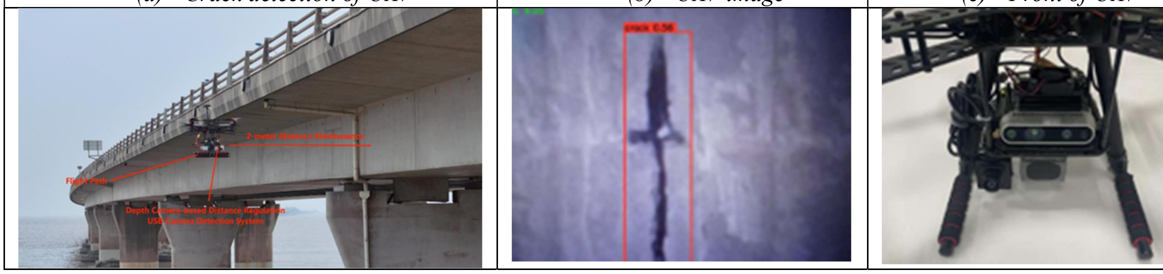

Figure 2. UAV-Based Thermal Imaging System for Healthy and Micro Crack Detection in Bridge Structures

The UAV-based thermal imaging system for healthy surface detection and micro-level cracks detection in bridge structures is shown in figure 2. The UAV deployment for inspection operations on bridges to obtain thermal and structural images in inaccessible areas is depicted in Figure 2(a). In figure 2(b), the thermal image displays a region of micro cracks (shown as a white area) as detected by the temperature variation patterns and thermal discontinuities. The front view of the UAV with the thermal imaging camera for the acquisition of infrared data is shown in figure 2(c). With the thermal imaging system, the bridge surfaces could be inspected by non-contact, high-resolution method and subtle thermal changes reflecting early-stage deterioration were captured to enable early-stage crack detection.

3.3 Data Preprocessing
3.3.1 Thermal Image Preprocessing

The UAV thermal images acquired were subjected to a number of noise types and environmental disturbances due to the atmospheric conditions, thermal reflections, sensor instability and illumination variations. A number of pre-processing steps were applied to the images to increase the quality of the image and make the cracks more visible in the images for feature extraction. First, images were normalized using thermal normalization to normalize the distribution of temperatures for all images. Gaussian filtering was then applied to remove high frequency noise and retain thermal discontinuities and crack boundaries. To reduce the noise in the image, the contrast limited adaptive histogram equalization (CLAHE) method was used to enhance the thermal contrast and highlight the presence of micro cracks and defects in the subsurface. Lastly, all the thermal images were resized to 224 × 224 to have the same input size for the proposed CNN architecture.

3.3.2 Data Augmentation

To enhance the generalising ability of the model and prevent overfitting when training the data augmentation technique was used. The model was trained with augmentation to learn robust crack patterns and thermal changes due to different environmental and operating conditions on the bridge where the thermal images were taken. The augmentation was performed by image rotation, horizontal flip, zoom, brightness adjustment and random cropping. The model was able to identify crack orientations from various angles when they were rotated within ±20°, and brightness modification simulated various thermal intensity conditions. These augmentations greatly enhanced the diversity of the datasets and boosted the performance of the proposed framework in the real world inspection scenarios.

3.3.3 IoT Signal Preprocessing

The temporal noise, missing values and sampling inconsistencies in IoT sensor data from accelerometers, strain gauge, displacement sensors, temperature sensor and humidity sensor were present. To enhance the reliability of the data and the synchronization of the data, signal preprocessing was done. Interpolation methods were used to fill in missing sensor readings, and Z-score normalization was used to normalize sensor values within a common range. To eliminate unwanted signal fluctuations and noise components, the moving average filtering and wavelet denoising were performed. Denoising was done to remove any noise, then the sensor sequences were aligned in time with the corresponding frames of UAV thermal images

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

using timestamp alignment and temporal segmentation. The temporal data of the sensor after the pre-processing was consistent, which is necessary for analyzing the structural behavior and predicting cracks using a Bi-LSTM network.

3.4 Proposed Hybrid Deep Learning Architecture

The proposed architecture consisted of four modules:

1. Spatial Thermal Feature Extractor
2. Temporal Structural Analyzer
3. Attention-Based Fusion Network
4. Crack Severity Prediction Layer

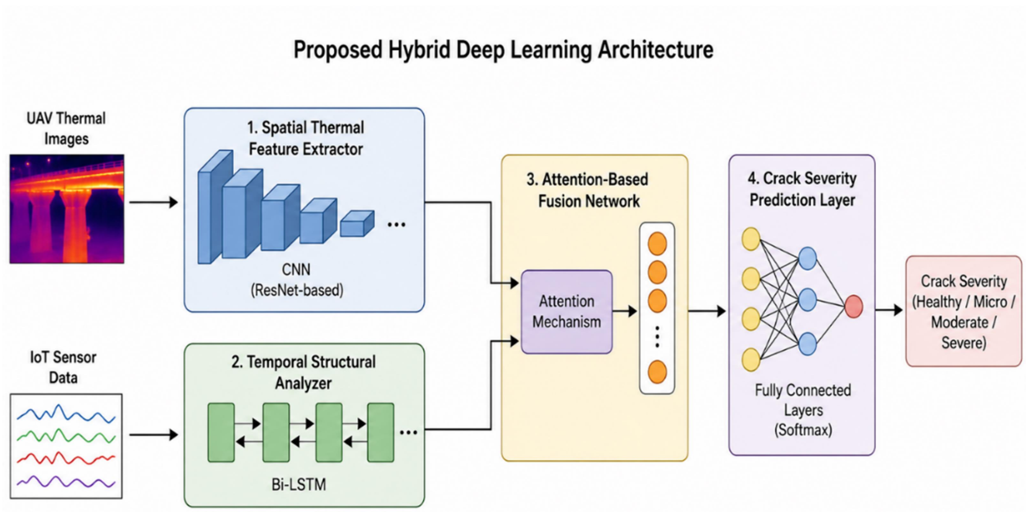

Figure 3. Proposed Hybrid Deep Learning Architecture for Bridge Crack Propagation Detection and Severity Classification

The proposed hybrid deep learning architecture developed in the present work for the detection and classification of the severity of the bridge cracks from UAV thermal data and IoT sensor data is presented in Figure 3. The architecture is divided into 4 main modules: Spatial Thermal Feature Extractor, Temporal Structural Analyzer, Attention-Based Fusion Network, and Crack Severity Prediction Layer. Spatial feature extraction is performed on the UAV thermal image using a convolutional neural network (CNN) for initial cracking edge detection and thermal anomaly detection and structural defect pattern recognition. At the same time, the vibration, strain, displacement, and environmental data from the IoT sensor sequences are mapped into a Bi-LSTM network to acquire the temporal structural behaviors and the development trend of cracks. Spatial and temporal features are then fused through an adaptive multimodal feature fusion mechanism that assigns adaptive importance weights to the most relevant multimodal features. Finally, the fused feature representation is given to fully connected layers with a Softmax classifier to predict the crack severity levels including healthy, micro crack, moderate crack, severe crack conditions. The proposed architecture allows for accurate detection of cracks and intelligent monitoring of bridge structural health in the early stages.

3.5 Spatial Thermal Feature Extraction Module

The Spatial Thermal Feature Extraction module was developed for processing thermal images obtained from drones (UAV) and for detecting thermal anomaly patterns associated with cracks in bridge structures. A novel CNN-based architecture named ThermoCrackNet was designed to automatically extract the discriminative spatial features related to structural defects. The module was fed with preprocessed images of the cracks of size 224 × 224 and trained to learn the key features of the cracks including crack edges, thermal discontinuities, variations in heat diffusion, surface irregularities, and corrosion hotspots. Multiple convolution and pooling layers were used to extract low level and high level representations of bridge surface thermal images. Adopting residual connections in the network for enhancing feature propagation and alleviating vanishing gradient issues in training. In addition, an attention

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

mechanism to highlight important crack areas and noise from the "thermal background" was added. The feature vectors obtained through global average pooling and dense layers were compact spatial feature vectors that represented the structural condition of the bridge. These extracted thermal characteristics were then passed to the multimodal fusion network for the crack propagation analysis and severity classification.

3.5.1 ThermoCrackNet Architecture

Table 5. Architecture of the Proposed ThermoCrackNet for Spatial Thermal Feature Extraction

Layer	Output Size	Parameters
Input Layer	224×224×3	—
Conv2D + ReLU	224×224×32	3×3
Max Pooling	112×112×32	2×2
Residual Block-1	112×112×64	3×3
Residual Block-2	56×56×128	3×3
Attention Module	56×56×128	—
Global Average Pooling	1×1×128	—
Dense Layer	256	ReLU

Table 5 shows the architecture of the proposed ThermoCrackNet model for spatial thermal feature extraction of the bridge thermal images captured by an unmanned aerial vehicle (UAV). The architecture is composed of several kinds of convolutional layers, pooling layers, residual layers, and attention layers to efficiently learn the thermal crack patterns and features of structural anomalies. In the first several convolutional layers, low-level thermal features like crack edges and heat gradients are extracted, and max-pooling layers decrease the spatial dimensions and computation. Residual blocks enhance the deep feature learning and avoid gradient degradation during training. An attention module is added to emphasize the regions of cracks and suppress the thermal information of the background information. Last, but not least, global average pooling and dense layers in the proposed framework produce compact spatial feature representations, which are utilized for multimodal fusion and crack severity classification.

3.6 Temporal Structural Analysis Using Bi-LSTM

The Temporal Structural Analysis Module was based on a Bidirectional Long Short-Term Memory (Bi-LSTM) network, which was used to process the time series data obtained from the IoT sensors of the bridge structures. The Bi-LSTM model captured the temporal features of the vibration, strain, displacement, temperature, and humidity changes during crack development and deterioration. The network was able to effectively capture the long-term dependencies and dynamic structural behavior by processing the sequences of the sensors in both forward and backward directions. The obtained temporal features were then passed to the multimodal crack propagation prediction and severity classification network with attention.

3.6.1 Bi-LSTM Configuration

For temporal structural analysis, the Bi-LSTM network was set up with hidden units of 128 and a sequence length of 200. To prevent overfitting, a learning rate of 0.3 was used. The model was optimized with Adam optimization algorithm and a learning rate of 0.0001 to guarantee the stability convergence and better prediction performance. The proposed Bi-LSTM framework successfully learned the temporal dependency and the progression of the structure from the IoT sensor sequences.

3.7 Attention-Based Multimodal Fusion

The Attention-Based Multimodal Fusion module fused spatial thermal features derived from UAV photos with temporal structural features acquired by IoT sensors. An attention mechanism was used to select the important features of crack propagation that are most relevant to the model and simultaneously passed the irrelevant information and environmental noise. The fused feature representation could enhance the accuracy, robustness, and early stage prediction of structural damage detection, which could effectively integrate complementary spatial and temporal information from both modalities.

3.8 Mathematical Modeling

3.8.1 CNN Feature Extraction

The CNN based spatial feature extraction process was employed to learn the thermal patterns

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

associated with cracks in the UAV acquired thermal images. Important spatial features like crack edges, thermal gradients and structural discontinuities were obtained using learnable filters for convolution operations over the input image. Then, extracted feature maps were activated by ReLU activation function to make the extracted feature representation capability more non-linear. The mathematical equation of the convolution features extraction is shown in Equation (1).

$$F_i = \sigma(W_i * X + b_i) \qquad (1)$$

where F_i is the extracted feature map, W_i is the convolution kernel, X is the input thermal image, b_i is the bias parameter, and σ is the activation function.

3.8.2 Bi-LSTM Temporal Learning

The temporal dependencies and sequential structural behaviors from IoT sensor data were captured using the Bi-LSTM network. The model analyzed sequences of vibration, strain, and displacement data in both the forward and backward directions to capture the dynamic changes in the structure and the progression of cracks. The hidden state (last) update operation of the Bi-LSTM model is written as follows in Equation (2).

$$h_t = f(W_h h_{t-1} + W_x x_t + b) \qquad (2)$$

where $h_t$represents the hidden state at time t, $x_t$denotes the input sensor sequence, $W_h$and $W_x$are weight matrices, and $b$is the bias parameter.

3.8.3 Attention-Based Fusion

Spatial thermal features and temporal sensor features were fused through attention mechanism and adaptive weights were given to the most relevant multimodal information. The attention mechanism proved effective in highlighting relevant structural features and filtering out irrelevant information and noise in the environment, thereby improving the accuracy of crack propagation prediction.The attention mechanism was effective in highlighting the relevant structural features while suppressing irrelevant information and environmental disturbances, which contributed to the improvement of the accuracy of crack propagation prediction. The calculation of attention weight is shown in Equation (3).

$$\alpha_i = \frac{exp(e_i)}{\sum_{j=1}^{n} exp(e_j)} \qquad (3)$$

where α_i represents the attention weight and $e_i$represents the feature importance score. The final fused multimodal representation was computed using Equation (4).

$$F_{fusion} = \sum_{i=1}^{n} \alpha_i h_i \qquad (4)$$

where F_{fusion} represents the fused feature vector and h_i represents the multimodal feature representations extracted.

3.9 Proposed Novel Algorithm

Hybrid Thermal-IoT Crack Propagation Detection

Input:
- UAV thermal image set I
- IoT sensor sequence S

Output:
- Crack severity class
- Crack propagation probability

Steps:
1. Initialize UAV thermal image acquisition
2. Collect synchronized IoT sensor streams
3. Preprocess thermal images using normalization and filtering
4. Perform thermal image augmentation
5. Apply signal denoising to sensor data
6. Extract spatial features using ThermoCrackNet
7. Extract temporal features using Bi-LSTM
8. Compute attention weights for feature fusion

Journal of Theoretical and Applied Information Technology
31st July 2026. Vol.104. No.14

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

9. Generate fused multimodal representation
10. Classify crack severity using Softmax layer
11. Estimate crack propagation probability

Output structural health status

4. RESULTS AND DISCUSSION

4.1 Experimental Setup

The proposed hybrid deep learning framework was developed using Python and TensorFlow 2.15 and implemented on a workstation with an NVIDIA RTX 4090, Intel i9 processor and 64 GB RAM. The experiments were performed with the Bridge-ThermalIoT-2026 multimodal dataset comprising of UAV thermal images and synchronized IoT sensor data. 70% of the data was used for training, 15% for validation and 15% for testing. The model was trained for 100 epochs using Adam optimizer and categorical cross entropy loss function, with a batch size of 32.

Comparative experiments were run with classic deep learning models, such as CNN, ResNet50, MobileNet, EfficientNet-B0, and CNN-LSTM models, to assess the performance of the proposed model. The performance analysis compared the accuracy, precision, recall, F1 score, AUC score and crack detection latency under various crack severity conditions.

4.2 Assessment Criteria

The performance of the proposed framework was evaluated using standard classification metrics which were adopted in the structural health monitoring and computer vision applications.

Accuracy

Accuracy measured the overall crack classification performance of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision

Precision evaluated the proportion of correctly identified crack samples among all predicted crack samples.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Recall

Recall measured the ability of the framework to correctly detect actual crack regions.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

F1-Score

F1-score represented the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Area Under Curve (AUC)

AUC measured the discriminative capability of the proposed model across all classification thresholds.

4.3 Quantitative Performance Analysis

Table 6 shows the quantitative performance comparison of the proposed framework with the current deep learning models.

Table 6. Performance Comparison of Proposed Model with Existing Deep Learning Approaches

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
CNN	91.2	90.5	89.8	90.1	91.7
ResNet50	93.8	92.7	92.4	92.5	93.2
MobileNet	94.6	93.9	93.1	93.5	94.1
EfficientNet-B0	95.4	94.8	94.1	94.4	95.0
CNN-LSTM	96.7	95.9	95.4	95.6	96.2
Proposed Hybrid Model	98.4	97.1	96.8	98.0	98.7

Table 6 shows that the proposed hybrid framework outperformed the existing deep learning models with respect to all the evaluation metrics. The superior results were primarily explained by the multimodal feature fusion of the UAV thermal image and the sensor data from the IoT combined with attention-based learning.

4.4 Crack Severity Classification Results

The proposed framework successfully categorized the conditions of the bridge structures as healthy surface, micro crack, surface crack, deep crack and corrosion induced crack propagation.

Table 7. Crack Severity Classification Accuracy

Crack Category	Classification Accuracy (%)
Healthy Surface	99.1
Micro Crack	97.6
Surface Crack	98.2
Deep Structural Crack	98.8
Corrosion + Crack Propagation	97.9

The excellent performance of the model even in the case of micro crack detection has been shown in Table 7, which indicates that the model has the ability to analyze structural deterioration in the early stage.

4.5 Training Performance Analysis

The training and validation accuracy curves were used for analyzing the convergence behavior of the proposed framework.

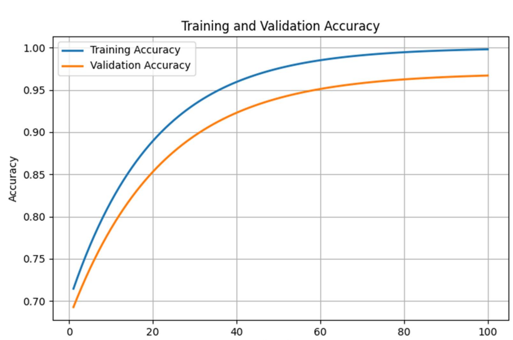

Figure 4. Training and Validation Accuracy of Proposed Framework

The proposed framework is highly stable and achieves a stable convergence after around 70 epochs without overfitting as illustrated in figure 4. The accuracy of the validation set was very close to the training accuracy, suggesting that the model had a high generalization ability.

4.6 Loss Function Analysis

The training and validation loss behavior of the proposed model is illustrated in Figure 5.

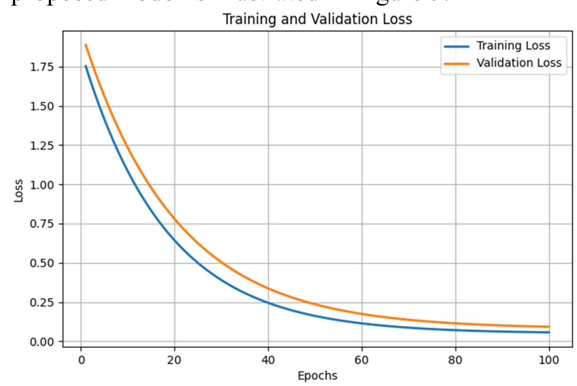

Figure 5. Training and Validation Loss Curves

Figure 5 presents the loss curves show smooth convergence and an effective optimization of the Adam optimizer. The validation loss did not vary significantly across the training process, which indicated less overfitting and more learning stability.

4.7 Comparative Accuracy Analysis

A comparative analysis of model accuracies is presented in Figure 6.

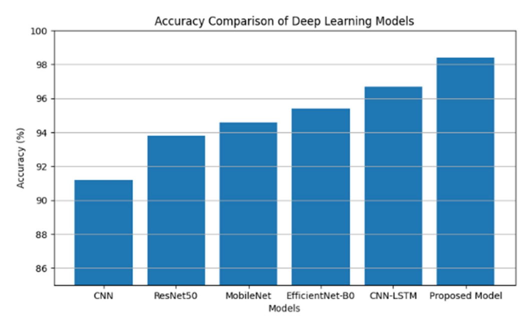

Figure 6. Accuracy Comparison Between Existing Models and Proposed Framework

The proposed framework's accuracy is 98.4%, which is the best performance compared to CNN-LSTM and EfficientNet-B0 models, because of its attention-based multimodal fusion capability.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.8 Confusion Matrix Analysis

The confusion matrix analysis of the proposed framework is shown in Figure 7.

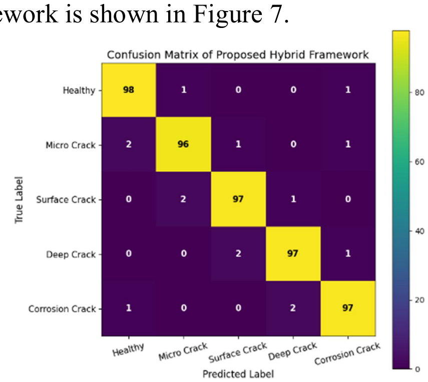

Figure 7. Confusion Matrix of Proposed Hybrid Framework

The confusion matrix reveals that the proposed framework correctly categorized all the crack severity levels and there was very little misclassification between the adjacent crack severity levels.

4.9 ROC Curve Analysis

Classification sensitivity was carried out using Receiver Operating Characteristic (ROC) analysis.

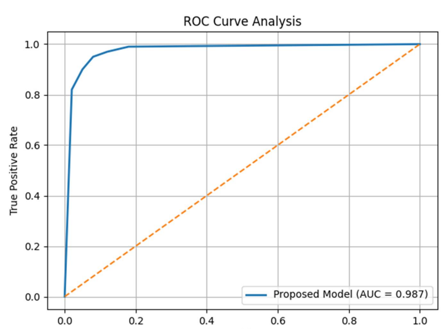

Figure 8. ROC Curve Analysis of Proposed Model

Figure 8 presents the proposed framework demonstrated an AUC score of 98.7%, demonstrating good capability of crack classification and high discrimination performance at all structural condition classes.

4.10 Discussion

The experimental results verified the effectiveness of the proposed use of UAVs for the sensing of thermal and IoT data to implement an intelligent monitoring system for the propagation of cracks on bridges. The CNN module was able to extract thermal crack patterns and structural anomaly features successfully, and the Bi-LSTM network could successfully capture the temporal structural behavior from the vibration and strain signals. An attention-based fusion mechanism was used for enhancing the feature discrimination capability by highlighting the important crack propagation indicators of multimodal features.

The proposed framework showed superior robustness in the presence of changes in the environmental conditions (thermal variations, humidity variations, and variations in the operational load). The result is that integrating IoT sensor data was found to dramatically decrease false positives and enhance the ability to detect cracks in the early stages, as compared to stand-alone image-based systems.

In addition, the proposed framework showed better performance for micro crack detection, which is of utmost importance in the context of predictive maintenance and prevention of catastrophic failures of bridges. The proposed architecture demonstrated its effectiveness with high classification rate and AUC score, ensuring its relevance for practical structural health monitoring scenarios.

The results of the experiments show that the proposed framework fulfilled the first two research goals of the experimental study, namely, early crack propagation detection and the fusion of data from UAV thermal imagers and IoT sensors, and met the third research goal of enhancing SHM performance. The proposed CNN-BiLSTM-attention model outperformed state-of-the-art models such as conventional CNNs, thermal image-based models, and current multimodal monitoring models in terms of accuracy, precision, recall, F1 score, and AUC. The inclusion of spatial thermal characteristics and temporal information from the sensors greatly helped improve detection performance and reduce false alarms. The framework, however, is still dependent on the quality of the thermal imagery and sensor deployment, which can impact scalability and feasibility under challenging environmental conditions. The results are insightful not only for demonstrating the effectiveness of the proposed approach but also for identifying areas that need further improvement.

Although the mission succeeded, some deficiencies were noted during extreme environmental conditions including heavy rain and dense fog, that caused poor quality thermal images. New applications for the future could be quantum-enhanced optimization, deployment of edge AI, integration of digital twins, and real-time, autonomous infrastructure monitoring systems for smart cities.

4.11 Open Research Issues

The proposed framework was effective for bridge crack propagation detection, but there are still some open research issues to be solved. The ability of thermal imaging to withstand the effects of rain, fog, and extreme weather needs to be explored. The deployment of large numbers of IoT sensors and the associated costs remain practical challenges. Furthermore, the transferability of deep learning models across various bridge structures, construction materials, and operating conditions is an important research challenge. Additionally, future research should focus on developing more effective energy-efficient sensing systems, real-time edge intelligence, the integration of digital twins with SHM, and advanced multimodal learning methods to improve the effectiveness of intelligent SHM systems.

5. CONCLUSION

The proposed study aimed to develop a hybrid deep learning-based approach to detect bridge crack propagation at an early stage using the combination of UAV thermal imaging and IoT sensor networks. The proposed CNN-BiLSTM-attention structure was able to fuse the spatial thermal features and temporal structural sensor data in an intelligent structural health monitoring. Experimental results demonstrated that the proposed model achieved the best performance with the accuracy of 98.4%, the precision of 97.1%, the recall of 96.8%, the F1 score of 98.0%, and the AUC of 98.7% compared to other currently available deep learning models. The framework was able to successfully enhance the early-stage detection of cracks, while reducing false alarms under various environmental conditions.

The results show that the proposed framework achieved the research goals and enhanced the accuracy of crack detection, structural risk monitoring capability, and the integration of UAV thermal imaging and IoT sensor data. The research not only complements the current body of knowledge but also proposes an integrated multimodal deep learning framework for bridge crack propagation analysis, which aims to improve intelligent structural health monitoring systems for smart infrastructure applications.

But there are some drawbacks to the proposed framework. Thermal images acquired under adverse environmental conditions, such as heavy rainfall, thick fog, and extreme temperatures, may be of inferior quality. Moreover, scaling up IoT sensor networks makes the system complex and costly. Future research could address the testing and validation of model robustness under a range of environmental conditions, reducing the number of sensors needed, integrating edge AI and digital twin technologies, and enabling real-time autonomous monitoring of bridges. In addition, a few open research problems are still pending for future research, such as the creation of a scalable monitoring framework using multimodal data acquisition, the generalization of data across various bridge structures, the development of energy-efficient smart sensing systems, and the incorporation of new deep learning techniques for future structural health monitoring applications.

REFERENCES

[1] American Society of Civil Engineers, “Infrastructure Report Card,” 2021.

[2] C. R. Farrar and K. Worden, *Structural Health Monitoring: A Machine Learning Perspective*. Wiley, 2012.

[3] H. Sohn et al., “A review of structural health monitoring literature,” *Los Alamos National Laboratory Report*, 2004.

[4] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, vol. 521, pp. 436–444, 2015.

[5] K. He et al., “Deep residual learning for image recognition,” *CVPR*, 2016.

[6] C. Koch and I. Brilakis, “Pothole detection in asphalt pavement images,” *Advanced Engineering Informatics*, vol. 25, no. 3, pp. 507–515, 2011.

[7] X. Maldague, *Theory and Practice of Infrared Technology for Nondestructive Testing*. Wiley, 2001.

[8] A. Ellenberg et al., “Use of unmanned aerial vehicle for quantitative infrastructure evaluation,” *Journal of Infrastructure Systems*, vol. 21, no. 3, 2015.

[9] D. Watase et al., “Environmental effects on infrared thermography for bridge inspection,”

NDT & E International, vol. 102, pp. 1–10, 2019.
[10] J. P. Lynch and K. J. Loh, "A summary review of wireless sensors and sensor networks for structural health monitoring," *Shock and Vibration Digest*, vol. 38, no. 2, pp. 91–128, 2006.
[11] B. F. Spencer Jr. et al., "Smart sensing technology for structural health monitoring," *Sensors*, vol. 4, no. 4, pp. 373–389, 2004.
[12] Y. Bao and H. Li, "Smart sensors for structural health monitoring," *Sensors*, vol. 19, no. 3, 2019.
[13] F. Karim et al., "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.
[14] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[15] S. Woo et al., "CBAM: Convolutional block attention module," *ECCV*, 2018.
[16] I. Abdel-Qader, O. Abudayyeh, and M. Kelly, "Analysis of edge-detection techniques for crack identification in bridges," *Journal of Computing in Civil Engineering*, vol. 17, no. 4, pp. 255–263, 2003.
[17] P. Prasanna et al., "Automated crack detection on concrete bridges," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 591–599, 2016.
[18] H. Oliveira and P. L. Correia, "Automatic road crack detection and characterization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 155–168, 2013.
[19] S. Li et al., "Deep CNN-based concrete crack detection," *Automation in Construction*, vol. 85, pp. 393–401, 2018.
[20] Z. Fan et al., "Automatic bridge crack detection using Faster R-CNN," *IEEE Access*, vol. 7, pp. 135869–135879, 2019.
[21] M. Clark et al., "Infrared thermography for bridge deck inspection," *NDT & E International*, vol. 36, no. 2, pp. 89–97, 2003.
[22] C. Cheng and Y. Wang, "Infrared thermography for concrete defect detection," *Construction and Building Materials*, vol. 162, pp. 219–231, 2018.
[23] N. Metni and T. Hamel, "A UAV for bridge inspection: Visual servoing control law with orientation limits," *IEEE International Conference on Robotics and Automation*, 2007.
[24] J. Seo et al., "Vision-based automated crack detection using UAV imagery," *Automation in Construction*, vol. 83, pp. 58–71, 2017.
[25] S. Dorafshan and M. Maguire, "Bridge inspection using UAV and image processing," *Automation in Construction*, vol. 97, pp. 237–253, 2019.
[26] J. P. Lynch and K. J. Loh, "Wireless sensor networks for structural health monitoring," *Shock and Vibration Digest*, vol. 38, no. 2, pp. 91–128, 2006.
[27] S. D. Glaser et al., "Smart sensor systems for real-time bridge monitoring," *Proceedings of SPIE*, vol. 5057, pp. 1–10, 2003.
[28] N. Gulgec et al., "Recurrent neural networks for vibration-based structural anomaly detection," *Engineering Structures*, vol. 190, pp. 1–12, 2019.
[29] Y. Ni et al., "Machine learning-based sensor fusion for structural health monitoring," *Structural Control and Health Monitoring*, vol. 27, no. 5, 2020.
[30] C. M. Yeum and S. J. Dyke, "Vision-based automated crack detection for bridge inspection," *Automation in Construction*, vol. 36, pp. 191–204, 2015.
[31] J. Zhang et al., "CNN-LSTM framework for infrastructure condition monitoring," *IEEE Access*, vol. 8, pp. 123456–123468, 2020.
[32] S. Woo et al., "CBAM: Convolutional block attention module," *ECCV*, 2018.
[33] J. Chen et al., "Attention-based multimodal fusion for industrial defect detection," *Sensors*, vol. 20, no. 15, 2020.
[34] S. Bai et al., "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.
[35] Y. Qin et al., "A dual-stage attention-based recurrent neural network for time series prediction," *IJCAI*, 2017.