January 2021

Validating NDE Dataset And Benchmarking Infrared Thermography For Delamination Detection In Bridge Decks

Eberechi Orie Ichi

VALIDATING NDE DATASET AND BENCHMARKING INFRARED THERMOGRAPHY FOR DELAMINATION DETECTION IN BRIDGE DECKS

By

Eberechi Orie Ichi, Msc.

A Thesis
Submitted to the Graduate Faculty

of the

UNIVERSITY OF NORTH DAKOTA

In partial fulfillment of the requirements

for the degree of

Master of Science
Civil Engineering
College of Engineering and Mines

Grand Forks, North Dakota

December
2021

This thesis, submitted by Eberechi Ichi in partial fulfillment of the requirements for the Degree for Master of Science in Civil Engineering from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

_____

Sattar Dorafshan

_____

Daba Gedafa

_____

Iraj Mamaghani

This thesis is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

_____

Chris Nelson
Associate Dean, School of Graduate Studies.

_____

Date

Permission

Title: Validating NDE dataset and benchmarking infrared thermography for delamination detection in bridge decks.

Department: Civil Engineering

Degree: Master of Science (MSc.)

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Eberechi Ichi

11/18/21

Table of contents

List of Tables

List of Figures

List of Acronyms

AI	Artificial intelligence
ACC	Accuracy
ASTM	American Society of Testing Materials
ASCE	American Society of Civil Engineers
Brisque	Blindness/reference less image spatial quality evaluator
CLAHE	contrast limited adaptive histogram equalization
CNN	Convolutional neural networks
DOT	Departments of Transportation
FHWA	Federal Highway Administration
FR_SB	Forest River south bound
FR_NB	Forest River north bound
GPR	Ground penetrating radar
IE	Impact echo
IRT	Infrared thermography
FPR	False Positive Rate
FNR	False Negative Rate
LTBP	Long-Term Bridge Performance
NDDOT	North Dakota Department of Transportation
NDT	Nondestructive testing
NDE	Nondestructive evaluation
NBE	National Bridge Element
NBI	National Bridge Inventory
Niqe	Naturalness image quality evaluator
Pique	Perception based image quality evaluator
PR_SB	Park River south bound
PR_NB	Park River north bound
PR M	Park River median
PPV	Positive Predictive value
RGB	Red, Green, Blue
SDNET2021	Annotated dataset containing IE, GPR and IRT data
TPR	True positive rate
TNR	True negative rate
UAS	Unmanned aerial systems
UAV	Unmanned aerial vehicles

Acknowledgments

I wish to express my sincere appreciation to the members of my advisory committee for their guidance and support during my time in the master's program at the University of North Dakota. Foremostly, my advisor Dr. Sattar Dorafshan contributed immensely to the successful completion of my master's program. Working closely with him has afforded me a rare opportunity to be mentored by such a patient, distinguished advisor, and researcher. His passion for innovative research and keenness for details has instilled in me a growing interest and sense of duty to contribute to the body of knowledge in my chosen career. His high expectations have likewise allowed me to explore more hidden capabilities. I am extremely grateful for all his contributions and believe that my experience with him will undoubtedly make me a renowned researcher. Many thanks also to Dr. Daba Gedafa, the Chair of the department, and Dr. Iraj Mamaghani, who made valuable contributions to the success of this study

I would like to acknowledge the North Dakota Department of Transportation (NDDOT) for funding this research project. I also wish to appreciate several others, such as Amrita Das, Bruce Dockter, who assisted in various ways. This study will not have been possible without the BDI and SkySkopes staff crew, who assisted in the data collection phase of this project.

I acknowledge my wife and son, who gave me all the moral, spiritual, and mental support throughout this journey. You are very much appreciated. My immediate and entire family members, Redeemed Christian Church of God Lord's Temple parish leader, Pastor Soji Haastrup, and members and all who supported me with their prayers are all well appreciated.

Dedication

This work is dedicated to GOD, the father of Light, whom only has immortality dwelling in the Light which no man has seen, nor can see and He is before all things, and by Him all things consist. To him be Honor and power everlasting.

Abstract

We collected and developed an SDNET2021 dataset, a rare combination of annotated non-destructive evaluation (NDE) data, from five in-service bridge decks in Grand Forks as a feed for validating, benchmarking, developing, training, and testing artificial intelligence (AI) models to evaluate, monitor, and assess bridge conditions. The developed dataset, which serves as ground truth, contains sound concrete as class 1 and delaminated sub-surface concrete as classes 2 and 3. The SDNET2021 dataset consists of 488 delaminated (class 2 and 3) and 1,448 sound (class 1) Impact Echo (IE) signals, 214,943 delaminated (class 2 and 3) and 448,159 sound (class 1) Ground Penetrating Radar (GPR) signals, and 1,718,083 delaminated (class 2 and 3) and 2,862,597 sound (class 1) pixels from Infrared thermography (IRT). SDNET2021 is publicly freely available. The dataset was annotated autonomously to reduce human errors and increase reliability. This study presents an adaptative image processing-based model for bridge deck sub-surface delamination evaluations. The proposed method adopts the IRT dataset and annotated ground truth generated from the in-service bridge decks. The model was developed by iterating the sensitivity (s) parameter and optimized by selecting s-values based on performance evaluation metric interactions. The evaluated 2- and 3-clustered optimized s-values ranged from 0.365 to 0.38 and 0.459 to 0.486, respectively. An average accuracy of 69% was obtained for the model. The study revealed that several factors, such as delamination depth and spatial dimensions, ambient weather conditions such as wind speed, temperature, and humidity, and mosaic image quality affect the IRT model's performance.

CHAPTER 1. Introduction and Background

One in five miles, or 173,000 total miles of highways and major roads, as well as more than 45,000 bridges in the US, are in poor condition. There are more than 617,000 bridges across the United States. Of these, 42% are at least 50 years old, and 7.5% are considered structurally deficient. Significant nationwide bridge repair funding deficiencies are also approximately $125 billion, with a 58% projected increase in annual rehabilitation funding necessary to improve conditions. These repairs will take until 2071 to complete at the current expenditure rate. An overwhelming increase in additional deterioration over 50 years has been projected by the ASCE (2021).

This issue has motivated the federal government to propose a $1.2 trillion infrastructure bill. Approximately $110 billion has been assigned to rehabilitate large bridges, highways, roads, and main streets, and modernize approximately 20,000 miles of road that have a critical need for reconstruction and repair. The bill, which the Senate passed on August 10, 2021, will mitigate bridge deterioration (Secretary of Transportation 2022 budget Highlights, 2021 & Holland & Knight Alert, 2021).

Approximately 86% of the existing 4,377 bridges are in fair or good condition, while 14% are classified as structurally deficient in North Dakota alone, according to a 2019 infrastructure report card by the ASCE of the North Dakota section (ASCE Report, 2019). Bridge element deterioration is caused by the continuous imposed traffic load, exposure to environmental elements, uncommon harsh frost action, and de-icing chemicals.

In-service bridge maintenance requires significant bridge deck repair and rehabilitation, since these elements are more susceptible to deterioration than other bridge parts. A section of existing bridge in ND, Forest River North Bound, that is scheduled for deck repair is depicted in Figure 1.

Continued bridge distress and deterioration is becoming a growing concern. Stakeholders in bridge management face the challenges of developing and deploying feasible techniques, from conceptual stages to construction and maintenance phases, to ensure that bridges are safe, serviceable, durable, and stable.

It is critically important for government agencies, Departments of Transportation (DOT's), and all other concerned parties to create a systematic bridge evaluation, defect detection, monitoring, repair, and maintenance program. The repair and rehabilitation of existing deteriorating bridge elements must be effectively prioritized, which will ensure the proper and timely disbursement of funds and prevent bridges from further deterioration and distress.

Physical and visual inspections have shortcomings in reliability and accuracy, with significant variability in condition rating assignments by different inspectors. Factors that affect inspection results are traffic fears, near visual acuity, color vision, formal training on bridge inspections, dependency on inspector skill or experience, accessibility, and complexity (Graybeal et al., 2002; Dorafshan et al., 2018a&b; Rens et al.,1997). Assessments using these methods are usually through contact, which involves significant setbacks, lane closures, and traffic disruptions.

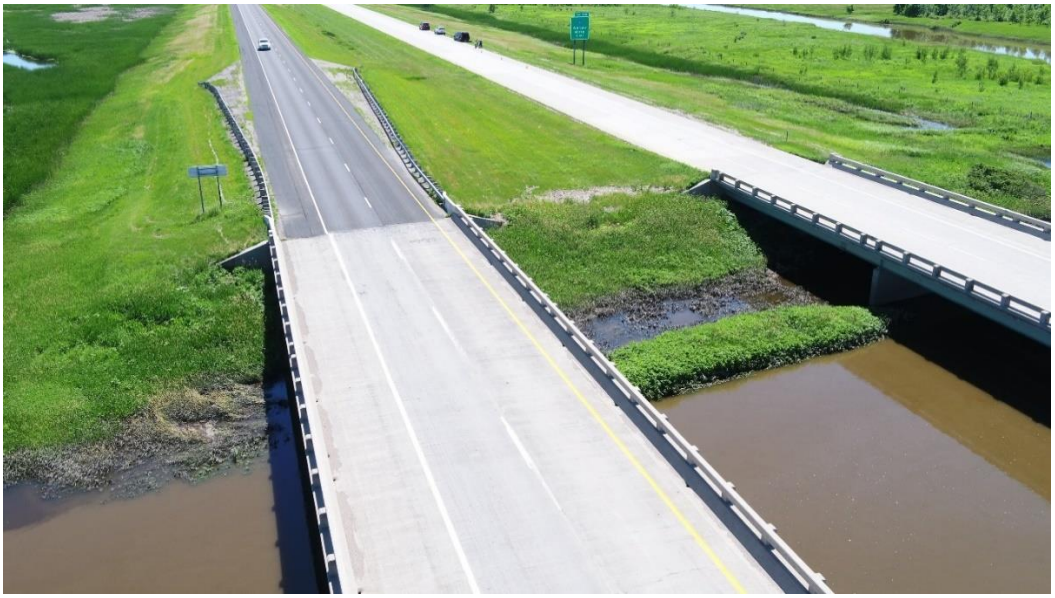


Figure 1. Approach and concrete bridge deck captured using a UAS during Summer 2020

1.1 Motivation and Problem Statement

Traditional bridge evaluation techniques are destructive, which generally involves collecting bridge deck samples by coring through the bridge, compromising the bridge's integrity. These samples, collected from localized portions of the bridge decks, are used as the basis for evaluating the entire bridge's condition. These evaluations, if not well interpreted by an expert, may result in false recommendations and imminent challenges. Destructive tests using core extractions may provide useful information about a deck's condition; however, they are labor-intensive, time-consuming, alter the structure's condition, and cannot be applied ubiquitously over large bridge deck areas.

Other traditional inspection methods, such as physical and visual, chain dragging, and sounding, rely on the inspector's experience and subjective judgment to differentiate the defective sub-surface or surface regions from the sound deck parts. Other concerns are difficulties in accessing bridge parts that are out of reach by physical inspection, safety issues during inspection due to heights, or open traffic and lane closures. The interpretation of field data inspection results can be cumbersome with a likelihood of errors and inconsistencies. Other advanced non-destructive evaluation (NDE) techniques, such as Infrared Thermography (IRT), Ground Penetrating Radar (GPR), and Impact Echo (IE), which have been adopted in past bridge projects and studies, have not been well validated over entire bridge deck areas, resulting in varying degrees of accuracy when detecting delaminations. NDE studies have been primarily performed on laboratory specimens, with very few studies performed on in-service bridge decks.

Several drawbacks have led to the adoption of advanced NDE techniques and remote structural evaluation methods, and the collection of reliable ground truth annotated datasets for detecting, assessing, and monitoring structural sub-surface bridge deck defects.

Recent advances in Unmanned Aerial Systems or Vehicles (UAS's/UAV's), robotics, and Artificial Intelligence (AI) can address these limitations; however, the development and adoption of proper NDE techniques for quality data collection, the development of reliable annotated datasets, processing, and interpretation is essential when evaluating in-service bridges. A combination of technologically advanced tools, such as infrared and visual camera sensors mounted on UAS to collect thermal and visual images, allows professionals and inspectors to perform bridge condition assessments at high altitudes without disrupting traffic flow.

Data collection and the development of reliable ground truth datasets is an important aspect of bridge deck NDE testing. The NDE dataset serves as an input for AI model training, validating, and testing. Datasets of surface defects exist; however, "ground truth" datasets have been limited to laboratory specimens and destructive methods. We have developed a reliable annotated dataset for subsurface defect detection using IRT, GPR, and IE NDE methods. The data were collected from five in-service Grand Forks, ND bridges scheduled for repair and maintenance. The collected IRT data were benchmarked for delamination detection using image processing and machine learning techniques.

1.2 Study Goals and Overview

Two broad bridge evaluation aspects were explored for this work: NDE data acquisition, validation, and annotation, and benchmarking the IRT dataset with the validated ground truth for delamination detection using image processing and machine learning techniques.

1.3 Organization of the Thesis

The thesis includes four chapters:

Chapter One consists of the introduction, motivation and problem statement, study goals, overview, and thesis organization.

Chapter Two consists of the study's first part, entitled "SDNET2021: An Annotated Dataset of Subsurface Structural Defects in Concrete Bridge Decks." This section contains the reports for the NDE dataset acquisition, conditions and methodology, data quality, data annotation, data validation, and study summary. This portion of the study has been submitted to a peer-reviewed journal.

Chapter Three consists of the second part of this study, entitled "Investigating the Effectiveness of Infrared Thermography in Delamination Detection for Concrete Bridge Decks." This section contains the study's goals and objectives, delamination detection methodology, model development and image enhancement, model evaluation metrics, results, and discussion. This portion of the study has also been submitted to a peer-reviewed journal.

Chapter Four consists of the conclusion, recommendations, and future work.

CHAPTER 2. SDNET2021 Dataset Acquisition and Annotation

2.1 An Annotated Dataset for Subsurface Structural Defects in Concrete Bridge Decks.

This section contains the developed SDNET2021 dataset and reports the NDE dataset acquisition methods, conditions and methodology, data quality, data annotation, data validation, and study summary. This portion of the report has been submitted to a peer-reviewed journal for publication and is publicly freely available.

2.2 Introduction and Past Study Review

The Code of Federal Regulations mandates periodic bridge inspections to ensure public safety. Inspectors must periodically inspect four billion square meters of reinforced concrete bridges (Dorafshan & Azari, 2020a). The recent demand for automation in infrastructure construction, inspection, and planning has led to artificial intelligence (AI) implementation for bridge evaluations. Supervised deep learning networks, such as convolution neural networks (CNNs), are the most promising of these AI methods due to their high accuracy. A comprehensive bridge deck evaluation requires the detection of both surface and subsurface defects. Many AI models have been developed for surface defect detection; however, effective AI models for subsurface defect detection have been hindered. AI-enabled bridge deck subsurface defect detection evaluations require annotated realistic NDE datasets. Datasets for surface defects, such as concrete cracks (Dorafshan et al., 2018), exist, where images are labeled to a class, such as cracked or intact. Other datasets for surface defects were annotated at the pixel level, where each pixel was assigned to a class (Liu et al., 2019), while others attempted to capture surface defects using a bounding box (Mei & Gül, 2019). NDE datasets are rare. McLaughlin (et al. 2019) developed a dataset consisting of 500 IRT images with 512x640 pixel resolution taken from four reinforced concrete bridges. The images consisted of 261 images with potential delamination and 239 with no delamination. The images were annotated using a semantic pixel-wise method into two classes, sound and

delaminated; however, the annotation was not based on actual delaminations. Dorafshan & Azari (2020a, 2020b) presented an annotated impact echo dataset (IE2020) of laboratory-created specimens for studying deep learning models to evaluate concrete bridge decks. The IE data was subdivided into three major categories: Sound (S), Defective (D), and De-Bonded (DB), with 736, 715, and 2092 samples in each class, respectively.

IE2020 was an effective dataset for deep learning model development; however, the defects were artificially made, which could negatively affect the accuracy of the models tuned on this dataset if used to classify real IE data.

Kalogeropoulos et al. (2013) collected GPR data for 0.08 m thick concrete slabs exposed to chloride contamination. Cores were taken from the concrete slabs using a drilling rig, and the free chloride ion content of eight slices that were 0.01 m thick was calculated using the water extraction procedure to validate the dataset. Dinh et al. (2016) used GPR data from twenty-four in-service bridge decks that was collected during the Federal Highway Administration's (FHWA's) Long-Term Bridge Performance (LTBP) Program. The data were collected with a ground-coupled 1.5-GHz GPR antenna on cast-in-place concrete bridge decks. The study's objective was to characterize the corrosive environment and create an overall bridge deck condition assessment; however, this dataset was validated with other NDE methods, not the actual bridge defect state. Liu et al. (2020) collected GPR data and converted them into segmented grayscale images of 300×300 pixels. The final dataset contained 3,992 images of 13,026 rebar targets, 2,370 of which were utilized for training and 1,622 for testing. The dataset had only two categories, hyperbola and background, and was labeled using the bounding box method.

Publicly available datasets specifically designed to evaluate crack and delamination detection algorithms are limited. Most of the datasets have been processed and simplified since they do not depict real life scenarios (Eisenbach et al., 2017). Some of these datasets manually

exclude any disturbance and focus only on pavement surfaces using static images (Amhaz et al. 2016, Shi et al., 2016, Yang et al. 2020), and others are not publicly available or are not validated with ground truth (Chambon & Moliard 2011 & Amhaz et al. 2016). A summary of existing structural defect datasets and their descriptions are listed in Table 1. There are few open-source datasets, and those that exist are primarily visual images for crack detection with sparse pixel-level annotations. Mei & Gül (2020) claimed that the bounding box method is not appropriate for defect annotation because of the irregular crack shapes. Too many details are lost if a rectangular bounding box is used to cover these irregular cracks. These open subsurface defect datasets are predominantly created using laboratory specimens.

Table 1. Review of NDE datasets

References	Data Type and Description	Defect types	Material or Structure	Annotation method	Limitation
RGB Images (Surface Defects)					
SDNET2018 Dorafshan et al. (2018)	Image-56,000 sub-images (256×256 px)	Crack (widths from 0.06 to 25 mm)	Concrete bridge decks, walls, and pavements	Labeling	Limited to crack defects. Not validated with ground truth
Özgenel et al.(2018)	40,000 images having 227x227 pixels generated from a 4,032x3,024 resolution camera	Cracks on buildings	METU Campus Buildings	Labeling	Dataset is based on buildings only
CrackForest Dataset and AigleRN. (Fan et al., 2020)	CFD contains 118 RGB, and the AigleRN database contains 38 gray-level images.	Cracks	Asphalt Pavements	Labeling	Only surface defects
EdmCrack600 (Mei et al. 2019)	600 RGB images	Crack	Pavement	Pixel level annotation	Only surface defects
COncrete DEfect BRidge IMage (CODEBRIM) dataset. (Mundt et al. 2019)	Over 17,754 RGB images	Cracks, spall, exposed bars, corrosion stain	Concrete bridges	Bounding box labeling	Only surface defects
Zhang et al. (2019)	6,500 3D pavement images	Cracks	Asphalt pavement	Labeling	Not publicly available and limited to asphalt pavement
Majidifard et al. (2020)	7,237 RGB images of pavement sections extracted from Google maps	Structural cracks	Asphalt pavement surface	Bounding box labeling	Not publicly available, dataset not validated with ground truth, without delamination defect
NDE (IRT/IE/GPR) Subsurface Defects					
IE2020 Dorafshan et al (2020a and b)	Impact Echo-2,016 IE signals	Debonding and subsurface defects	Laboratory concrete specimens	Signal labeling	Limited to laboratory specimens, dataset not validated with ground truth
Kalogeropoulos et al. (2013)	GPR signals	Chloride migration detection	Laboratory concrete decks	Signal labeling	Limited to laboratory specimens, dataset was validated using a destructive method on core samples

Dinh et al. (2016)	GPR signals collected within the FHWA's LTBP Program	Characterize the corrosive environment	Asphalt and Concrete bridge decks	Signal labeling	Dataset not validated with ground truth but was validated with other NDE methods and bridge decks
Liu et al. (2020)	GPR signals converted into 3,992 grayscale images	Detection and localization of rebar	Residential buildings under construction	Bounding box labeling	Dataset not validated with ground truth
McLaughlin et al. (2019)	500 infrared images	Sub-surface delamination	Reinforced concrete bridges	Semantic pixel-wise image labeling	Dataset is not publicly available and was not validated

SDNET2021 is the only publicly available NDE dataset annotated based on different classes of delamination in reinforced concrete bridge decks and validated by the actual state of delamination in the field, to the best of the author's knowledge. SDNET2021 was collected to represent challenges faced by bridge inspectors in the field, such as **i)** Change in weather conditions, **ii)** Significant environmental effects and noise, such as shadows, occlusion, stains, texture difference, and low contrast due to overexposure, **iii)** Blurring effects due to UAS movement and poor lighting conditions, **iv)** Inclusions of disturbances during data collection, and **v)** Manual annotation limitations.

SDNET2021 contains three types of NDE data; IRT, IE and GPR for existing in-service bridges, validated by ground truth. The combination of these rare and state-of-the-art validated datasets, made publicly available, will be useful to benchmark, develop, train, and test AI models for bridge evaluation, monitoring, and condition assessment.

2.3 Objectives of the Study

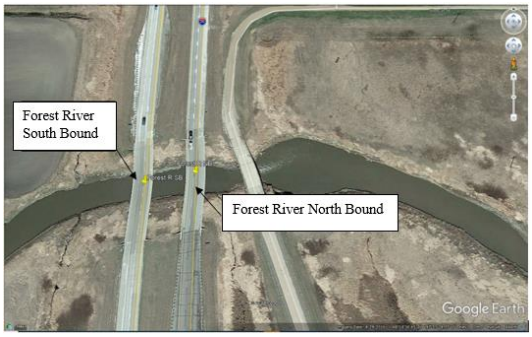
The main goal for developing the novel SDNET2021 dataset is to provide a means for autonomous bridge deck evaluations using NDE. The research activities performed to achieve this goal include: (1) A literature review on existing NDE datasets, (2) Brief description and current bridge deck conditions, (3) NDE data acquisition of field IE, GPR, and IRT field data, (4) NDE data quality control, and (5) NDE data annotation.

2.4 Experimental Program

A brief description of the bridges investigated, and the NDE data acquisition methods are discussed in this section.

2.4.1 Investigated Bridges.

The North Dakota Department of Transportation (NDDOT) scheduled five bridges for deck repair during the Summer of 2020. These bridges were between 47 and 49 years old at the time of repair and were built along the I-29 corridor, except for the Park-River median designed to provide access to the rest area (Figures 2a and 2b). The bridge lengths range from 64 m (Forest River bridge) to 142m (Park River bridge). An information summary is listed in Table 2. Inspectors rated the bridges according to the National Bridge Inventory (NBI) and National Bridge Element (NBE) level guidelines (Nebme, 2019), and reported the Health Index (HI) and Condition State (CS). The concrete bridge decks are supported on steel beam girders with expansion joints at appropriate intervals. This investigation only focused on concrete bridge decks and did not include the steel girder supports, vertical supporting piers, and sub-structure.



(a)

(b)

Figure 2. Google earth map depicting the (a) Forest River bridges (b) Park River bridges. (Source: Google Earth)

Table 2. Summary of bridge condition ratings according to The NBE and NBI Rating System, 2019. (Source: FHWA infobridge)

Bridge ID	Structure Number/Year Built	Width x Length (m^2)	Last Inspection	Deck and Bridge Rate	Deck Area (sqm)	Condition State (CS)	Health Index (HI)
FR_SB	0029168629 L (1971)	12.7x64	Nov. 2019	NBI: 6/Satisfactory Bridge condition: Fair	816	CS1- Good 100%-816	HI-100%
FR NB	0029168632 R (1971)	12.7x64	Nov. 2019	NBI: 6/Satisfactory Bridge condition: Fair	816	CS1-Good-99% - 813.2, CS2-Fair-1%-2.8	HI-99.8%
PR NB	0029179087 L (1973)	11.3x141.7	Dec. 2019	NBI: 5/Fair Bridge condition: Fair	1,806	CS1-Good-99%-1,794, CS2-Fair-1%-12.1	HI-99.78%
PR M	0029179123 M (1973)	7.3x111.3	Dec. 2019	NBI: 5/Fair Bridge condition: Fair	977	CS1-Good-92% - 907, CS2-Fair-7%-70	HI-97.65%
PR SB	0029179147 R (1973)	14.9x120.4	Dec. 2019	NBI: 5/Fair Bridge condition: Fair	1,974	CS1-Good-99%-1,966, CS2-Fair-1%-8	HI-99.87%

Note - FR SB: Forest River South Bound, FR NB: Forest River North Bound, PR NB: Park River North Bound, PR SB: Park River South Bound, PR M: Park River Median.

Bridge condition has continually declined from good in 2015 to fair in 2019, according to the NBE and NBI rating System; therefore, the NDDOT planned to rehabilitate these decks during the Summer and Fall of 2020.

2.4.2 IRT Data Acquisition

IRT is commonly used to detect subsurface delaminations without contact. IRT cameras convert emitted electromagnetic radiation to temperature. The rate at which this energy is emitted is a function of the material's temperature and emissivity. A material's emissivity defines the correlation between the true kinetic temperature and the object's radiant temperature (Robert 1982).

$$M = \varepsilon \sigma T^4 \quad (1)$$

where M = Total energy emitted from the surface of a material, ε = Emissivity, σ = Stefan-Boltzmann constant, and T = Temperature of the emitting material in Kelvin. Defective areas will have different electromagnetic radiation than the intact areas, visualized as a change in pixel intensity. The thermography reveals that all objects with a temperature higher than absolute zero emit radiation in the infrared range, at wavelengths of 700 nm–1 mm

corresponding to frequencies of 430 THz–300 GHz, which is between visible radiation and the microwave range.

The IRT data was collected using a UAS mounted with a thermal camera at an average altitude of 18m. The environmental details for the IRT data collection are listed in Table 3.

Table 3. IRT data collection ambient weather condition on the inspection date

Bridge ID	Time	Temperature (°C)	Humidity (%)	Wind Speed (kmph)
FR_SB	9:55-10:25am	26.0	47.0	10.5
FR_NB	10:26-10:44am	26.7	44.0	12.9
PR_NB	11:36-11:55am	27.0	47.0	12.9
PR_MD	12:09-12:32pm	27.8	44.5	14.5
PR_SB	12:34-12:55pm	27.8	45.0	16.1

The UAS and camera specifications used are presented in Table 4. Figure 3(a) illustrates the crew collecting IRT using an Unmanned Aerial System.

Table 4. UAS and camera specifications

DJI Matrice 210 UAV	
UAS Specification	Description
Takeoff Weight Maximum	6.14 kg.
Maximum Flight Altitude	2,500 m (above sea level)
Maximum Flight Time	Thirty-two minutes (approx.) with TB55 batteries and no payload
Maximum Tilt Angle P-mode (GPS)	35° (Forward Vision System enabled: 25°)
Hovering Accuracy (GPS Mode)	Vertical: ±0.5 m (±0.1 m, with Downward Vision System enabled) Horizontal: ±1.5 m (± 0.3 m, with Downward Vision System enabled)
Operating Temperature	-20 to 45°C
Dimensions Unfolded	887x880 x408 mm
FLIR XT V2 Thermal Camera	
Characteristics	Specifications
Thermal Resolution	640×512 pixels
Full Frame Rates	30 Hz (NTSC) 25 Hz (PAL)
Spectral Band	7.5 - 13.5 μm
Pixel Pitch	17 μm
Thermal Imager/ Detector type	Uncooled VOx Microbolometer
Digital Zoom	2x, 4x
Field of View	24°×19°

2.4.3 IE and GPR Data Acquisition

IE is a widely used nondestructive testing (NDT) method that has been used successfully to evaluate concrete structures (Kee et al., 2012, Dorafshan et al., 2020). IE uses elastic waves to identify and characterize delaminations in concrete structures. This method uses the transient vibration response of a plate-like structure subjected to a mechanical impact. IE was implemented for deck evaluation by conducting point testing on a pre-defined grid. The

transient time response of a solid structure is measured with an accelerometer placed on the surface close to an impact source.

Similarly, GPR can be used to qualitatively assess bridge decks by detecting suspected delamination or corrosive deterioration. The GPR technique works by transmitting electromagnetic radio waves through the concrete, with frequencies ranging from 10 MHz to 2.5 GHz. The reflected electromagnetic waves are then recorded when the reflector's dielectric constants, such as in rebar or delamination, are different from the concrete. The GPR data usually consists of changes in reflection strength and the arrival times of specific reflections, source wave distortions, and signal attenuations.

IE and GPR data were collected from portions of aforementioned bridge decks ermarked for repairs. The evaluation was caried out while the decks were closed to traffic flow. The environmental conditions during these data collection are presented in Table 5. Average values for each inspection day were used to represent temperature and humidity during data collection.

Table 5. IE and GPR data collection ambient weather and deck condition

Bridge ID	Date Collected	Temperature (°C)	Humidity (%)	Deck Condition
FR_SB	7/6/2020	18.3	67.0	Dry
FR_NB	7/7/2020	27.8	43.0	Dry
PR_NB	7/8/2020	25.6	66.5	Dry
PR_MD	7/6/2020 - 7/7/2020	23.1	55.0	Dry
PR_SB	7/9/2020	22.8	56.0	Dry

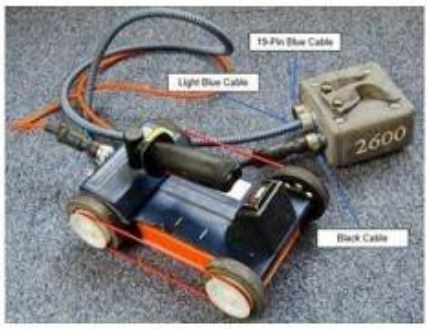
GPR data were collected during the investigation by scanning along the longitudinal and transverse sections of the bridges using GSSI GPR Equipment with a ground-coupled antenna**.** Other information associated with GPR data collection and equipment used is grid arrangement (600x3,000mm), a GSSI SIR-3000 Data Acquisition System, a GPR Antenna Frequency of 2600MHz, Horizontal Parameters (Scans/meter – 200), Vertical Parameters (Samples/Scan-512), Bits per Sample (16-bit), Dielectric Constant (6.25), Range (12.00ns),

Depth to direct coupling (2.1ns), Surface Type, and the condition of a concrete surface free from debris.

IE signal data were collected with a PCB 353B15 Accelerometer. IE equipment and specifications include a NI-USB-4431 USB DAQ System**,** Laptop with LabView software, a PCB 353B15 Accelerometer, Samples per Scan (204800)**,** Sampling Rate (102.4 kHz., Grid), and 300mmx300mm test grid. Figures 3b and 3c depict the IE and GPR field data collection setup.

(a)

(b)



(c)

Figure 3. (a) UAS flight crew collecting IRT data, (b) Crew collecting IE data on bridge deck, and (c) GPR scanning equipment for data collection

2.4.4 NDE Data Acquisition Conditions

NDE data collection needs to be properly performed to maximize efficiency. Several standardized codes and past publications were reviewed to provide a guide for effective NDE data acquisition (Table 6).

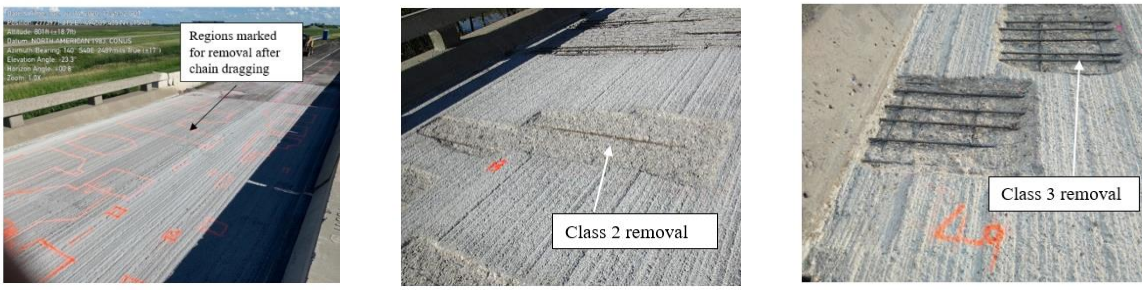
Table 6. Comparison of parameters for data collected in this study with past studies

Method	Data Collection for Past Studies	Data Collection in this Study	Reference
IRT	Weather conditions (Ambient Tempt. Change, Tempt., RH, wind speed): 8°C-10°C, 6-19 km/h. 47%-71%	-4.4° C -27.8° C, 1.6kmh-22.5 km/hr, 49-96.5% .	(Washer et al., 2009; Washer et al., 2013, ASTM 2013)
	Solar loading: An ambient temperature rises of 20°F with four hours of sun and winds less than 24 km/h on PCC.		
	No testing: Ambient air temperatures are less than 0 °C, and wind velocity exceeds 50 km/hr.		
	Camera specifications: Thermal IR 640 × 480 pixels res.	FLIR 640 × 512 pixels res.	(Hiasa et al., 2017; Vaghefi et al., 2012)
	Altitude: 2-8 meters above	18m AGL	(Hing & Halabe, 2010, Farrag et al. 2016, Tran et al., 2018, Vaghefi et al. 2012)
	Overlapping rates: 60% -65% for end lap and 30% ±15% for side lap.	65-80%	(Vaghefi et al. 2013),
	Angles and offset distances and number of images taken: Angles ranging from 0 to 45°, offset distances of 20–40 m, and a total of 295 images taken in 3 flights.	Angles ranging from 0 to 35°, and a total of 1064 thermal images taken in 3 flights.	(Vaghefi et al. 2012),
GPR	Surface Type and condition: Air-dried surface of the asphalt concrete layer	Concrete surface-dried cleared of debris.	(Abouhamad et al. 2017, Rhee et al. 2019).
	Weather and grid arrangement: RH 76%, 10 ºC -14ºC	600x3000mm spacing grid RH 43-67%, 18 ºC -27ºC	(Abouhamad et al. 2017, Rhee et al. 2019)
	Antenna Type and Frequency: For Ground coupled, 1.5GHz, GPR inspection passes should either be parallel or perpendicular to the direction of the traffic	2600MHz.Equipment: GSSI SIR-3000 Data Acquisition System. GPR Antenna	(Abouhamad et al. 2017, Hing et al. 2010, Maierhofer 2003)
	Scanning rate: 12 scans/m in the longitudinal direction at a speed of 80–100 km/h. Samples per scan 512 Samples/Scan	Horizontal Parameters: (20scans/m). Vertical Parameters: 512 Samples/Scan	(Diamanti et al. 2017, ASTM 2010)
	GPR sampling window: 8 -12 ns for a ground-coupled, and up to 40 ns for an air-coupled antenna.	12 ns range, 6.25 Dielectric Constant, 2.1ns Depth to Direct Coupling, 16-bit/sample.	(Rhee et al. 2019)
IE	Grid: 300mm × 300 mm - 900x900mm Contact time: 20 to 100 µs	300mm x 300mm test grid.	(Gucunski et al. 2008, ASTM 2010).
	Sampling frequency: 200, 500 kHz	Samples/Scan: 204800. Sampling Rate: 102.4 kHz., Grid:	(Azari & Lin 2019, Zhang et al. 2016, Gucunski et al. 2008).
	Surface condition: Concrete polished carefully with sandpaper and then rinsed with water.	Concrete surface-dried cleared of debris	(Azari & Lin 2019, Zhang et al. 2016, Gucunski et al. 2008).

2.4.5 Ground Truth for Data Validation

The NDDOT surveyed all five decks to identify subsurface delamination locations and sizes using chain dragging. The locations susceptible to delamination were marked, and their locations were mapped using GPS (Figure 4a). The asphalt overlay of all five bridge decks was removed by milling off the top 75 mm of each deck prior to chain dragging. The marked regions were then prepared for removal by jackhammering to a depth just above the top of the

reinforcement bars. These areas were classified as class two (2) removal (Figure 4b) and chain dragged again to detect possible deeper delamination. Observed delaminated areas within the class 2 regions were removed to a depth below the rebar reinforcement, called class three (3) removal (Figure 4c). These definitions were used to define delamination classes for annotating the data collected during this study: **Class 1** Sound**,** with no delamination; **Class 2** Delamination, with delamination above the top bar mat; and **Class 3** Delamination, with delamination below the top bar mat. A typical cross-bridge section depicting the levels of delamination removal is illustrated in Figure 5.



(a) (b) (c)

Figure 4. (a) Markings of delaminated portions, (b) Class 2 sub-surface delamination removal, and (c) Class 2 sub-surface delamination removal

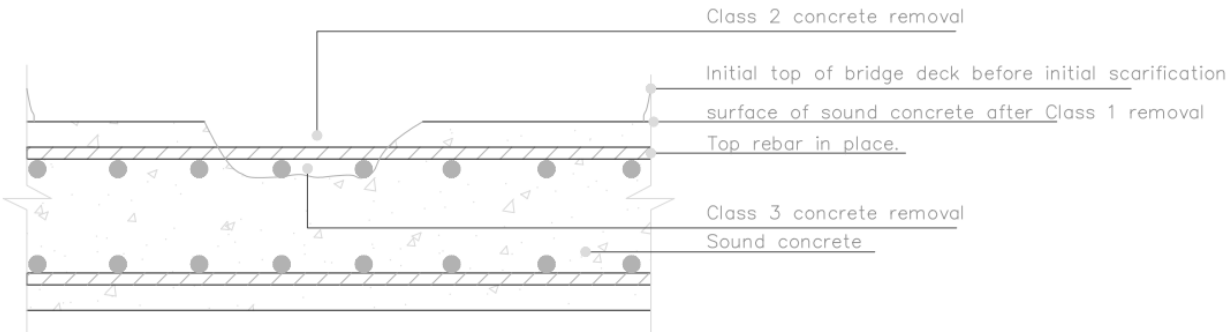


Figure 5. Bridge section depicting a delaminated sub-surface with embedded rebar

A set of delamination maps that indicated each delamination's location and class, were generated for each bridge deck and used as ground truth for NDE data classification. Figure 6 depicts the ground truth map for Forrest River South Bound as an example. Classes 2 and 3 were distinguished from sound concrete by dashed and solid boundaries, respectively.

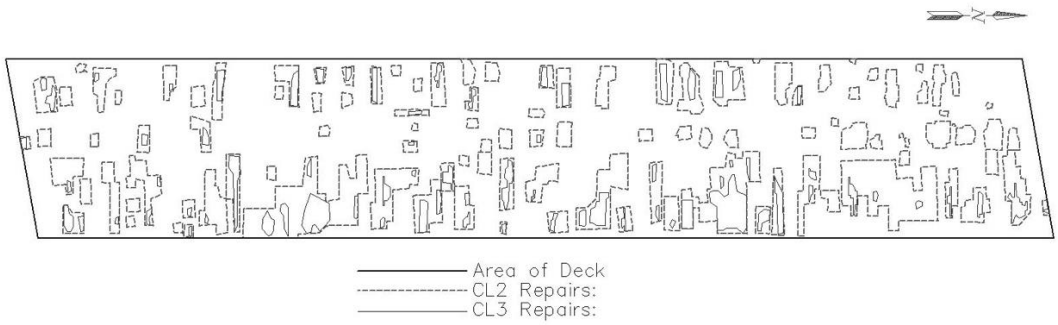


Figure 6. Layout plan depicting the delamination survey for the Forest River South Bound bridge

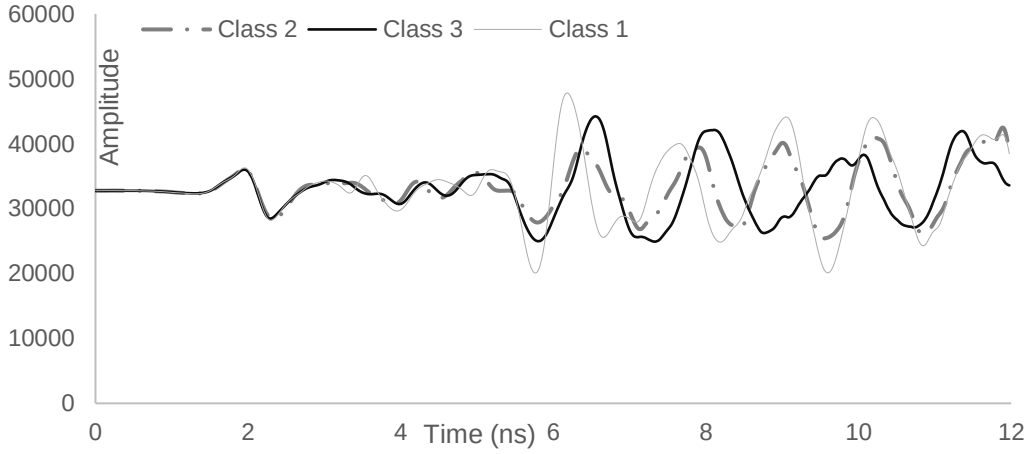
2.4.6 NDE Data Quality Control

The performance of an AI model constructed using a dataset is often influenced by the data's quality; therefore, a set of known quality assessment metrics were used to assess the IRT, IE, and GPR data. The collected data was classified into time series and image data; therefore, appropriate quality control methods were used for each type.

The output of the IE signals is time and acceleration (g) with 204,800 rows and each testing point saved in an 'lvm' file format. A total of 1,936 IE test points (files) were acquired for the entire bridge sections. The time duration varied from a minimum value of zero (0) to a maximum value of 1.99990 with time steps of 9.765625E-6 seconds, approximately 0.00001 seconds. The IE data were checked for null, duplicate, void or missing values, correctness, repetitiveness, and other pre-processing operations deemed fit for quality checks. The same quality checks were conducted on the GPR raw data. Each testing signal was saved in an 'MS Excel' format, for a total of 663,102 signals. The signals have a vertical time sale of 12 ns with 512 samples per trace. The longitudinal signal scans along the length of the bridges resulted in an output file with 16,383 amplitudes. The transverse signals, which scan across a bridge's width, result in an output file with approximately 1,225 amplitude signals. All files and signal outputs were checked for null, missing, and duplicate values, and the quality was satisfactory with no observed irregularities in the data files. A preliminary investigation was performed to ensure that the collected IE and GPR data were consistent with the class. An example of this investigation is presented in Figures 7 (a) and (b), where IE and GPR signals

collected from areas with different degrees of subsurface delamination have similar origins and waveform matches.

(a)

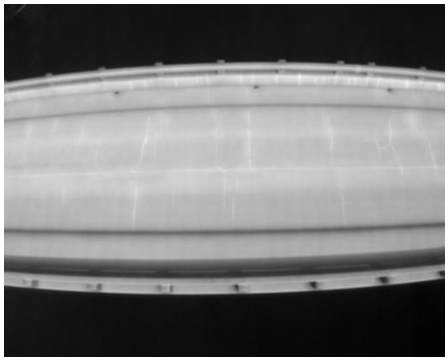


(b)

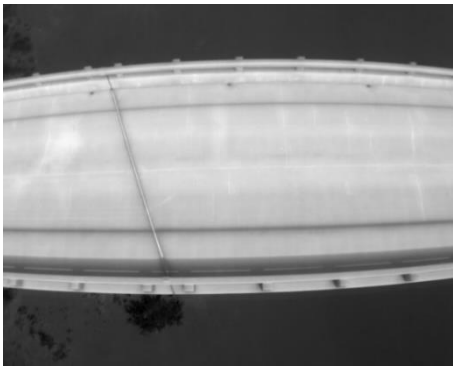
Figure 7. Selected samples of (a) IE signal for Park River Median classes 1, 2, and 3, and (b) GPR signal for Park River file_004 for Forest River SB.

Subjective and objective methods are the two broad classifications for image quality assessment. Non-reference objective assessment metrics, such as Perception based image quality evaluator (Pique), Naturalness image quality evaluator (Niqe), and Blindness/reference less image spatial quality evaluator (Brisque), were selected to evaluate IRT image quality. Brisque compares an image to a default model computed from images of

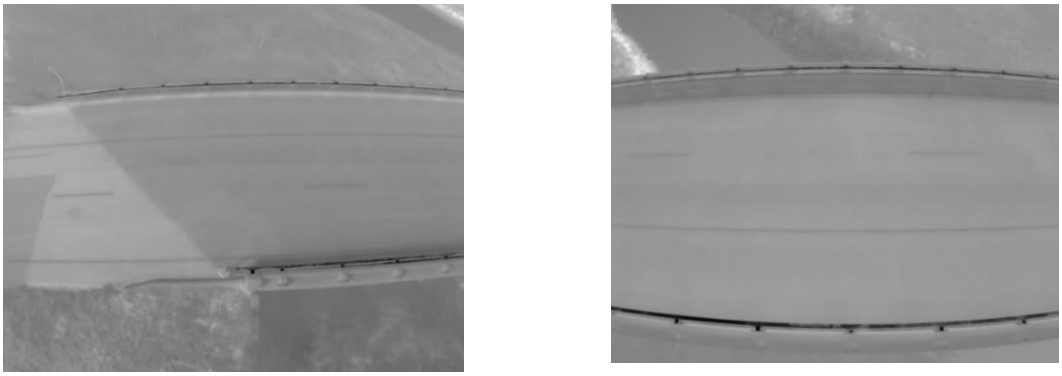
natural scenes with similar distortions. A smaller score indicates better perceptual quality. The metric scores for Brisque range from Excellent [0,20] to Bad [81,100] (Taylor et al., 2020). Pique is a local feature-based image quality assessment tool that calculates the Mean Subtracted Contrast Normalized (MSCN) coefficient. Brisque and Niqe are based on spatial features derived from natural scene statistics. Niqe compares an image to a default model computed from images of natural scenes. The pixel intensity distribution of natural images differs from distorted images. This difference in distribution is significantly pronounced when pixel intensities are normalized and the distribution is calculated. The pixel intensities of natural images after normalization follow a Gaussian distribution, while pixel intensities of unnatural or distorted images do not. The deviation of the distribution from an ideal bell curve measures the amount of distortion in the image. A lower Niqe score indicates better perceptual quality, ranging from 1 (very unnatural) to 7 (very natural). Figures 8 (a) and (b) depict sample IRT images collected at different ambient weather conditions with excellent and good Brisque scores for their quality categories. Figures 8 (c) and (d) depict fair and poor Pique scores of thermal image samples. Figure 9 (a) depicts a thermal image with poor natural or unnaturalness, and 9 (b) depicts thermal image samples with excellent or good naturalness.



(a)

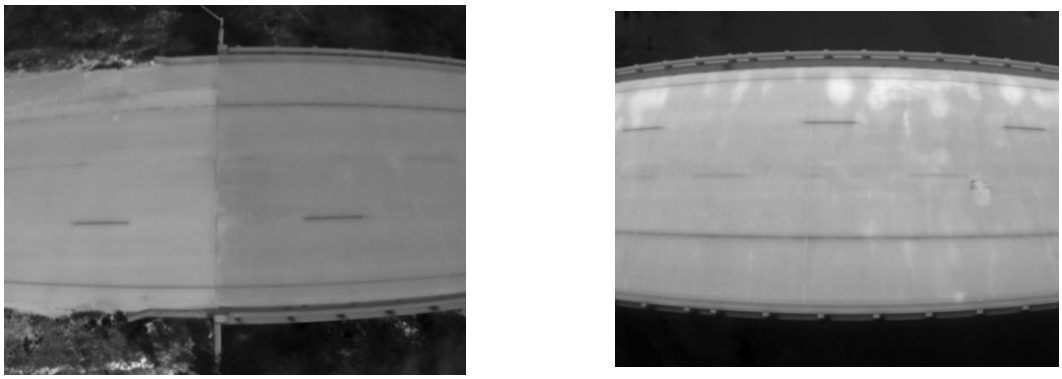


(b)



(c) (d)

Figure 8. Samples of IRT images with (a) Excellent or (b) Good Brisque scores, and (c) fair or (d) poor Pique Scores



(a) (b)

Figure 9. Sample of IRT images with (a) poor natural or unnaturalness, and (b) Excellent or good naturalness

The quality assessment results for IRT, IE, and GPR collected for the studied bridges are summarized in Table 7. The IRT data indicates that Piqe, Niqe, and Brisque quality metrics are satisfactory, with excellent or good values ranging from 82-100%.

Table 7. Summary of quality assessment of NDE data

Metrics	Excellent/Good%	Fair %	Poor/Bad %
Image data quality			
PIQE	97	3	0
NIQE	82	18	0
BRISQUE	100	0	0
GPR and IE signals			
Null Values	100	0	0
Missing values	100	0	0
Duplicate values	100	0	0

2.5 Data Annotation

Data annotation is the process of assigning collected NDE data to one of the defined classes of delamination using the ground truth maps. AI training data must be categorized appropriately and annotated for autonomous bridge deck defect detection. An overwhelming majority of AI models used or applied to bridge deck evaluations rely heavily on the availability of high-quality and accurately labeled training data.

2.5.1 IRT Image Annotation

Image annotation can be completed in three different ways: 1) image labeling, where an entire image is assigned to a specific class, 2) object detection using bounding boxes, where a rectangular box is placed around a group of pixels in each class, and 3) semantic segmentation, where each pixel is assigned to a specific class. Semantic segmentation provides the most information about the data; however, it is the most time-consuming since every pixel must be labeled. A pixel-based semantic annotation method was developed to annotate IRT images autonomously. The primary reason for this methodology was to superimpose the ground truth maps to thermal stitched maps for each bridge. Fast and accurate image annotation in a semantic manner remains a problem in computer vision and related fields; however, the procedure developed in this study can be effectively used to assign delamination classes to each pixel accurately.

Other semantically segmented image datasets rely on image labelers to assign pixels (Wang et al. 2014 and Qin et al. 2018). These methods can be time-consuming and labor-intensive proportional to the level of detail required, and possible inconsistencies between different annotators can occur. The proposed annotation method can effectively remove the role of IRT image labeler.

The primary steps devised for annotating IRT images are depicted in Figures 10 and 11 (a-g). Individual images are stitched together to generate a single-view presentation of each thermal

image in the entire deck. The authors used commercial software, Agisoft 2021 © Professional Version, to create appropriate stitched maps for each bridge deck (Figure 11 (a and b)). Stitched maps were generated by adopting relevant metadata for the set of selected images that produced the highest quality. This process can also be completed using computer vision techniques to remove lens distortion, extract features, and then stitch them together. The generated stitched maps for each bridge were aligned with their corresponding ground truth maps (Figure11 (c and d)). This method requires using the geometrical transformation described in Equations 2 through 4.

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ X & Y & 1 \end{bmatrix} \quad (2)$$

$$R = \begin{bmatrix} cosd & sind & 0 \\ -sind & cosd & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$S = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

where T is Translation, R is Rotation, and S is scale. Equations 2, 3, and 4 determine the Affine transformation matrix for translation, rotation, and scale used. ‘X’ and ‘Y’ are displacements along the x- and y-axis, ‘d’ is the angle of rotation, and ‘a’ and ‘b’ are scale factors along the x- and y-axis. Table 8 summarizes the Affine transformation tuned parameters adopted prior to image registration.

Table 8. Summary of Affine transformation tuned parameters for registration.

Bridge ID	Translation (X, Y) (Pixels)	Rotation (Degree)	Scale
FR_SB	[-474, 220]	-4.5	1.33
FR NB	[-200, 105]	-1.9	1.12
PR_NB	[-105 -5]	-87.4	2.6
PR_M	[0 0]	2.8	1
PR_SB	[-360 -20]	-2	2.88

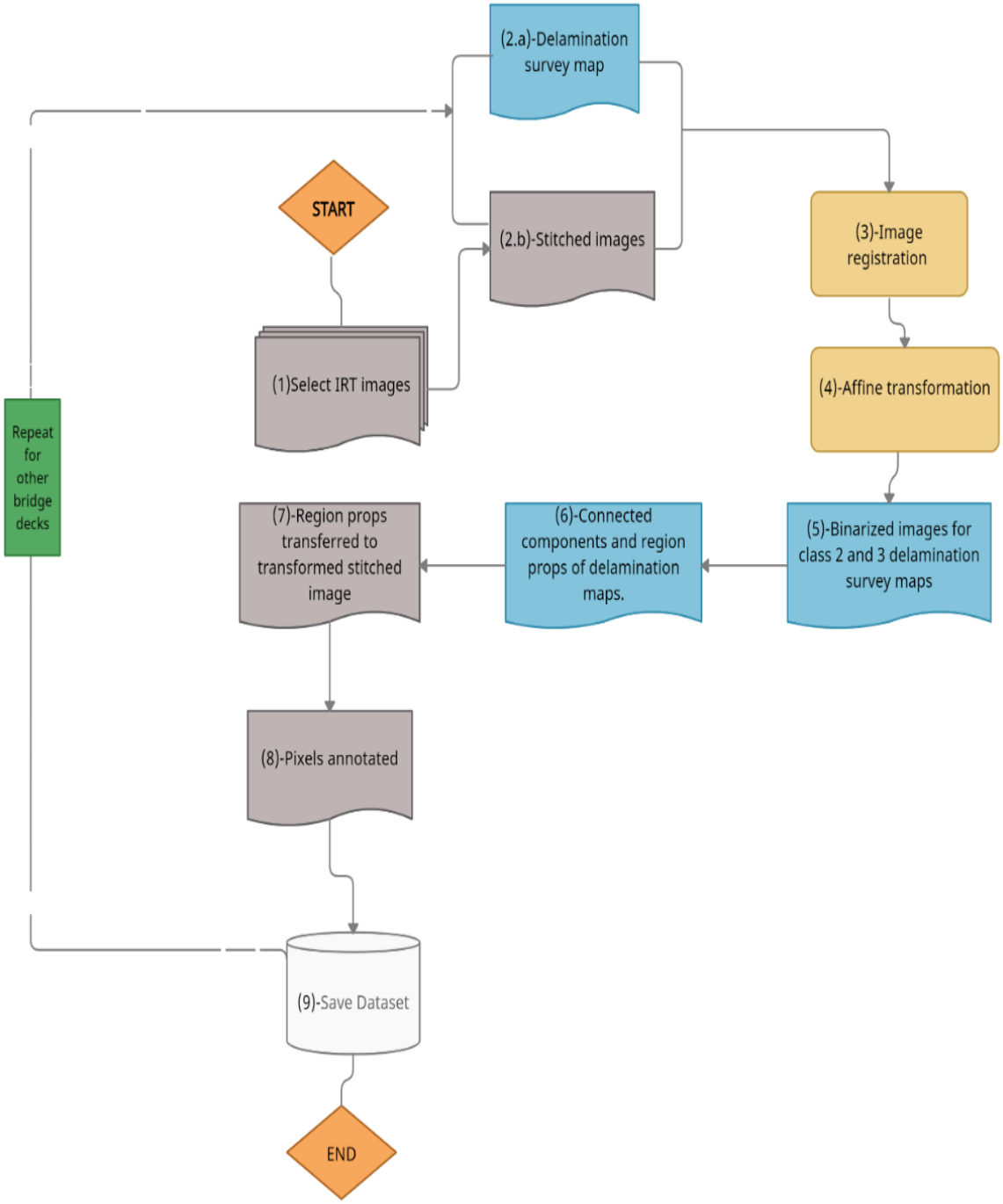


Figure 10. Flowchart for image processing and IRT annotation

An algorithm was developed to register thermal stitched maps against the ground truth images. Affine transformation, including scaling, translation, and rotation matrices were then developed for each bridge through iterations. This process resulted in the generation of a set of stitched maps with the same geometric properties as the ground truth maps (Figure 11b). The original stitched maps were completed in computer aided design format; however, they have been converted to RGB images for computer vision processing. Define all terms on first use.

The location of the pixels within these regions were extracted for each class, resulting in two binary images that represent classes 2 and 3 (Figure 11e-f). The stitched maps were aligned with the ground truth maps; therefore, these pixels depict the actual locations where classes 2 and 3 were removed on the stitched maps (Figure 11e-f), which were later superimposed on the IRT maps (Figure 11g). The pixels in the final image were annotated pixels-wise as G (0 255 0) for all class 2 delamination pixels and R (255 0 0) for class 3 delamination pixels, while the remaining pixels were annotated as sound. The class 2 and 3 pixels are denoted as green and red respectively.

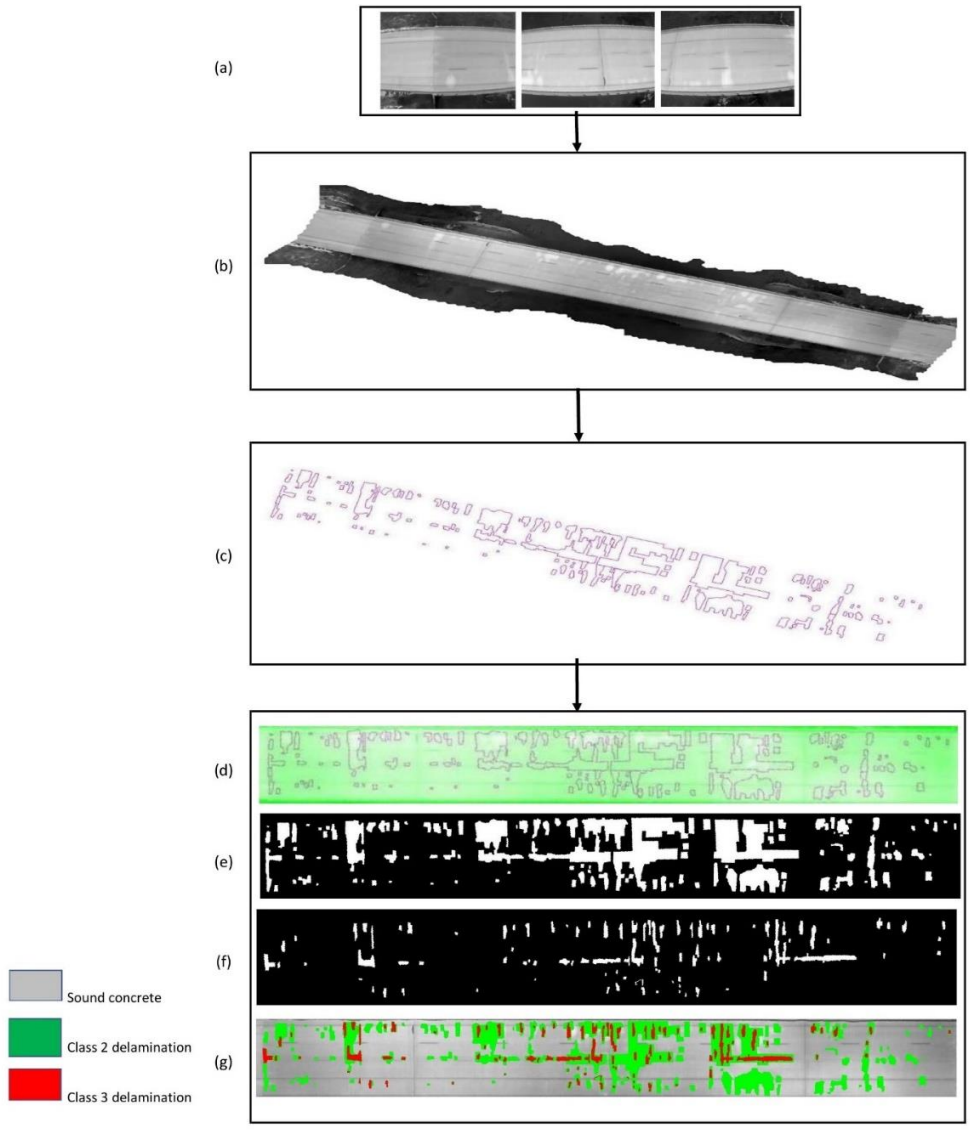


Figure 11. Workflow for annotation (a) input n-number of raw images, (b) stitched image, (c) ground truth survey map for class 2, and (d) aligned and registered ground truth with the stitched map.

2.5.2 IE Annotation

The IE datapoints are point-wise measurements, unlike IRT images. The collected IE data were annotated and validated by cross-referencing the location of the IE signals against the ground truth. Figure 12a-d depicts the IE test regions for the Forest River NB bridge deck layout with removal classes. The IE tests were performed at each point of a set of 3mx3m grids with 0.3m intervals on each bridge. The IE test defined three regions, A, B, and C, on the Forest River North Bound bridge. The exact locations of the IE test points and GPR scans were mapped on the ground truth layout. The class for each signal was determined with the

aid of a computer program, using the ground truth maps for each bridge deck (Figure 12e-g). The annotation output was cross-referenced with the mapped regions on the ground truth layout for consistency.

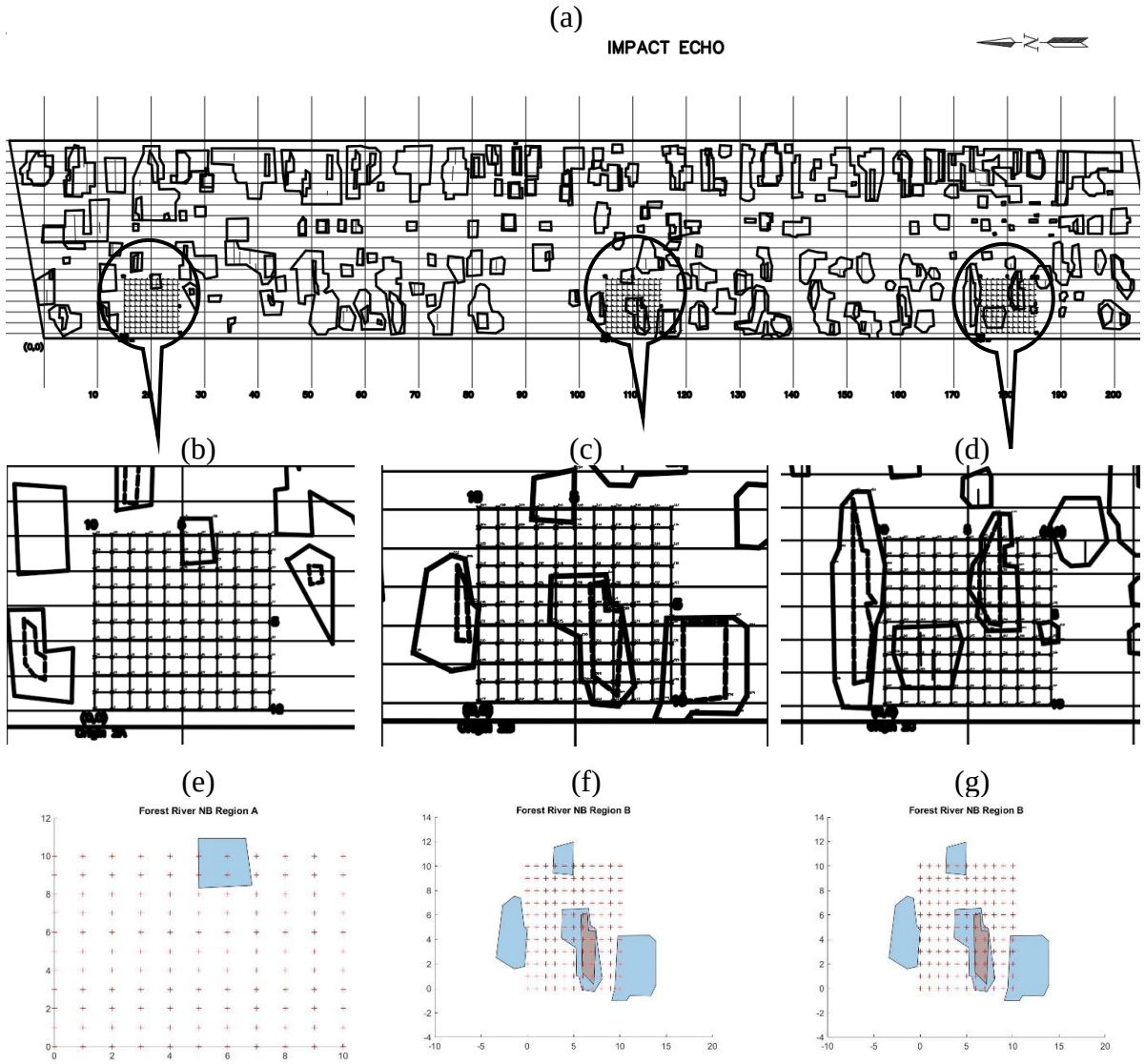


Figure 12. (a) IE test points, (b), (c), and (d) Regions A, B, and C IE test regions on ground truth, (e), (f), and (g) computer outputs of the IE annotation, and (i) computer outputs of the GPR annotation.

2.5.3 Ground Penetrating Radar (GPR) Annotation

A similar approach to IE data annotation was used for GPR signals; however, the GPR device scanned each bridge deck along a designated scan line, unlike the method used for IE. The coordinate of each GPR signal was extracted before annotation using Equations 5 and 6.

$$d_{ux} = \frac{L_x}{n} \qquad d_{uy} = \frac{L_y}{n} \tag{5}$$

$$X_i = X_{x-i} + d_{ux} \qquad Y_i = Y_{y-i} + d_{uy} \tag{6}$$

where,

d_{ux} , d_{uy} = signal sub-divisions for longitudinal and transverse signals, respectively.

L_x , L_y = length of signal scans for longitudinal and transverse signals, respectively.

n = number of signal amplitudes;

X_{x-i} , Y_{y-i} = initial coordinates of the longitudinal and transverse scans, respectively.

X_i , Y_i = cumulative coordinates of the longitudinal and transverse scans, respectively.

The scan lines were plotted on the ground truth maps for each bridge (Figure 13a-b). The intersections of each GPR line scan with the class 2 and 3 delamination regions were extracted automatically with the aid of computer programs and computer-aided design applications.

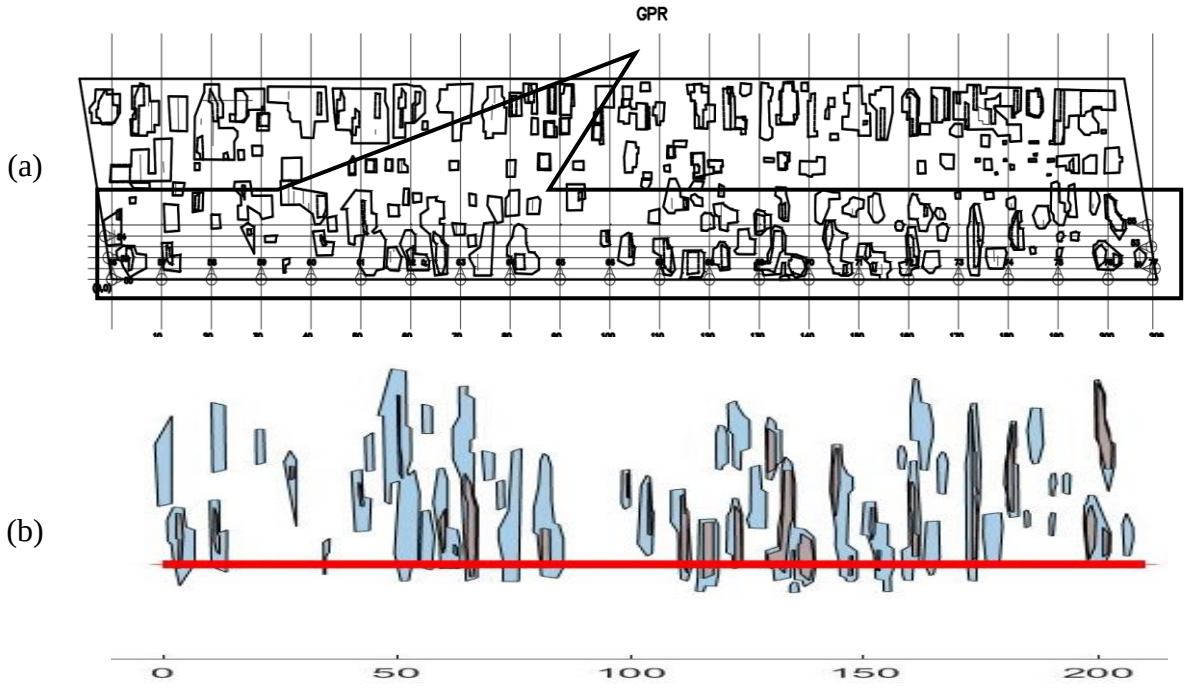


Figure 13. (a) GPR scan signals for Forest River NB, (b) Magnified section of 51 signal scans lines in Forest River NB

A summary of SDNET2021 annotation results is presented in Table 9. The annotation consists of a large dataset with 448,159, 177,483, and 37,460 GPR signals; 1,448, 426, and 62 IE signals; and 2,862,597, 1,252,313, and 465,770 IRT pixels associated with classes 1 (sound), 2 and 3 (delaminated), respectively.

Table 9. Summary of SDNET2021 annotation

GPR						
Classes of delamination	PR_M	FR_NB	PR_NB	PR_SB	FR_SB	Total signals
class 1	171,085	66,334	94,978	61,732	54,030	448,159
class 2	56,528	39,577	26,590	38,510	29,885	177,483
class 3	13,478	6,945	443	6,674	7,392	37,460
Total	241,091	141,500	141,500	106,916	91,307	**663,102**
IE						
Classes of delamination	PR_M	FR_NB	PR_NB	PR_SB	FR_SB	Total signals
class 1	291	301	273	326	257	1,448
class 2	61	49	74	146	96	426
class 3	12	13	16	11	10	62
Total	364	363	363	483	363	**1,936**
IRT						
Classes of delamination	PR_M	FR_NB	PR_NB	PR_SB	FR_SB	Total pixels
class 1	898,758	344,771	802,348	572,455	244,265	2,862,597
class 2	298,544	189,280	215,113	411,147	138,229	1,252,313
class 3	79,294	80,619	49,640	200,968	55,249	465,770
Total Pixels	1,276,596	614,670	1,067,101	1,184,570	437,743	**4,580,680**

2.5.4 Significance and Potential Use of Dataset

Bridges in North Dakota and the US are deteriorating due to the initiation and propagation of surface and subsurface defects that undermine the stability and serviceability of the bridge elements. Data from LTBP Info bridge reveals the condition state of Park River bridge at about 48 years in-service (Figure 14). The condition has been declining steadily since 2015 due to a combination of high average daily traffic, freeze-thaw cycles, and harsh environmental conditions. The current rating for approximately 50% of ND bridges is fair (Figure 15b). Only 54% of ND bridges are in good condition: in the next 20-40 years, over 90% of the existing bridges will be older than 50 years. The condition rating for bridge decks has declined since 2015 from very good, or a condition rating of 8, to fair, or a condition rating of 5. Bridge maintenance, repair budget, and expenditures will rise exponentially if not

well planned; therefore, the NDDOT and other DOT's must prioritize bridge repairs accordingly and monitor deterioration for effective maintenance.

A combination of reliable NDE and AI applications can help bridge stakeholders as they plan for more effective inspections, maintenance, and asset management. This publicly available dataset will help advance studies and research to improve maintenance management systems when prioritizing bridges for repair.

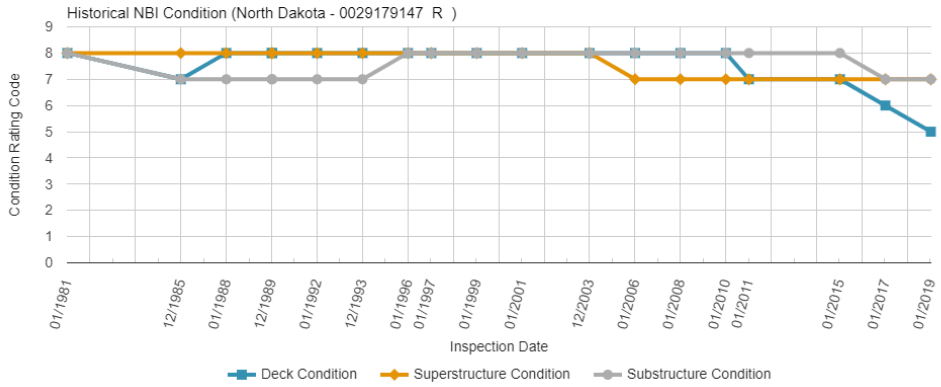


Figure 14. Bridge condition rating history (Source: LTBP FHWA infobridge data)

The role of validated, annotated datasets for AI is critical when benchmarking and developing effective and feasible models with high accuracy.

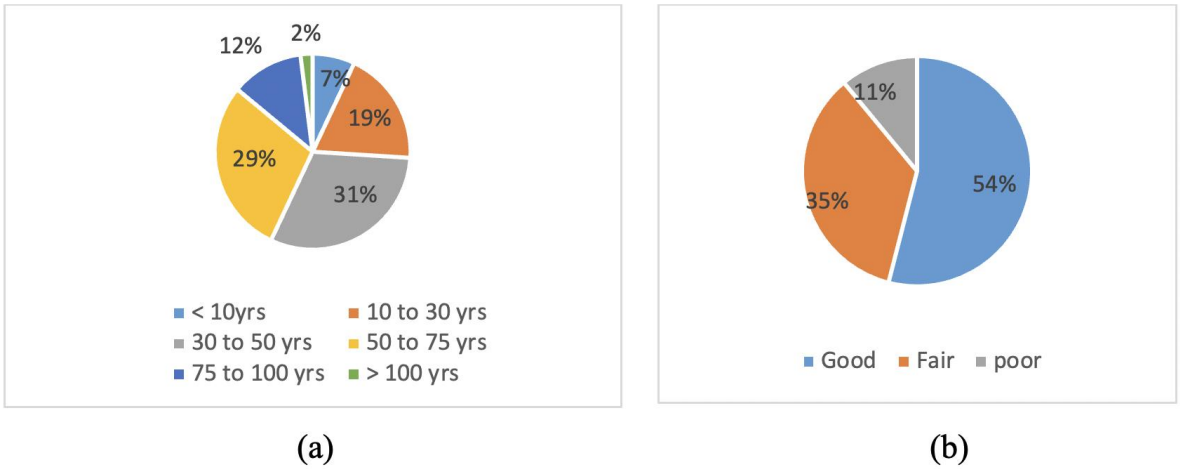


Figure 15. (a) Bridge ages and (b) Bridge condition to FAST Act condition-based performance management system for North Dakota in 2018 (adapted from ASCE, 2019).

Figure 16 demonstrates a delamination survey layout for the Forest River SB bridge. Ground truth can be useful as a benchmark when assigning a more reliable condition rating for the bridge decks compared to the rating developed by NBI and NBE, which is based on visual inspection.

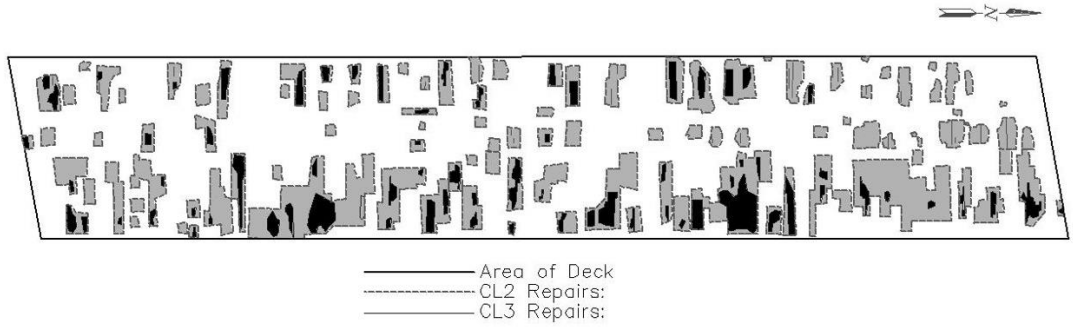


Figure 16. Delamination map for Forest River SB

SDNET2021 will play a significant role in developing and benchmarking AI for bridge deck evaluation using NDE. Some of the applications of this dataset are listed below:

- SDNET2021 will provide data for in-service bridge decks in contrast with most available NDE data generated from laboratory models and specimens.
- The IRT, IE, and GPR datasets have been annotated with reliable and validated ground truth. This dataset will create a benchmark for evaluating bridge deck sub-surface defects.
- The dataset will be essential for the continued development of concrete bridge deck evaluations with the aid of AI models, especially convolutional neural networks (CNNs) training, which is still in its explorative stages. CNNs have significant potential for providing unbiased and inexpensive methods to analyze and interpret bridge evaluation data without operator input, compared to conventional methods.
- The dataset will provide reliable data for investigating the relationships between the concrete deck surface and subsurface defects using AI models.

- The dataset will be a remarkable resource for the development of data fusion for the different types of NDE datasets, allowing researchers to investigate the reliability and precision of one method relative to the other.
- SDNET2021 can be used for algorithm training, validation, and benchmarking for autonomous concrete crack detection.
- A deep learning model trained on SDNET2021 can identify sub-surface delaminations of varying sizes and depths.

2.6 Summary

SDNET2021 contains the following annotated signals from five in-service bridge decks in Grand Forks, ND, USA:

- IE
    - Class 1 - 1,448 sound
    - Class 2&3 - 488 delaminated
- GPR
    - Class 1 - 448,159 sound
    - Class 2&3 - 214,943 delaminated
- IRT image pixels
    - Class 1 - 2,862,597 sound
    - Class 2&3 - 1,718,083 delaminated

All NDE data were collected before the bridges were repaired. The quality of the collected data was evaluated using standard data quality metrics for images and signals. The evaluation indicated that the data presented in SDNET2021 were high-quality.

The dataset was validated and annotated using a set of ground truth maps obtained as the investigated bridges were repaired, illustrating the deck removal class. The delaminated area GPS location, size, and removal class were collected during the repair. The ground truth was

developed on site prior to repair to establish the status of areas as: **Class 1,** sound with no delamination; **Class 2,** shallow delamination with delamination above the top bar mat; and **Class 3,** deeper delamination with delamination below the top bar mat.

IRT images of each bridge deck were stitched together to create an Orthomosaic map. The maps were aligned with the ground truth maps using computer vision techniques. The ground truth maps were analyzed using a developed image processing algorithm to find pixels in each removal class successfully, then each pixel was labeled based on a ground truth map location, providing a unique semantically segmented IRT map for each bridge deck. The ground truth data developed during the investigation has also created the basis for comparing the condition state developed by NBE to the results obtained from our study. The HI of the NBE rating indicates that the investigated ND bridge values range from 97.7% -100%. The HI values is contrary to the motivation for carrying out the repairs work by the NDDOT. Our bridge deck evaluation, based on the generated ground truth delamination survey, yields an HI range of 86-90%, which is more representative of the bridge's condition.

The IE and GPR signal data were annotated and validated with the ground truth maps. The locations of each IE and GPR signal were determined based on the geometrical properties of the decks and data acquisition information. The delamination classes were assigned to each of the 1,936 IE and 663,102 GPR signals. The dataset represents actual on-site conditions such as light sources, disturbances, varying weather conditions, occlusions, and other field conditions expected during the investigation.

SDNET2021 will be a significantly important dataset for further AI development studies to classify and detect concrete bridge deck delaminations and defects. The development of these models is critical for continued research in advanced NDE and structural health monitoring. SDNET2021 is publicly freely.

We have introduced an extensive data set for the Park and Forest River bridges in North

Dakota. The data acquired during the project have been organized and uploaded to the UND repository platform for research data and are available for download. The DOI given in this paper will be permanently linked to the database and will direct users to the latest version of the data. The version history will be visible on the repository, and previous versions will be accessible.

CHAPTER 3. Delamination Detection

3.1 Effectiveness of IRT in Delamination Detection.

This chapter presents the IRT technique adopted for the sub-surface delamination detection of in-service bridge decks and reports the goals, objectives, methodology, model development, and evaluation metrics for delamination detection. This part of the report has been submitted to a peer-reviewed journal.

3.2 Introduction and Review of Past Studies

The demand of NDE techniques for bridge assessment appears to have increased compared to visual and physical inspections (Lee et al. 2014; Rolander et al. 2001). NDE techniques are used to assess and evaluate the condition of bridge decks without human involvement, possibly even autonomously. These evaluations include detecting surface and sub-surface defects, such as corrosion, cracks, delamination, and other anomalies, and evaluating structural properties, such as modulus (La et al. 2017; Lim et al. 2011). Concrete bridges may be evaluated using contact and non-contact NDE techniques, such as Ultrasonic Testing (UT), GPR, IE, IRT, Radiography Testing (RT), and the half-cell method to avoid the shortcomings of visual and physical inspections (Dorafshan et al., 2018b). DOTs and concerned stakeholders are continuously seeking new non-contact techniques (NCT) for automatic defect detection in place of contact techniques (CT). Past studies have indicated that NCTs, such as laser scanning, close range photogrammetry, air-coupled vehicle-mounted NDT, robotic platforms such as Robotic Crack Inspection and Mapping, Robotic Assisted Bridge Inspection Tools, and Unmanned Aerial Systems (UASs) mounted with several spectral sensor types can be used for the NDE of defective bridge decks (Dorafshan et al., 2018b; Gucunski et al., 2014; Dabous et al. 2020).

Unmanned and autonomous bridge inspections are a potentially feasible and effective alternative to physical and manned inspections compared to manned inspections, which are

costly, time-consuming, and labor-intensive (Dorafshan et al. 2018b). UAS and robots used for image or video data acquisition are usually mounted with sensors such as visual or RGB, thermal infrared, Light Detection and Ranging (LiDAR), multispectral and hyperspectral, magnetic, sound navigation and ranging (SONAR), radio detection and ranging (RADAR), or synthetic aperture RADAR (SAR) sensors depending on the mission requirements.

RGB sensors are mounted on a UAS to collect images for crack detection and other surface defects, which can then be evaluated and processed further (Dorafshan et al., 2021). The spectral range of these sensors is in the visible range, from wavelengths of 390–700 nm. Special sensors, such as thermal cameras, are mounted on a UAS to collect thermographs for sub-surface defects that are not within the visible light rays, which can then be evaluated and processed further.

Image-based evaluations require images to be processed and reviewed using several methods, including: 1) raw image inspection, which is usually performed by an experienced inspector, 2) image enhancements, which are typically performed using one or more of several image processing operations, such as filtering, smoothening, denoising, binarization, thresholding, and region of interest to magnify the visibility of delamination within images, and 3) autonomous image processing, which refers to the use of machine learning (ML) algorithms or other artificial intelligence (AI) models to automatically detect delaminations within images.

IRT is an image-based technique that utilizes a UAS mounted with thermal sensors. IRT sensors are used as a non-contact method to detect and identify sub-surface defects such as delaminations and voids. (Hiasa et al., 2016).

Past studies have used thermal images to identify sub-surface concrete delaminations in laboratory specimens based on temperature differences between damaged and undamaged

areas depicted in the image. The false color image temperature matrix is extracted and transformed into grayscale, then normalized to pixel values ranging from 0 to 255 (Montaggioli et al., 2021). The delamination results are presented in plan-view images displaying portions of the defective bridge deck with colored maps or contour lines (Gucunski et al. 2013). Defect detection using IRT is performed using visual, image enhancement, or automatic methods. Table 10 presents a summary of selected studies for IRT delamination detection.

Table 10. Summary of past delamination detection using IRT

References	Nature of Defects	Method of Detection	Data Type/Material or Structure	Validation Method	Limitation
Farrag et al. (2018)	Delamination, voids, cracks, corrosion, honeycombing, and missing bars	Visual	IRT images of 16 nos. of 1.22-mx1.22-mx20-cm reinforced slabs laboratory specimens	Sound laboratory sample	Laboratory specimen used and no ground truth for validation
Omar et al. (2017)	Delamination	Image enhancement (stitching, segmentation)	IRT images (640×512 pix resolution) for two full-scale in-service RC bridge decks in Ontario, Canada	Validated using NDE data (HCP and hammer)	No ground truth for validation
Omar et al (2018)	Delamination	Image enhancement (filtering, enhancement, thresholding, clustering, image registration)	IRT images (640×480 pix resolution) for full-scale in-service RC bridge decks in Quebec, Canada	Validated using IRT data from three bridge decks located in Wisconsin, USA	No ground truth for validation
Montaggioli et al. (2021)	Delamination	Image enhancement (thresholding)	IRT false-color images in temperature matrix converted to grayscale images	Validated with IRT dataset from different concrete structures	No ground truth from the same bridge deck
McLaughlin et al. (2020)	Spalls and delamination	Convolutional Neural Network (CNN)	Platform mounted with RGB and IRT sensors for mobile data collection, generate point cloud maps and sets of synchronized and calibrated images of in-service bridge	Validated with IRT images from a beam/girder concrete structure	No ground truth for the inspected bridge
Gu et al. (2021)	Delamination	Image enhancement and noise reduction algorithm	IRT images of 100×80×30 cm carbon fiber reinforced polymers (CFRP) laboratory specimens	No validation	Laboratory specimen used and no ground truth for validation
Pozzer et al (2021)	Delamination, cracks, spalling, and patches	CNN	Optical and IRT images of a concrete dam in Thunder Bay, Canada	Validated with IRT dataset from two concrete bridge decks in Tapejara, Brazil	No ground truth from the inspected bridge

3.3 Research Goals and Objectives

The primary goal of this research was to develop an automated adaptive image processing-based algorithm to detect subsurface delaminations of in-service concrete bridge decks under ambient environmental conditions.

The following objectives were pursued to achieve this goal:

(a) Develop an optimized image-based processing model with appropriate operations to detect sub-surface delaminations based on pixel gradient.

(b) Benchmark IRT delamination detection with reliable ground truth generated from in-service bridge dimensions, locations, and data.

(c) Consider the effects of delamination depth area, image quality, and ambient weather conditions, such as temperature, humidity, and wind speed, on the model's performance.

3.4 IRT for Delamination Detection

IRT technology is based on the principle of heat flow through two different material media with different thermal properties. The radiant energy can be dissipated through absorption, transmission, and reflection. Materials with null transmissivity and reflectivity are called blackbodies because the radiant energy is absorbed. Most objects with temperatures above absolute zero (-273.15^{o} Celsius or 0 Kelvin) emit infrared radiation (Kuenzer et al. 2013, Madding 1983). Infrared radiation lies between the visible and microwave portions of the electromagnetic spectrum, approximately 0.8 to 14 μm. This range can be further subdivided into the near-infrared (0.8–1.5 μm), short-wavelength infrared (1.5–2.5 μm), mid-wavelength infrared (2.5–8 μm), and long-wavelength infrared (8–14 μm). The intensity of the infrared radiation emitted by an object is a function of the material's temperature and its emissivity, according to the Stefan-Boltzmann equation (Madding 1983).

A perfect black body has an emissivity value of unity, but for concrete the value is typically greater than 0.92 (Chen et al., 2016).

The defects and subsurface delaminations disrupt the heat flow and transfer through the concrete medium, influencing the amount of radiant energy emitted and measured by the IRT sensor. Radiant energy from the sun causes the defective portions to warm up or cool down faster than the surrounding deck due to the difference in the thermal properties of the delaminated and sound portions during the day and at night (Washer et al. 2009). The presence of subsurface delaminations can produce a thermal contrast (ΔT) between delaminated and sound regions. A thermal contrast of 0.2-0.5^{o}C is suitable to detect subsurface delaminations using IRT. A grayscale image is generated from the temperature matrix of the IRT image in false color, then the pixel values are normalized to between 0 (black pixels) to 255 (white pixels) utilizing the camera's parameters. Human IRT image inspection is often subjective and depends on an operator's visual assessment (ASTM-D4788 2013, Hiasa et al. 2017, Montaggioli et al. 2021).

Solar loading, temperature, and wind speed affect delamination detection accuracy when using IRT. Its testing is not suitable at ambient air temperatures less than 0^{o}C and wind velocities greater than 50 km/hr (ASTM-D4788 2013, Washer et al. 2009). Condition and subsurface defect characteristics, such as depth, delamination thickness, materials present in the delamination, concrete properties, and the presence of asphalt overlays are important considerations for IRT testing (Hiasa et al. 2017, Washer et al. 2013).

IRT images were evaluated for sub-surface delaminations using direct prediction and refinement by applying a conditional random field (CRF) method for augmented and non-augmented data. Automatic segmentation based on a deep learning model was adopted to detect delaminations. Direct prediction established values of 15.8, 21.09, and 42.14% for IoU, precision, and recall for non-augmented data, respectively; however, these values

increased to 36.9, 44.01, and 69.61%, respectively, for augmented data (Cheng et al. 2020). An image-based algorithm for automatically detecting delaminations that takes the temperature difference between damaged and undamaged parts in a thermal image into account was developed to evaluate bridge decks. Canny edge detector, local intensity weighing operator, and image processing techniques, among others, were used in deck assessment (Montaggioli et al., 2021). Most of these studies adapted contrast difference as the primary criterion for developing delamination detection models (Table 10).

3.5 IRT Data Acquisition

An FLIR XT V2 infrared thermal camera was used to capture IRT images for the NDE of defective and delaminated sub-surface regions of five in-service reinforced concrete bridge decks in Grand Forks, North Dakota. The thermal sensor was mounted on a DJI Matric 210 UAS at an average altitude of 17m above ground level (AGL). Five mosaic stitched images of the bridge decks were generated using commercial mosaic stitching software. The specifications of the FLIR XT V2 infrared thermal camera used for IRT data collection are summarized in Table 11. The number of pixels computed within the region of interest (ROI) for the delaminated and sound regions of the mosaic-generated IRT images are presented in Table 12. SDNET2021 (Ichi and Dorafshan, 2021a&b) was used in this study. The UAS and IRT camera specifications, weather conditions, data quality, and annotation details are listed in Tables 4, 5,7 & 9

A sample map illustrating delaminated and undelimitated portions, as well as the color representation of the sound and delaminated pixels, is illustrated in Figure 17.

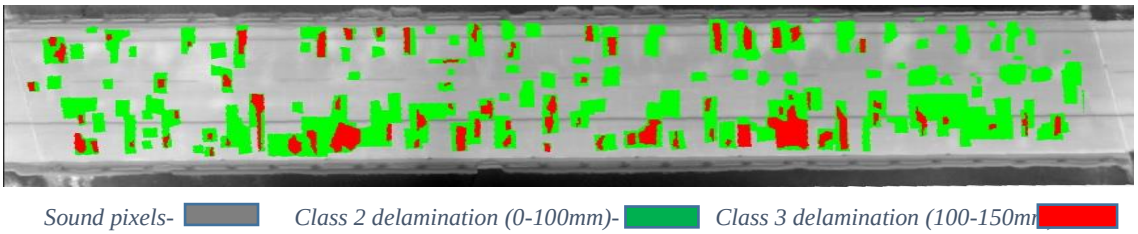


Figure 17. Ground truth map for Forest River SB

3.6 Research Methodology

This section further discusses the experiment, model development, image enhancement, and model performance evaluation metrics sections.

3.6.1 Model Development

All computations were performed on a desktop computer with a 64-bit operating system, 32 GB of memory, and a 3.80 GHz processor running an intel ® Core™ i7-9800X CPU. Image processing operations were programmed and performed in MATLAB R2020a. A novel image processing technique was developed to detect delaminations in the IRT images of the bridge decks. The images collected were benchmarked and validated with ground truth on a pixel-by-pixel basis. Extracting quantitative information, such as the area and location of the delaminated areas, from a set of IRT images is a critical component of thermal data analysis. Developing an automated procedure and algorithm capable of detecting subsurface defects regardless of temperature or pixel intensity range was challenging. The procedure developed to achieve this goal consists of several steps.

The methodology to achieve the objectives of this study were developed using the following MATLAB script:

a) Input image.

b) Read the ground truth and IRT image to perform several initial pre-processing operations, such as the rotation and conversion of the image to grayscale.

c) Pre-processing.

   i) Ground truth: generate a binarized and segmented thermal image of the ground truth by creating a mask based on chosen thresholds for each bridge deck. The ground truth image was converted into a binary coded image of “0” (black for sound pixels) and “1” (white for delaminated pixels) by changing the intensities of the ground annotated pixels.

ii) IRT image: image enhancement operations, such as smoothening, contrast limited adaptive histogram equalization (CLAHE), an adaptive histogram equalization operator, and a fast-local Laplacian filter were applied to an IRT image for higher pixel intensity gradient.

d) Image Processing for delamination detection.

i) Image thresholding and segmentation: the gray image pixels for the entire image were assigned a value between 0 and 1 after pre-processing, which reduced the computation complexity for subsequent computation and segmentation. An adaptative thresholding operation was adopted for segmenting the binary IRT images into two clusters of sound (black, “0”) and delaminated pixels (white, “1”) based on the pixel intensity contrast. The sensitivity factor for adaptive thresholding, specified as the comma-separated pairs consisting of 'Sensitivity' and a number between [0, 1], was iterated and its effect investigated. A stepwise increase of 0.001 was iterated for the sensitivity factor, implying that each image was iterated over 1,001 sensitivity values. A high sensitivity value leads to thresholding more pixels as the foreground at the risk of including some background pixels. An optimum value was determined.

ii) Morphological operations: appropriate morphological operations for further IRT image cleaning and processing were adopted.

iii) Condition map: a condition map for the detected delaminations was created, the number of delaminations by pixel were quantified, and the accuracy and other metrics of the generated model were evaluated.

e) Metric evaluation:

i) Region of interest (ROI): the ROI for the ground truth was created, the IRT image was processed, and the image pixels were mapped.

ii) Quantify delamination and Confusion matrix: the predicted and actual image sets were mapped, and the pixel index numbers were compared. The true positives (TP), true negatives (TN), false positives (FP), and true negatives (FN) were determined for each bridge deck. The developed MATLAB script read every white (delaminated) pixel as 1 (positive pixel), and black (sound) pixels as 0 (negative pixel). Figure 18 depicts the flow chart used for methodology and model development.

If BW_im(t)=1 & BW_gt(t)=1); TP,

If BW_im(t)=1 & BW_gt(t) =0; FP,

If BW_im(t)=0 & BW_gt(t)=1; TN,

If BW_im(t)=0 & BW_gt(t)=0); FN,

BW_im = IRT image for delamination detection; BW_gt = IRT ground truth image for validation; t=pixel index within common ROI.

a) The results from the confusion matrix were used to develop the performance evaluation metrics: i) Accuracy (ACC), (ii) Positive Predictive value/Precision (PPV), (iii) F1-value, (iv) Sensitivity/Recall/True positive rate (TPR), (v) Specificity/True negative rate (TNR), (vi) False Positive Rate (FPR), (vii) False Negative Rate (FNR).
b) Selected sensitivity-values and optimization: The metrics were plotted against the sensitivity (s) values for 1,001 points. Various intersection points from the graphs generated were observed and recorded. An optimized value was selected from the selected sensitivity values. The images for the optimized delamination detection were generated and compared to the ground truth.

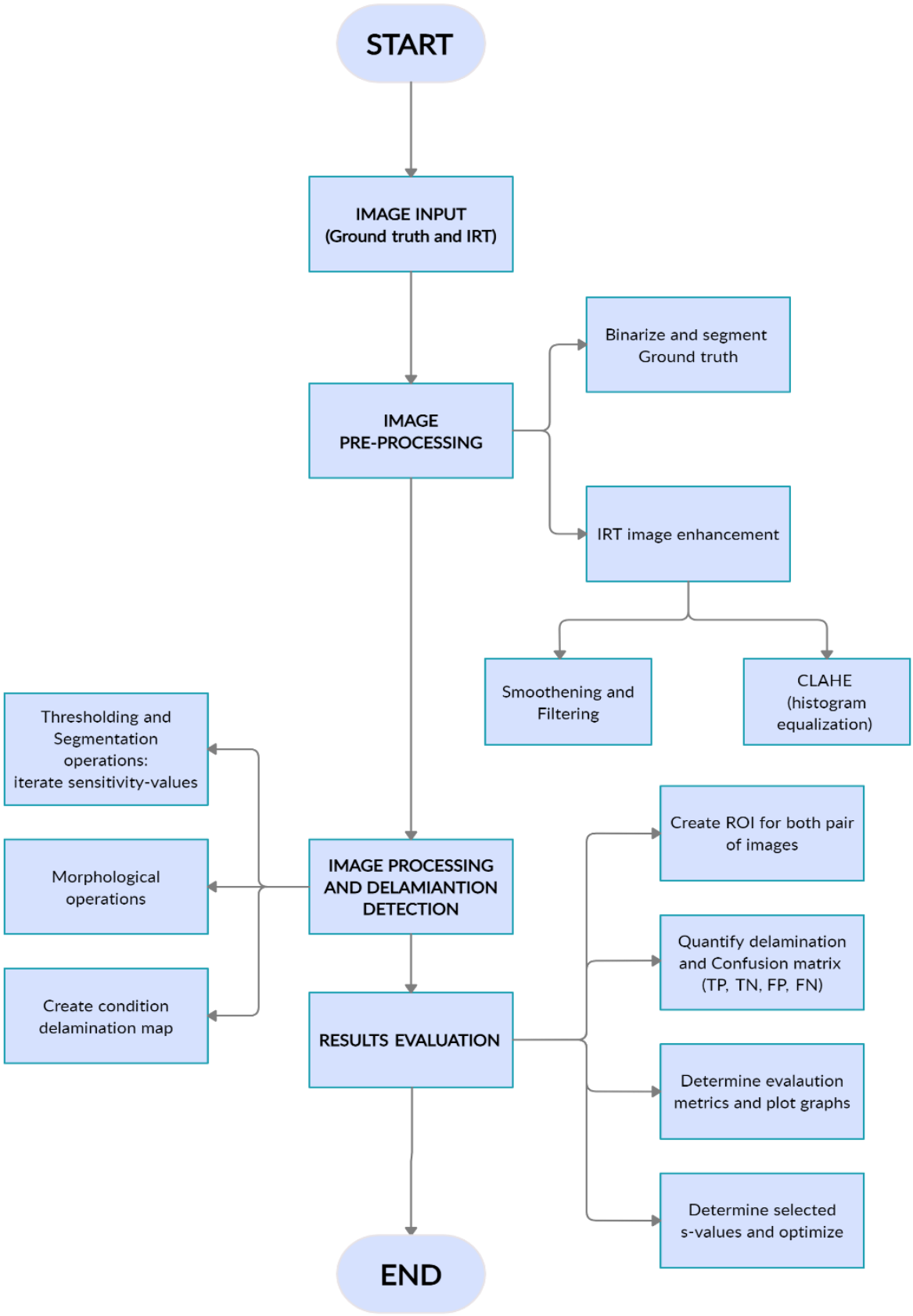


Figure 18. Flow chart for the methodology and model development

3.6.2 Image Enhancements

Several image enhancement operations used in this study were binarization, image thresholding and segmentation, filtering, and Contrast Limited Adaptive Histogram Equalization (CLAHE). CLAHE was introduced to improve contrast and generate an enhanced IRT image, while others were primarily used to segment the features of interest as foreground or background (Hou et al., 2016). Thresholding is one of the most widely used methods, partly due to its effectiveness (Sezgin et al., 2004). Some known thresholding methods include Otsu's method, minimum error method, and maximum entropy method.

The Otsu (1979) thresholding method is a non-parametric and unsupervised automatic threshold selection method for image segmentation. An image f(x,y) of size M×N is represented in L gray levels. Its corresponding average image g(x,y) is defined by Equations (3)-(5).

$$g(x,y) = \frac{1}{k^2} \sum_{a=-(k-1)/2}^{(k-1)/2} \sum_{b=-(k-1)/2}^{(k-1)/2} f(x+a, y+b) \quad (3)$$

We obtain a gray level pair (i,j) for each pixel by denoting the gray level at pixel (x,y) in image f(x,y) and g(x,y) as i and j, respectively. Let F_{ij} be the frequency of the pair (i,j). Its joint probability is given by

$$pi = \frac{\mathrm{Fij}}{\mathrm{MxN}} \quad (4)$$

where i,j=0,1,……..,L-1 and

$$\sum_{i=0}^{L-1} . \sum_{j=0}^{L-1} Pij = 1 \quad (5)$$

The contrast enhancement phase is completed using the contrast limited adaptive histogram equalization (CLAHE), which is a refinement of the adaptive histogram (Zuiderveld et al.

1994). The CLAHE method reduces the noise and edge-shadowing effect produced in homogeneous areas. This method has been used for enhancement to remove noise and reduce the edge-shadowing effect in the pre-processing of digital mammograms (Wang et al. 2005 & Ball et al. 2007).

The histogram of a digital image with intensity levels in the range (0, L – 1) is a discrete function:

$$h(r_k) = n_k \quad (6)$$

Where r_k is the kth intensity value and n_k is the number of pixels in the image with an intensity r_k. A normalized histogram is given by:

$$P_r(r_k) = n_k / MN \quad (7)$$

where k=0,1,2……..., L-1

is an estimation of the occurrence probability for the intensity level r_k in an image. The sum of all components of a normalized histogram is equal to 1.

The histogram equalization is obtained by Equation 8:

$$Sk = (L - 1) \sum_{J=0}^{k} . P_r(r_j) \quad (8)$$

where s_k is the new distribution of the histogram.

3.6.3 Model Evaluation Metrics

The model was evaluated with the performance evaluation metrics of (i) Accuracy (ACC), (ii) Positive Predictive Value/Precision (PPV), (iii) F1-value, (iv) Sensitivity/Recall/True Positive Rate (TPR), (v) Specificity/True Negative Rate (TNR), (vi) False Positive Rate (FPR), and (vii) False Negative Rate (FNR).

The metrics are calculated using Equations (9)-(13):

$$ACC = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \quad (9)$$

$$PPV = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (10)$$

$$F1 - Score = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \quad (11)$$

$$TPR = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (12)$$

$$TNR = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \quad (13)$$

The model's ACC indicates the correction detection rate for the delaminations with respect to the total detections. The PPV and TPR are metrics often used to increase the understanding of the prediction outcome. PPV measures what fraction of the detected delaminations have been correctly detected given the ground truth, and TPR indicates what fraction of the real damages are correctly detected among the positives (Kumar et al. 2018). Low PPV rates indicate a high number of false positives, such as when many areas are incorrectly classified as delaminated. Models with high PPV results are preferred to minimize false detections. Low TPR rates indicate a high number of false negatives, such as when many of the existing delaminations were missed.

3.7 Results and Discussions

A summary of the results for the image-based delamination detection for the five in-service bridges is presented in Table 14. The evaluation metrics for the developed model are pixel-based, unlike other image-based studies that visually evaluate delaminations by comparing the pixel contrast between the delaminated and un-delaminated portions. This study also evaluates and validates the delamination detection of the studied IRT images that are well annotated and with reliable ground truth data. Table 11 presents the performance evaluation

metrics for selected optimized s- values with corresponding intersecting curves: ACC, PPV, F1-Score, TPR, TNR, FPR, FNR.

Table 11. Summary of delamination metrics

Bridge Name	Intersecting Curves	Nominated s-value	ACC	PPV	F1-score	TPR	TNR	FPR	FNR
FR NB	TPR/PPV/ F1 score	0.452	0.703	0.254	0.211	0.180	0.851	0.149	0.820
	FPR/PPV	0.464	0.699	0.262	0.228	0.201	0.840	0.160	0.799
	FPR/F1 score	0.486	0.691	0.275	0.260	0.248	0.816	0.184	0.753
	TPR/FNR	0.560	0.626	0.276	0.337	0.433	0.680	0.320	0.567
FR SB	TPR/PPV/ F1 score	0.459	0.716	0.140	0.099	0.077	0.879	0.121	0.924
	FPR/TNR	0.581	0.5634	0.202	0.265	0.386	0.609	0.391	0.614
	TPR/TNR/ ACC	0.584	0.560	0.204	0.270	0.399	0.601	0.394	0.601
	FNR/ F1 score	0.625	0.500	0.224	0.325	0.590	0.477	0.523	0.410
PR MD	TPR/PPV/ F1 score	0.380	0.655	0.141	0.131	0.123	0.799	0.201	0.877
	FPR/TPR	0.544	0.564	0.207	0.267	0.373	0.616	0.384	0.627
	TPR/FNR	0.545	0.615	0.240	0.293	0.375	0.680	0.321	0.625
	ACC/FNR	0.566	0.549	0.213	0.283	0.420	0.584	0.417	0.581
	FNR/TPR	0.580	0.593	0.248	0.319	0.450	0.631	0.369	0.550
PR NB	TPR/PPV/ F1 score	0.480	0.696	0.330	0.287	0.254	0.836	0.164	0.746
	FPR/PPV	0.514	0.666	0.310	0.314	0.315	0.778	0.223	0.685
	FPR/ F1 score	0.532	0.646	0.3000	0.324	0.352	0.739	0.261	0.648
	TPR/FNR	0.559	0.607	0.283	0.335	0.411	0.670	0.330	0.589
	TPR/TNR/ ACC	0.572	0.587	0.277	0.341	0.443	0.632	0.368	0.557
	FPR/FNR	0.572	0.587	0.277	0.341	0.443	0.632	0.368	0.557
PR SB	FPR/PPV	0.345	0.740	0.156	0.134	0.117	0.869	0.131	0.883
	TPR/PPV/ F1 score	0.365	0.731	0.161	0.148	0.136	0.853	0.147	0.864
	TPR/TNR/ ACC	0.580	0.529	0.169	0.248	0.446	0.546	0.454	0.554
	FPR/FNR	0.580	0.529	0.169	0.248	0.446	0.546	0.454	0.554
	ACC/FNR	0.590	0.517	0.169	0.248	0.465	0.527	0.473	0.535

The performance metrics for the bridge decks were plotted against the s-values, which varied from 0 to 1.0 at 0.001 steps. The generated curves are illustrated in Figures (19)-(23).

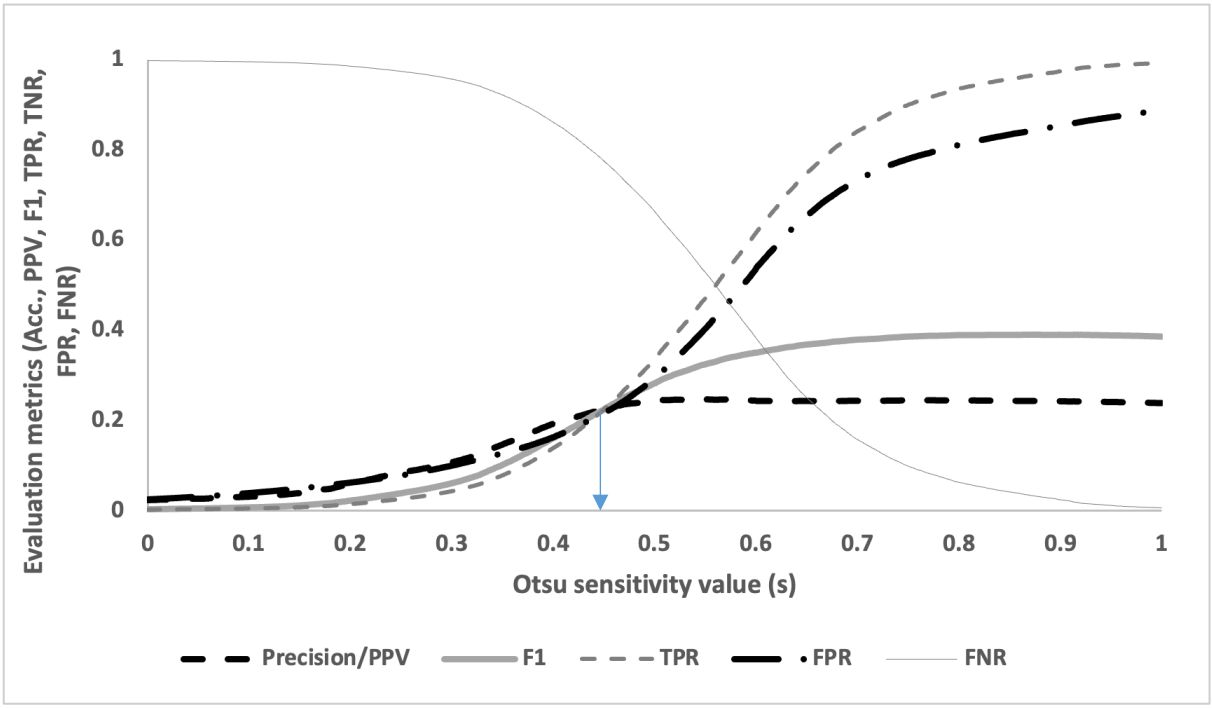


Figure 19. Forest River NB evaluation metrics vs. sensitivity values

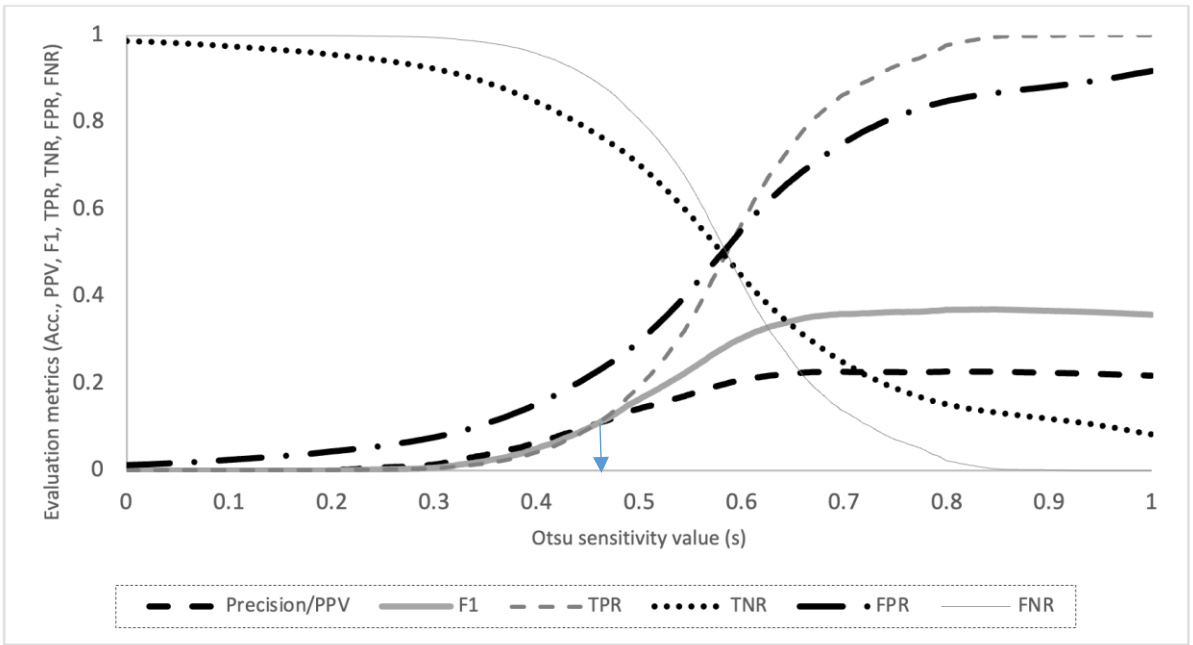


Figure 20. Forest River SB evaluation metrics vs. sensitivity values

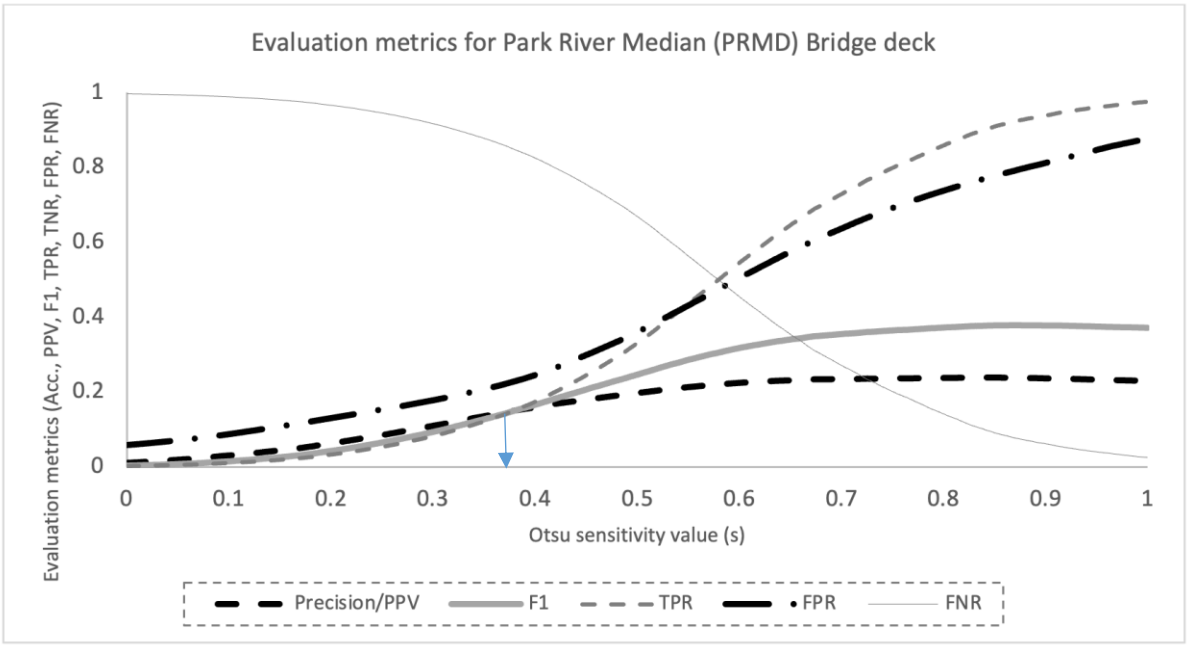


Figure 21. Park River Median bridge evaluation metrics vs. sensitivity values

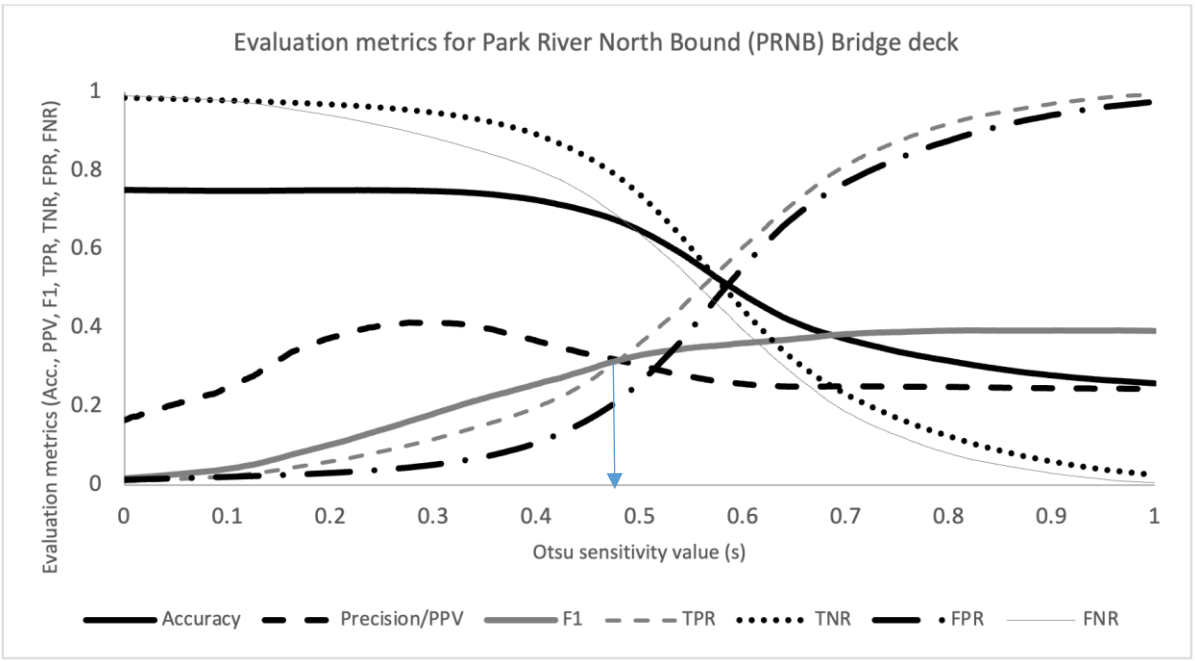


Figure 22. Park River NB bridge evaluation metrics vs. sensitivity values

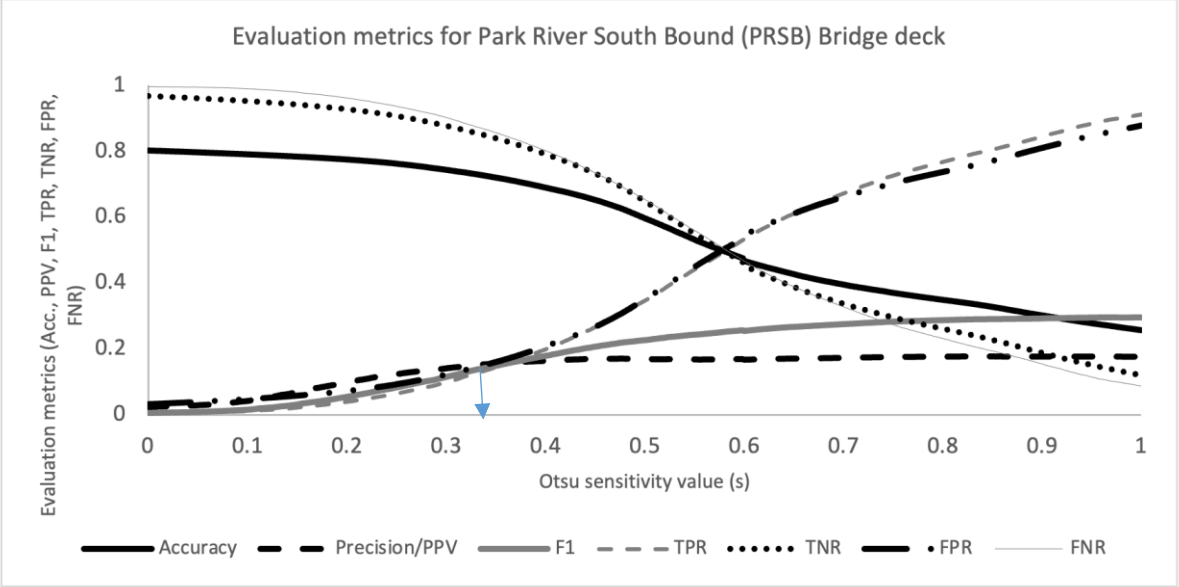


Figure 23. Park River SB bridge evaluation metrics vs. sensitivity values

The performance metric curves exhibit a similar sigmoid curve pattern. The FNR, TNR, and ACC decrease from 1 to 0 as the s-values increase from 0 to 1. The TPR and FPR increase from 0 to 1, conversely. The F1-score increases from 0 to a maximum steady value of 0.4, while the PPV slopes upwards to a maximum value of 0.4.

The bridges' selected s-values were plotted against the corresponding accuracies (Figure 24). This plot shows a correlation between the parameters, with an R-squared value of 0.7588. This shows a possible relationship between the s-values and accuracy of the model.

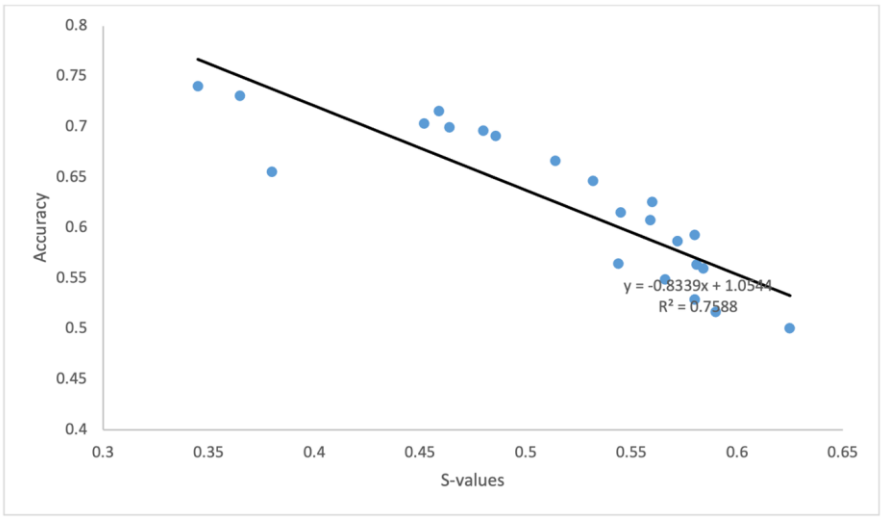


Figure 24. Graph of selected s-values for bridge decks vs. accuracy

3.7.1 Optimized S-Value(s) Selection

The s-value is a parameter in the adaptative Otsu thresholding operation that determines the number of pixels segmented as foreground and background. The s-values for each bridge were selected at various intersecting points of the various curves. The image with a higher sensitivity has higher pixels segmented as foreground or sound pixels (Figure 25); therefore, a higher sensitivity value leads to thresholding more pixels as foreground at the risk of including some background pixels, implying that higher s-values may not yield acceptable results since it may result in increased FP.

The s-vales were selected for FR NB where TPR and PPV intersect, FPR and PPV intersect, FPR and F1 scores intersect, and TPR and FNR intersect at s-values of 0.452, 0.464, 0.486, and 0.56, respectively. The s-values and intersecting curves for other bridges are depicted as well in Figure (21-23). The s-values for the models vary from the lowest value of 0.365 for PR MD to the greatest value of 0.625 for FR SB.

The intersection of the TPR, PPV, and F1-score curves was common for all five bridges evaluated. The s-values at these points were extracted and investigated further: 0.452, 0.459, 0.380, 0.480, and 0.365 for FR NB, FR SB, PR MD, PR NB, and PR SB, respectively. The PR SB had the lowest selected s-value of 0.365, while the PR NB had the highest value of 0.48. Other bridges possessed similar s-values at this point of intersection.

The images for the values generated by the model are presented for FR NB and FR SB in Figures 26 (1a-2c).

(a)

(b)

(c)

(d)

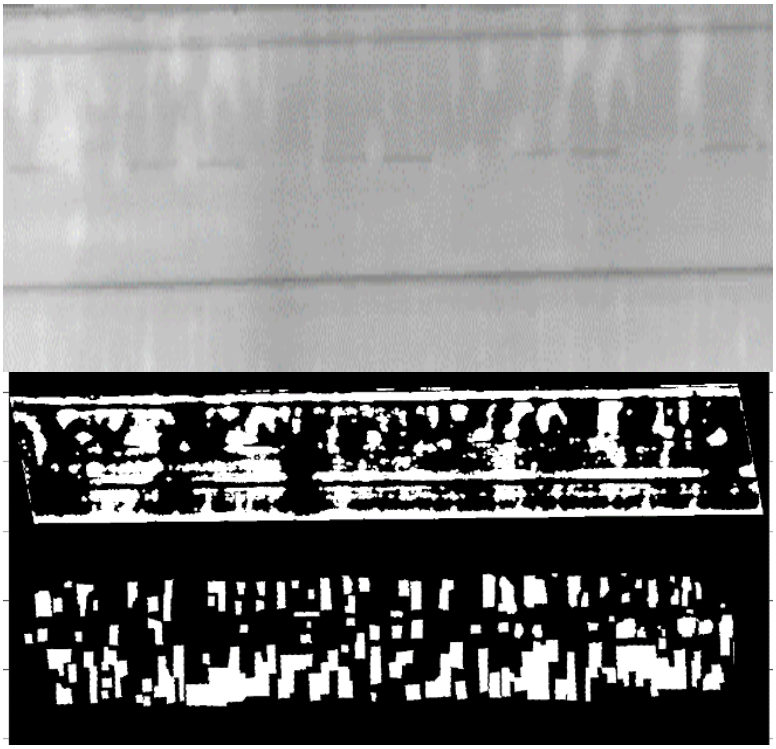
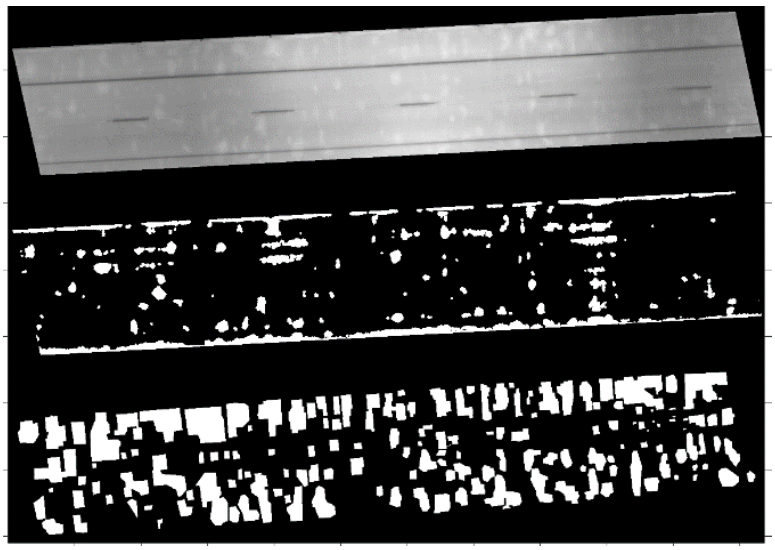


Figure 25. (a) IRT image for delamination detection, delamination detection for (b)s-value=0.459, (c) s-value=0.58, and (d) binarized ground truth.

1(a)

1(b)

1(c)



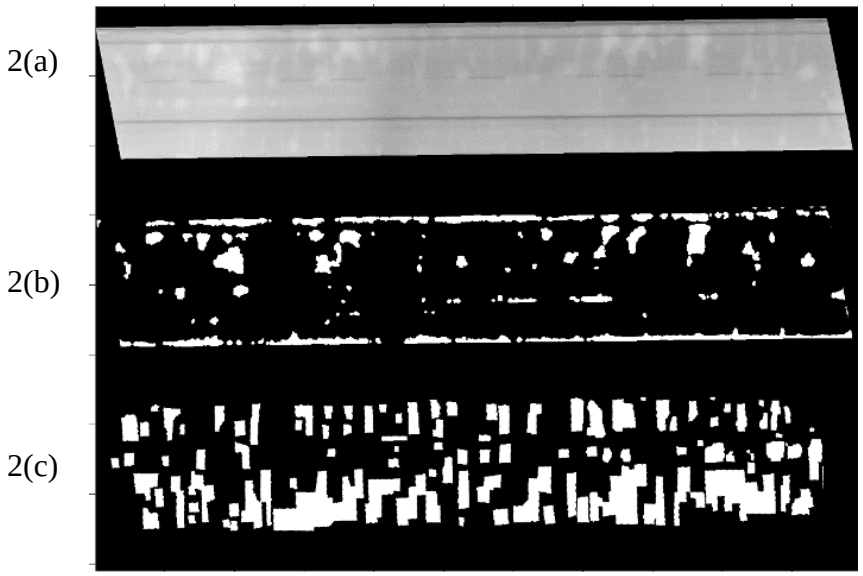


Figure 26. (1a&2a) IRT images for FR NB and FR SB, (1b&2b) Condition maps for 0.452 and 0.459 s-values, and (1c&2c) Ground truth images for FR NB and FR SB.

The selected s-vales for the studied bridges are the lowest, compared to the s-values generated at other points of intersection, except for the PR SB bridge deck, which had a value that was 0.02 higher than the lowest value. The FR NB s-value of 0.452 was the lowest for the intersecting curves of TPR, PPV, and F1-score compared to 0.464 for FPR and PPV, 0.486 for FPR and F1-score, and 0.56 for TPR and FNR. The selected optimized values for each bridge deck are depicted in Figure 27.

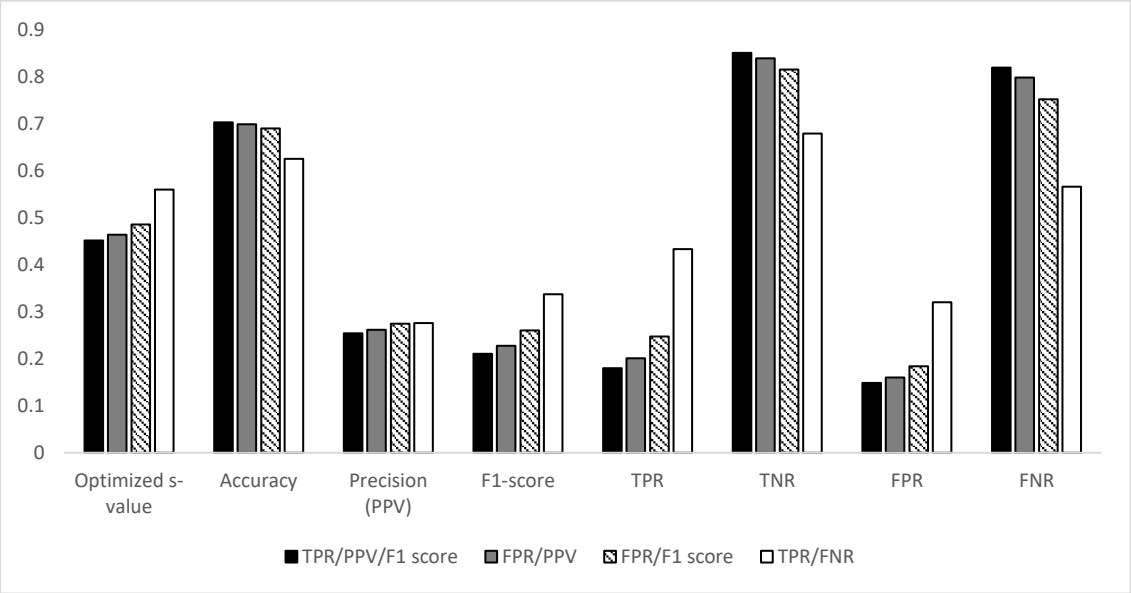


Figure 27. Selected s-values and evaluation metrics for Forest River NB

The FR SB s-value of 0.459 was the lowest at the intersecting curves TPR, PPV, and F1-score compared to 0.581 for FPR and TNR, 0.584 for TPR, TNR, and ACC, and 0.625 for FNR and F1-score, which is a similar trend for the PR MD, PR NB, and PR SB bridges.

3.7.2 Selected S-Value(s) and Evaluation Performance Metrics

We also considered the effect of optimized s-values on the performance evaluation metrics when selecting the value used. An increased s-value segments more pixels to the foreground, likely causing an increase in FP. The confusion matrix for the selected s-values of the five bridges investigated is presented in Table 12(i)-(v). The TP, FP, TN, and FN of the actual IRT image are listed in reference to the ground truth.

Table 12. Summary of confusion matrix for selected s-values

(i) Confusion matrix for s=**0.452** for FR-NB

		Predicted IRT pixels		
		1	0	
Actual Ground truth pixels	1	35,284 4%	160,372 18%	Total= 886,271 pixels
	0	102,825 12%	587,790 66%	

(ii) Confusion matrix for s=**0.459** for FR-SB

		Predicted IRT pixels		
		1	0	
Actual Ground truth pixels	1	10,088 2%	121,918 19%	Total= 647,077 pixels
	0	83,180 13%	431,891 67%	

(iii) Confusion matrix for s=**0.38** for PR-MD

		Predicted IRT pixels		
		1	0	
Actual Ground truth pixels	1	46,294 3%	350,913 20%	Total= 1,775,027 pixels
	0	281,625 16%	1,096,195 62%	

(iv) Confusion matrix for s=**0.480** for PR-NB

		Predicted IRT pixels		
		1	0	
Actual Ground truth pixels	1	67,347 6%	197,408 18%	Total= 1,100,012 pixels
	0	137,013 12%	698,244 63%	

(v) Confusion matrix for s=**0.3650** for PR-SB

		Predicted IRT pixels		
		1	0	
Actual Ground truth pixels	1	43,762 2%	277,084 15%	Total= 1,873,838 pixels
	0	227,879 12%	1,325,113 71%	

The change in the selected optimized s-values and their effect on the confusion matrix parameters was evaluated. A 2.7% increase in the FR NB s-value caused an increase in the TP and FP by 11% and 8%, respectively, and a reduction in FN and TN by 3% and 1%, respectively. A 27% increase, from 0.459 to 0.581, for the FR SB s-value resulted in a 405% and 142% increase in TP and FP, and a 34% and 127% reduction in FN and TN, respectively. There was also a significant decrease in the ACC values, from 0.716 to 0.563, resulting in a 15.3% reduction. An increase in the s-values for the PR MD bridge, from 0.38 to 0.544 (43%), caused an increase in TP and FP of 203% and 91%, respectively, and a 33% and 21% reduction in FN and TN, respectively. An increase in the PR NB s-value, from 0.48 to 0.514 (7%), caused an increase in TP and FP of 24% and 36%, respectively, and an 8% and 7% reduction in FN and TN, respectively. An increase in the PR SB bridge s-value, from 0.365 to 0.58 (59%), caused an increase in TP and FP of 227% and 210%, respectively, and a 36% reduction in FN and TN.

Changes in s-values caused significant changes in the performance evaluation metrics (Table 11). A 7.5% increase in the FR NB s-value, from 0.452 to 0.486, caused no significant reduction in ACC (1.8%). The F1-score, TPR, and FPR increased significantly from this change in s-value, by 24%, 38%, and 24%, respectively, while TNR and FNR reduced by 4% and 8%, respectively; therefore, a selected value of 0.486 was more appropriate since it caused a significant increase in TPR and FI-score and a drop in TNR, while ACC was relatively the same even though the value of 0.452 was the optimized value selected, with an ACC value of 70%. There was an increase in the FR SB F1-score, TPR, and FPR (173%, 422%, and 226%, respectively), while ACC, TNR, and FNR were reduced by 21.3%, 32%, and 35%, respectively. There was an increase in the PR MD F1-score, TPR, and FPR (123%, 205%, and 59%, respectively), while ACC, TNR, and FNR were reduced by 14%, 15%, and 29%, respectively. There was also an increase in the PR NB F1-score, TPR, and FPR (13%,

38%, and 59%) while ACC, TNR, and FNR were reduced by 4.3%, 11%, and 13%, respectively. There was an increase in the PR SB F1-score, TPR, and FPR (68%, 227%, and 210%, respectively), while ACC, TNR, and FNR were reduced by 1.3%, 36%, and 36%, respectively. The model's average ACC was 0.6976; therefore, the optimized s-values selected for FR SB, PR MD, PR NB, and PR SB were 0.459, 0.38, 0.48, and 0.365, respectively.

An R-squared (R^2) value of 75.88% was calculated after the range of the selected s-values were evaluated and compared to the accuracy, suggesting that the model was a relatively good fit for evaluating the model's ACC (Table 13). A modified ACC value can be calculated and determined using the generated graph and model depicted in Figure 24 (Table 13).

There is a slight variance in the modified ACC values generated using the model but with similar average values, implying that the ACC for the detection model is approximately 69%.

Table 13. Modified accuracy values from accuracy vs. s-values graph

	Optimized s-value	ACC	Modified ACC
FR NB	0.486	0.6907	0.6491
FR SB	0.459	0.7156	0.6716
PR MD	0.38	0.6552	0.7375
PR NB	0.48	0.696	0.6541
PR SB	0.365	0.7305	0.7500
Average		0.6976	0.6926

3.7.3 Factors Affecting the Performance of the Model

It is important to consider possible factors that may affect the model's performance for all bridges evaluated. The optimized s-values indicate a slight variance between FR NB, FR SB, and PR NB, while PR MD and PR SB exhibit similarly optimized s-values, indicating the relative uniformity of the model when detecting defects with two clusters. The average s-values for these clusters is 0.475 and 0.3725.

Several possible factors may have influenced the model's performance: depth of delamination; spatial dimension of delamination; ambient weather conditions such as wind speed, temperature, and humidity; and quality of the original and mosaic images (Table 14).

Table 14. Summary of factors affecting delamination detection

Classes of Delamination	FR_NB (1)		FR_SB (2)		PR_MD (3)		PR_NB (4)		PR_SB (5)	
class 1 (Sound)	344,771	56%	244,265	56%	898,758	70%	802,348	75%	572,455	48%
class 2	189,280	31%	138,229	32%	298,544	23%	215,113	20%	411,147	35%
class 3	80,619	13%	55,249	13%	79,294	6%	49,640	5%	200,968	17%
Total Pixels	614,670	100%	437,743	100%	1,276,596	100%	1,067,101	100%	1,184,570	100%
Optimized s-value	0.486		0.459		0.38		0.48		0.365	
Performance Evaluation Metrics										
ACC	0.6907		0.7156		0.6552		0.6960		0.7305	
Depth of Delamination										
% Class 3:2 ratio	0.426		0.400		0.266		0.231		0.489	
Ambient Weather Conditions										
Temperature (ºC)	26.7		26.0		27.8		27.0		27.8	
Humidity (%)	44.0		47.0		44.5		47.0		45.0	
Wind Speed (kmph)	12.9		10.5		14.5		12.9		16.1	
Image Quality										
Brisque	44.0 (Fair)		49 (Fair)		48 (Fair)		40 (Fair)		34 (Good)	
Piqe	68.0 (poor)		81 (Bad)		72 (Poor)		36 (Fair)		39 (Fair)	

An evaluation of the original reference image revealed Brisque and Pique results of 24 and 31, respectively; therefore, they are categorized as 'good' based on the evaluation metrics (Taylor et al., 2020). An evaluation of the mosaic image's quality, when compared to the original frames generated, indicated that the quality had decreased, from 'good' to 'poor/bad.'

The effect of wind on the model's performance was evaluated with respect to accuracy. A plot of the accuracy and ambient wind speed is illustrated in Figure 28. The plot indicates an inverse relationship with a high R-squared value. Similar studies completed by Raja et al. (2020) indicate that wind significantly impacts absolute thermal contrast (ΔT) development,

demonstrating that an increase in wind velocity resulted in a decreased thermal contrast. The Figure 28 indicates that the highest accuracy was derived from the lowest recorded wind speed

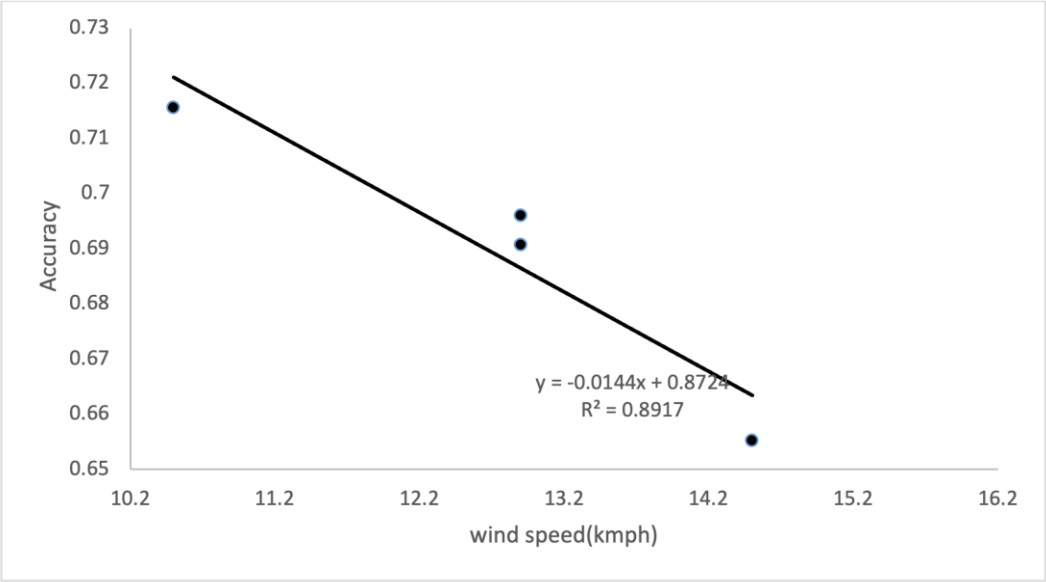


Figure 28. Effect of wind speed on model accuracy

The ambient temperature for the 2-bridge cluster, PR MD and PR SB, is 27.8 ^{o}C (Table 17). This cluster's wind speed and humidity range are 14.5-16.1 kmph and 44.5-45%, respectively. The values for the 3-bridge clusters, FR-NB, FR_SB and PR-MD, range from 26-27 ^{o}C, 44-47%, and 10.5-12.9 kmph for the temperature, humidity, and wind speed, respectively. Ambient weather conditions affected the s-values adopted for the studied bridges compared to the optimized s-value range for the 2-cluster and 3-cluster bridges (0.365-0.38 and 0.459-0.48, respectively).

The effect of the delamination depth area was also evaluated. The class 3:class 2 values for all studied bridges, in ascending order, are 0.231, 0.266, 0.40, 0.426, and 0.489 for PR NB, PR MD, FR SB, FR NB, and PR SB, respectively (Table 8). Figure 29 illustrates a positive R^2 value trend of approximately 50%, possibly due to the likelihood of the class 3:class 2 delamination proportion on the model's ACC.

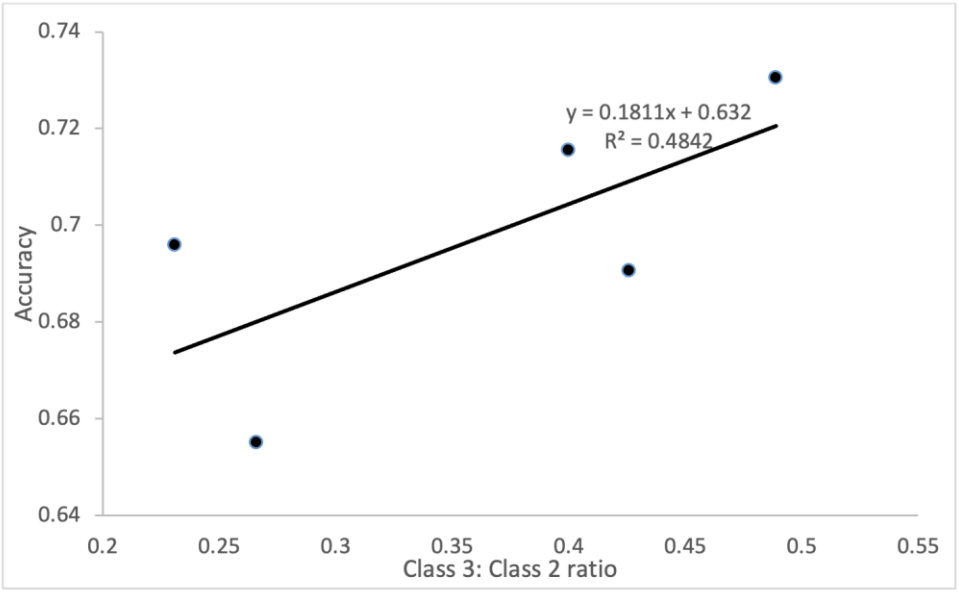


Figure 29. (a)&(b) Effect of delamination depth area on the model's performance

Similar past studies have revealed that the delamination size and the depth are the critical factors for IRT detection. Some studies indicate that as the size of delamination increases, the temperature gradient between the sound and delaminated pixel increases, improving the detectable depth of the sub-surface delaminations (Washer et al. 2010, Maierhofer et al. 2005, Cheng et al. 2008). Tran et al. (2017) also demonstrated that delamination detection increases with increased ambient temperature in large and shallow delamination than small and deep delaminations, and might be greater under a highly humid atmosphere.

Delamination size also has a significant effect on these outcomes. A 32 mm deep delamination (14.5mm^2) was not detected in the work completed by Abdel-qader, et al. (2008). Washer et al. (2010) revealed that a delamination area of 900 mm^2 was detected up to a 7.6 mm depth but was not detected at 127 mm. Most of the delamination studies were based on laboratory specimens and may not be indicative of results from field investigations, which are influenced by several factors such as occlusion, background interference, light control, quality and resolution of the image, and ambient weather conditions.

3.7.4 Summary

This study presents an adaptative image processing-based model for the sub-surface bridge deck delamination evaluation of five in-service concrete bridge decks in North Dakota. The proposed method is conceptually different than most of the recent image-based algorithms and models developed, which did not use reliable ground truth or were primarily based on laboratory specimens. We have investigated the effectiveness of IRT in delamination detection with reliable and validated ground truth data. The developed model was optimized by iterating the sensitivity (s) parameter and selecting s-values based on the interactions of the performance evaluation metrics: Accuracy (ACC), Positive Predictive value (PPV), F1-value, True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), and False Negative Rate (FNR). The 2-clustered and 3-clustered optimized s-values ranged from 0.365 to 0.38 and 0.459 to 0.486, respectively. An average accuracy of 69% was obtained for the model. The true positive rate and true negative rate ranged from 0.077 to 0.254 and 0.799 to 0.879, respectively. The study revealed that several factors, such as the depth and spatial dimensions of the delaminations; ambient weather conditions such as wind speed, temperature, and humidity; and mosaic image quality affect the model's performance. Our results indicate that adopting deep learning and data fusion methods to improve the model's accuracy and performance regarding delamination detection is promising.

One known limitation of the model is the adoption of low-resolution and quality mosaic IRT images. Investigating the use of a higher quality image with increased resolution may improve the model's performance, which will be necessary for future studies.

Data fusion of IRT with visual/RGB images or other NDE methods would be instrumental in developing and improving models for delamination detection. Past studies indicate that delamination can be observed by GPR and IRT and can be used together for accurate anomaly identification (Maser et al. 2009, Washer et al. 2013). Future work could explore

and combine other NDE data, such as visual images, impact Echo, and GPR with IRT data in the SDNET2021 dataset to increase TPR and ACC. Further studies could establish the effect of delamination depth by considering different delamination classes.

Exhaustively trained models are increasingly available as deep learning (DL)-based methods increase in popularity. Re-using these models and combining them with image processing methods to optimize delamination detection may be more feasible. The results from this study indicate that the future adoption of convolutional neural network (CNN) methods for image-based delamination detection in concrete bridge decks is promising.

CHAPTER 4. Conclusions and Future Work

4.1 Conclusions

Advanced non-destructive evaluation (NDE) techniques are feasible and established methods for bridge evaluation, monitoring, and assessment. This report has investigated existing in-service bridge decks and presented a potential framework and methodology for evaluating sub-surface defects, such as delaminations. The primary objective of this research was to assess the condition of the existing bridges. The assessment was preceded with reliable data acquisition, quality evaluation, validation, and annotation, which is the basis for evaluating the condition of the bridge deck's sub-surface.

NDE techniques, such as unmanned aerial systems/vehicles (UAS/UAV's) mounted with infrared thermography or visual sensors, GP), and IE, are useful for bridge deck condition assessments. A well-validated SDNET2021 dataset has been developed that contains the following summary dataset information

- IE
    - Class 1 - 1,448 sound
    - Class 2&3 - 488 delaminated
- GPR
    - Class 1 - 448,159 sound
    - Class 2&3 - 214,943 delaminated
- IRT image pixels
    - Class 1 - 2,862,597 sound
    - Class 2&3 - 1,718,083 delaminated

The dataset is publicly available. The study also developed a reliable ground truth and annotation technique for the collected NDE dataset, which is useful for validating Artificial

intelligence (AI) models. The dataset annotation was performed autonomously to avoid possible dataset errors or misrepresentations.

This study also confirms the feasibility of IRT applications for concrete bridge deck delamination detection. The method applied in this study is conceptually different than most recently developed image-based algorithms and models that have no reliable ground truth or were created using laboratory specimens. We have investigated the effectiveness of IRT when detecting delamination with reliable and validated ground truth data. The developed model, which was optimized by iterating the sensitivity (s) parameters and selecting s-values, revealed the ability to detect delaminations with an accuracy of approximately 70%. The study revealed that several factors, such as delamination depth and spatial dimension; ambient weather conditions such as wind speed, temperature, and humidity; and mosaic image quality may have affected the model's performance. The study indicates that adopting deep learning and data fusion methods to improve model accuracy and performance is promising.

One known limitation of the model is the adoption of low-resolution and quality mosaic IRT images. Investigating quality images with higher resolution may improve the model's performance, which is an important consideration for future studies.

Using the SDNET2021 dataset, IRT techniques, and other advanced techniques in bridge evaluation may significantly benefit the structural health monitoring (SHM) and longevity of bridge infrastructure.

4.2 Future Work

i. Fine-tune and improve the model to increase the accuracy, possibly using other evaluation metrics. The model can be optimized by fine-tuning the pre-processing parameters to yield an optimum delamination detection model.

ii. Adopt techniques, such as NDE data fusion between IRT with visual/RGB images or other NDE datasets to increase and explore the effectiveness of IRT techniques. The IRT technique will augment RGB sensors to aid in removing noise or unnecessary artifacts that may not be defects or delaminations.

iii. Improve existing deep learning (DL)-based models or develop new models that will effectively detect delaminations and defects, ensuring that they are benchmarked with a validation ground truth dataset, which has been developed.

iv. Evaluate the effect of weather conditions, such as wind speed, ambient temperature, humidity, and other prevailing conditions. Optimum ambient weather conditions exist that will yield optimum delamination detection accuracy results. The NDE dataset should be recorded under varying ambient weather conditions and evaluated to determine the effect of these conditions on the model performance.

v. Evaluate the effect of delamination sizes and depth on the model's performance. The size and depth of the delaminations can influence the prediction of the optimal time window for passive infrared thermography inspections. Both factors must be further studied to precisely predict the best time window for data collection.

vi. Investigate and develop an objective and reliable health index for bridge decks and elements. The health index developed for bridges should be reviewed and evaluated based on reliable and objective methods that have been well validated. This work will form a proper basis for bridge stakeholder management and policy decisions and help prioritize repairs and budget decisions, which are crucial aspects of DOT, government, and private institution bridge management systems.

References

Abouhamad, M., Dawood, T., Jabri, A., Alsharqawi, M., & Zayed, T. (2017). Corrosiveness mapping of bridge decks using image-based analysis of GPR data. Automation in Construction, 80, 104-117.

Abdel-qader, I., Yohali, S., Abudayyeh, O., Yehia, S.: Segmentation of thermal images for non-destructive evaluation of bridge decks. NDT E Int. 41, 395–405 (2008).

Amhaz, R., Chambon, S., Idier, J., & Baltazart, V. (2016). Automatic Crack Detection on Two-Dimensional Pavement Images: An Algorithm Based on Minimal Path Selection. *IEEE transactions on intelligent transportation systems*, *17*(10), 2718-2729. https://doi.org/10.1109/TITS.2015.2477675

ASTM D4788-03, Standard Test Method for Detecting Delaminations in Bridge Decks Using Infrared Thermography. ASTM International; West Conshohocken, PA, USA: 2013.

ASTM, C. (2010). Test method for measuring the P-wave speed and the thickness of concrete plates using the impact-echo method. Annual Book of ASTM Standards, 1383(04.02).

ASTM, Standard Test Method for Evaluating Asphalt-Covered Concrete Bridge Decks Using Ground Penetrating Radar, in ASTM D6087-08, West Conshohocken, PA., 2010.

ASCE. 2019 Report card for North Dakota's Infrastructure. American Society of Civil Engineers.

ASCE (2021) Infrastructure report card. https://www.infrastructurereportcard.org/americas-grades/. Accessed 19 Apr 2018.

Azari, H., & Lin, S. (2019). Evaluation of the Impact Echo Method for Concrete Bridge Decks with Asphalt Overlays. *Transportation research record*, *2673*(2), 436-444. https://doi.org/10.1177/0361198119828676

Ball, J. E., & Bruce, L. M. (2007, August). Digital mammogram spiculated mass detection and spicule segmentation using level sets. In 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4979-4984). IEEE.

Chambon, S., & Moliard, J.-M. (2011). Automatic Road Pavement Assessment with Image Processing: Review and Comparison. *International journal of geophysics*, *2011*, 1-20. https://doi.org/10.1155/2011/989354

Chen, H. Y., & Chen, C. (2016). Determining the emissivity and temperature of building materials by infrared thermometer. Construction and Building Materials, 126, 130-137.

Cheng, C.-C., Cheng, T., Chiang, C.: Defect detection of concrete structures using both infrared thermography and elastic waves. Autom. Constr. 18, 87–92 (2008).

Cheng, C., Shang, Z., & Shen, Z. (2020). Automatic delamination segmentation for bridge deck based on encoder-decoder deep learning through UAV-based thermography. NDT & E International, 116, 102341.

Dabous, S. A., & Feroz, S. (2020). Condition monitoring of bridges with non-contact testing technologies. Automation in Construction, 116, 103224.

Data - LTBP InfoBridge (dot.gov), https://infobridge.fhwa.dot.gov/Data/SelectedBridges

Diamanti, N., Annan, A. P., & Redman, J. D. (2017). Concrete bridge deck deterioration assessment using ground penetrating radar (GPR). *Environmental & engineering geoscience*, *22*(2), 121-132. https://doi.org/10.2113/JEEG22.2.121

Dinh, K., Gucunski, N., Kim, J., & Duong, T. H. (2016). Understanding depth-amplitude effects in assessment of GPR data from concrete bridge decks. *NDT & E international*

: independent nondestructive testing and evaluation, *83*, 48-58. https://doi.org/10.1016/j.ndteint.2016.06.004

Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks. Data in Brief, 21, 1664-1668. https://doi.org/10.1016/j.dib.2018.11.015

Dorafshan, S., & Azari, H. (2020a). Deep learning models for bridge deck evaluation using impact echo. *Construction and Building Materials*, *263*, Article 120109. https://doi.org/10.1016/j.combuildmat.2020.120109

Dorafshan, S., & Azari, H. (2020b). Evaluation of bridge decks with overlays using impact echo, a deep learning approach. *Automation in Construction*, *113*, Article 103133. https://doi.org/10.1016/j.autcon.2020.103133

Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Construction and Building Materials*, *186*, 1031-1045. https://doi.org/10.1016/j.conbuildmat.2018.08.011

Dorafshan, S., Thomas, R. J., and Maguire, M. (2018a). "Fatigue Crack Detection Using Unmanned Aerial Systems in Fracture Critical Inspection of Steel Bridges." Journal of Bridge Engineering, 23(10).

Dorafshan, S., & Maguire, M. (2018b). Bridge inspection: human performance, unmanned aerial systems and automation. Journal of Civil Structural Health Monitoring, 8(3), 443-476.

Dorafshan, S., Campbell, L. E., Maguire, M., & Connor, R. J. (2021). Benchmarking Unmanned Aerial Systems-Assisted Inspection of Steel Bridges for Fatigue Cracks. Transportation Research Record, 03611981211001073.

Ellenberg, A., Kontsos, A., Moon, F., & Bartoli, I. (2016). Bridge deck delamination identification from unmanned aerial vehicle infrared imagery. Automation in Construction, 72, 155-165.

Everingham, M., L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. 2012. “The PASCAL visual object classes challenge 2012 (VOC2012) results.” Accessed March 10, 2020. http://www .pascal-network.org/challenges/VOC/voc2012/workshop/index.html.Ge, F., Wang, S., & Liu, T. (2007). New benchmark for image segmentation evaluation. Journal of Electronic Imaging, 16(3), 033011.

Eisenbach, M., Stricker, R., Seichter, D., Amende, K., Debes, K., Sesselmann, M., & Gross, H. M. (2017, May). How to get pavement distress detection ready for deep learning? A systematic approach. In 2017 international joint conference on neural networks (IJCNN) (pp. 2039-2047). IEEE.

Elkins, G. E. (2003). *Long-term pavement performance information management system: pavement performance database user reference guide*. U.S. Dept. of Transportation, Federal Highway Administration [Offices of] Research, Development, and Technology, Turner-Fairbank Highway Research Center.

Fan, Z., Li, C., Chen, Y., Wei, J., Loprencipe, G., Chen, X., & Di Mascio, P. (2020). Automatic Crack Detection on Road Pavements Using Encoder-Decoder Architecture. *Materials*, *13*(13), 2960. https://doi.org/10.3390/ma13132960

Farrag, S., Yehia, S., & Qaddoumi, N. (2016). Investigation of Mix-Variation Effect on Defect-Detection Ability Using Infrared Thermography as a Nondestructive Evaluation Technique. *Journal of bridge engineering*, *21*(3), 4015055. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000779

Graybeal, B. A., Phares, B. M., Rolander, D. D., Moore, M., & Washer, G. (2002). Visual inspection of highway bridges. Journal of nondestructive evaluation, 21(3), 67-83.

Gu, J., & Unjoh, S. (2021). Image processing methodology for detecting delaminations using infrared thermography in CFRP-jacketed concrete members by infrared thermography. Composite Structures, 270, 114040.

Gucunski, N., & National Research Council. (2013). Nondestructive testing to identify concrete bridge deck deterioration. Transportation Research Board.

Gucunski, N., Boone, S. D., Zobel, R., Ghasemi, H., Parvardeh, H., & Kee, S. H. (2014, April). Nondestructive evaluation inspection of the Arlington Memorial Bridge using a robotic assisted bridge inspection tool (RABIT). In Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security 2014 (Vol. 9063, p. 90630N). International Society for Optics and Photonics.

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., & Agrawal, A. (2017). Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction & building materials*, *157*(C), 322-330. https://doi.org/10.1016/j.conbuildmat.2017.09.110

Gucunski, N., Slabaugh, G., Wang, Z., Fang, T., & Maher, A. (2008). Impact Echo Data from Bridge Deck Testing: Visualization and Interpretation. *Transportation research record*, *2050*(1), 111-121. https://doi.org/10.3141/2050-11

Hiasa, S., Birgul, R., & Catbas, F. N. (2016). Infrared thermography for civil structural assessment: demonstrations with laboratory and field studies. Journal of Civil Structural Health Monitoring, 6(3), 619-636.

Hiasa, S., Birgul, R., & Catbas, F. N. (2017). Effect of Defect Size on Subsurface Defect Detectability and Defect Depth Estimation for Concrete Structures by Infrared

Thermography. *Journal of nondestructive evaluation*, *36*(3), 1-21. https://doi.org/10.1007/s10921-017-0435-3

Hiasa, S., Birgul, R., & Catbas, F. N. (2017). Investigation of effective utilization of infrared thermography (IRT) through advanced finite element modeling. Construction and Building Materials, 150, 295-309.

Hing, C. L. C., & Halabe, U. B. (2010). Nondestructive Testing of GFRP Bridge Decks Using Ground Penetrating Radar and Infrared Thermography. *Journal of Bridge Engineering*, *15*(4), 391-398. https://doi.org/10.1061/(asce)be.1943-5592.0000066

Hou, J., Gao, H., & Li, X. (2016). DSets-DBSCAN: A parameter-free clustering algorithm. IEEE Transactions on Image Processing, 25(7), 3182-3193.

Ichi, E., Dorafshan, S.(2021)a. SDNET2021: Annotated NDE dataset for Structural Defects.

Ichi, E., Dorafshan, S.(2021)b. SDNET2021: An Annotated Dataset of Subsurface Structural Defects in Concrete Bridge Decks. ***(Under review in Journal of Structural Engineering, ASCE).***

Kalogeropoulos, A., van der Kruk, J., Hugenschmidt, J., Bikowski, J., & Brühwiler, E. (2013). Full-waveform GPR inversion to assess chloride gradients in concrete. *NDT & E international : independent nondestructive testing and evaluation*, *57*, 74-84. https://doi.org/10.1016/j.ndteint.2013.03.003

Kuenzer, C., & Dech, S. (Eds.). (2013). Thermal infrared remote sensing: sensors, methods, applications (Vol. 17). Springer Science & Business Media.

Kumar, S. S., D. M. Abraham, M. R. Jahanshahi, T. Iseley, and J. Starr. 2018. “Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks.” Autom. Constr. 91 (Jul): 273–283. https://doi.org/10.1016/j.autcon.2018.03.028.Kumar, S. S., M. Wang, D. M.

Abraham, M. R. Jahanshahi, T. Iseley, and J. C. Cheng. 2020. “Deep learning–based automated detection of sewer defects in CCTV videos.” J. Comput. Civ. Eng. 34 (1): 04019047. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000866.

La, H. M., Gucunski, N., Dana, K., and Kee, S. H. (2017). "Development of an autonomous bridge deck inspection robotic system." Journal of field robotics, 34(8), 1489-1504.

Lee, S., Kalos, N., and Shin, D. H. (2014). "Non-destructive testing methods in the U.S. for bridge inspection and maintenance." KSCE Journal of civil engineering, 18(5), 1322-1331.

Lehmann, B., Wakili, K. G., Frank, T., Collado, B. V., & Tanner, C. (2013). Effects of individual climatic parameters on the infrared thermography of buildings. Applied Energy, 110, 29-43.

Lim, R. S., Hung Manh, L., Zeyong, S., and Weihua, S. (2011). "Developing a crack inspection robot for bridge maintenance." IEEE, 6288-6293.

Liu, H., Lin, C., Cui, J., Fan, L., Xie, X., & Spencer, B. F. (2020). Detection and localization of rebar in concrete by deep learning using ground penetrating radar. *Automation in construction*, *118*, 103279. https://doi.org/10.1016/j.autcon.2020.103279

Liu, W., Huang, Y., Li, Y., & Chen, Q. (2019). FPCNet: Fast Pavement Crack Detection Network Based on Encoder-Decoder Architecture.

Maierhofer, C. (2003). Nondestructive Evaluation of Concrete Infrastructure with Ground Penetrating Radar. *Journal of materials in civil engineering*, *15*(3), 287-297. https://doi.org/10.1061/(ASCE)0899-1561(2003)15:3(287)

Majidifard, H., Jin, P., Adu-Gyamfi, Y., & Buttlar, W. G. (2020). Pavement Image Datasets: A New Benchmark Dataset to Classify and Densify Pavement Distresses.

Transportation research record, *2674*(2), 328-339. https://doi.org/10.1177/0361198120907283

Madding, R. P. (1983, March). Science behind thermography. In Thermosense V (Vol. 371, pp. 2-9). International Society for Optics and Photonics.Pratt, W. K. (1991). Digital image processing John Wiley & sons. Inc., New York.

Maser, K. R. (2009, June). Integration of ground penetrating radar and infrared thermography for bridge deck condition evaluation. In Non-Destructive Testing in Civil Engineering Conference, Nantes, France, June 30th-July 3rd.

Maierhofer, C., Brink, A., Ro, M., Wiggenhauser, H.: Quantitative impulse-thermography as non-destructive testing method in civil engineering—experimental results and numerical simulations. Constr. Build. Mater. 19, 731–737 (2005).

McLaughlin, E., Charron, N., & Narasimhan, S. (2019). Combining Deep Learning and Robotics for Automated Concrete Delamination Assessment. *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, *36*, 485-492. https://doi.org/10.22260/ISARC2019/0065

McLaughlin, E., Charron, N., & Narasimhan, S. (2020). Automated defect quantification in concrete bridges using robotics and deep learning. Journal of Computing in Civil Engineering, 34(5), 04020029.

Mei, Q., & Gül, M. (2019). A Cost Effective Solution for Road Crack Inspection using Cameras and Deep Neural Networks.

Mei, Q., & Gül, M. (2020). A cost effective solution for pavement crack inspection using cameras and deep neural networks. *Construction & building materials*, *256*, 119397. https://doi.org/10.1016/j.conbuildmat.2020.119397

Montaggioli, G., Puliti, M., & Sabato, A. (2021, March). Automated damage detection of bridge's sub-surface defects from infrared images using machine learning. In Health Monitoring of Structural and Biological Systems XV (Vol. 11593, p. 115932A). International Society for Optics and Photonics.

Mundt, M., Majumder, S., Murali, S., Panetsos, P., & Ramesh, V. (2019). Meta-learning Convolutional Neural Architectures for Multi-target Concrete Defect Classification with the COncrete DEfect BRidge IMage Dataset.

Nebme, J. (2019). A New Home for Bridge Data: FHWA's InfoBridge is a centralized gateway for efficient and quick access to performance-related data and information. *Public roads*, *83*(2), 4.

Omar, T., & Nehdi, M. L. (2017). Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography. Automation in Construction, 83, 360-371.

Omar, T., Nehdi, M. L., & Zayed, T. (2018). Infrared thermography model for automated detection of delamination in RC bridge decks. Construction and Building Materials, 168, 313-327.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

Özgenel, Çağlar Fırat (2019), "Concrete Crack Images for Classification", Mendeley Data, V2, doi: 10.17632/5y9wdsg2zt.2

Özgenel, Ç. F., & Sorguç, A. G. (2018). Performance comparison of pretrained convolutional neural networks on crack detection in buildings. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 35, pp. 1-8). IAARC Publications.

Qin, X., He, S., Zhang, Z., Dehghan, M., & Jagersand, M. (2018, March). Bylabel: A boundary based semi-automatic image annotation tool. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 1804-1813). IEEE.

Pozzer, S., Rezazadeh Azar, E., Dalla Rosa, F., & Chamberlain Pravia, Z. M. (2021). Semantic segmentation of defects in infrared thermographic images of highly damaged concrete structures. Journal of Performance of Constructed Facilities, 35(1), 04020131.

Pratt, W. K. (1991). Digital image processing John Wiley & sons. Inc., New York.

Raja, B. N. K., Miramini, S., Duffield, C., Sofi, M., Mendis, P., & Zhang, L. (2020). The influence of ambient environmental conditions in detecting bridge concrete deck delamination using infrared thermography (IRT). Structural Control and Health Monitoring, 27(4), e2506.

Rens, K. L., Wipf, T. J., & Klaiber, F. W. (1997). Review of nondestructive evaluation techniques of civil infrastructure. Journal of performance of constructed facilities, 11(4), 152-160.

Rhee, J.-Y., Choi, J.-J., & Kee, S.-H. (2019). Evaluation of the Depth of Deteriorations in Concrete Bridge Decks with Asphalt Overlays Using Air-Coupled GPR: A Case Study from a Pilot Bridge on Korean Expressway. *International journal of concrete structures and materials*, *13*(1), 1-17. https://doi.org/10.1186/s40069-018-0327-7

Rolander, D. D., Phares, B. M., Graybeal, B. A., Moore, M. E., Washer, G. A., and Trb (2001). "Highway bridge inspection - State-of-the-practice survey." Maintenance of Transportation Pavements and Structures: Maintenance(1749), 73-81.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic imaging, 13(1), 146-165.

Shi, Y., Cui, L., Qi, Z., Meng, F., & Chen, Z. (2016). Automatic Road Crack Detection Using Random Structured Forests. *IEEE transactions on intelligent transportation systems, 17*(12), 3434-3445. https://doi.org/10.1109/TITS.2016.2552248

Taylor, T. R., Chao, C. T., & Chiou, J. S. (2020). Novel Image State Ensemble Decomposition Method for M87 Imaging. Applied Sciences, 10(4), 1535.

Tran, Q. H., Huh, J., Kang, C., Lee, B. Y., Kim, I.-T., & Ahn, J.-H. (2018). Detectability of Subsurface Defects with Different Width-to-Depth Ratios in Concrete Structures Using Pulsed Thermography. *Journal of nondestructive evaluation, 37*(2), 1-11. https://doi.org/10.1007/s10921-018-0489-x

Tran, Q. H., Han, D., Kang, C., Haldar, A., & Huh, J. (2017). Effects of ambient temperature and relative humidity on subsurface defect detection in concrete structures by active thermal imaging. Sensors, 17(8), 1718.

Vaghefi, K., Oats, R. C., Harris, D. K., Ahlborn, T. M., Brooks, C. N., Endsley, K. A., . . . Dobson, R. (2012). Evaluation of Commercially Available Remote Sensors for Highway Bridge Condition Assessment. *Journal of bridge engineering, 17*(6), 886-895. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000303

Wang, T., Han, B., & Collomosse, J. (2014). Touchcut: Fast image and video segmentation using single-touch interaction. Computer Vision and Image Understanding, 120, 14-30.

Washer, G., Fenwick, R., Bolleni, N., & Harper, J. (2009). Effects of Environmental Variables on Infrared Imaging of Subsurface Features of Concrete Bridges. *Transportation research record, 2108*(1), 107-114. https://doi.org/10.3141/2108-12

Washer, G., Fenwick, R., & Bolleni, N. (2010). Effects of solar loading on infrared imaging of subsurface features in concrete. Journal of Bridge Engineering, 15(4), 384-390.

Washer, G., Fenwick, R., Nelson, S., & Rumbayan, R. (2013). Guidelines for Thermographic Inspection of Concrete Bridge Components in Shaded Conditions. *Transportation research record*, *2360*(1), 13-20. https://doi.org/10.3141/2360-02

Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., & Ling, H. (2020). Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. *IEEE transactions on intelligent transportation systems*, *21*(4), 1525-1535. https://doi.org/10.1109/TITS.2019.2910595

Zhang, A., Wang, K. C. P., Fei, Y., Liu, Y., Chen, C., Yang, G., . . . Qiu, S. (2019). Automated Pixel-Level Pavement Crack Detection on 3D Asphalt Surfaces with a Recurrent Neural Network. *Computer-aided civil and infrastructure engineering*, *34*(3), 213-229. https://doi.org/10.1111/mice.12409

Zhang, J.-K., Yan, W., & Cui, D.-M. (2016). Concrete Condition Assessment Using Impact-Echo Method and Extreme Learning Machines. *Sensors (Basel, Switzerland)*, *16*(4), 447-447. https://doi.org/10.3390/s16040447

Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In Graphics gems IV (pp. 474-485).