FEDERAL UNIVERSITY OF CEARA
CENTER OF TECHNOLOGY
DEPARTMENT OF TRANSPORTATION ENGINEERING
POST-GRADUATE PROGRAM IN TRANSPORT ENGINEERING

ALINE CALHEIROS ESPÍNDOLA

COMPUTER VISION AND CONVOLUTIONAL NEURAL NETWORKS APPLIED IN PAVEMENT EVALUATION

FORTALEZA

2024

ALINE CALHEIROS ESPÍNDOLA

COMPUTER VISION AND CONVOLUTIONAL NEURAL NETWORKS APPLIED IN
PAVEMENT EVALUATION

Thesis submitted to the Post-Graduate Program
in Transport Engineering of the Technology
Center at the Federal University of Ceará,
as a partial requirement of the Doctorate in
Transport Engineering. Area of concentration:
Transport infrastructure.

Advisor: Prof. Dr. Ernesto Ferreira Nobre
Júnior
Co-advisor: Prof. Ph.D Mujib Rahman

FORTALEZA

2024

ALINE CALHEIROS ESPÍNDOLA

COMPUTER VISION AND CONVOLUTIONAL NEURAL NETWORKS APPLIED IN PAVEMENT EVALUATION

Thesis submitted to the Post-Graduate Program in Transport Engineering of the Technology Center at the Federal University of Ceará, as a partial requirement of the Doctorate in Transport Engineering. Area of concentration: Transport infrastructure.

Evaluated on: 29/07/2024.

EXAMINATION BOARD

Prof. Dr. Ernesto Ferreira Nobre Júnior (Advisor)
Federal University of Ceará (UFC-PETRAN)

Profa. Dra. Arielle Elias Arantes
Federal University of Ceará (UFC-DET)

Prof. Dr. Elias Teodoro da Silva Júnior
Federal Institute of Ceará (IFCE-PPGCC)

Prof. Dr. Guilherme de Alencar Barreto
Federal University of Ceará (UFC-PPGETI)

Profa. Dra. Viviane Adriano Falcão
Federal University of Pernambuco (UFPE-PPGEC)

I dedicate this thesis to those who believe in science and the transformative power of education. May this work contribute, even modestly, to the advancement of knowledge in our field.

ACKNOWLEDGMENTS

The acknowledgments will be the only item in these contents in Portuguese language to enable the understanding of everyone recipients in the acknowledgment.

Agradeço, primeiramente, a Deus por me conceder a oportunidade e as condições necessárias para realizar este sonho.

À Universidade Federal do Ceará (UFC), sou profundamente grata pela oportunidade de realizar meu doutorado. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), expresso meu sincero agradecimento pelo apoio financeiro para o doutorado sanduíche.

Agradeço também à Universidade Federal de Alagoas (UFAL), especialmente aos professores Vladimir e Roberto, diretores do centro de tecnologia, ao prof. Barbirato, coordenador do curso de engenharia civil, e ao prof. Alexandre, colega de jornada em engenharia de transportes, pelo suporte e encorajamento ao longo desta trajetória.

Minha gratidão eterna é dedicada à minha família. Aos meus pais, Afonso e Alice, por proporcionarem todas as condições para que eu alcançasse esta conquista, por seu carinho, incentivo, motivação e preocupação constantes. Vocês são meus eternos exemplos a seguir, e a vocês, dedico todo meu amor e carinho.

Aos meus irmãos, Alisson e Alicia, meu agradecimento especial por sempre se orgulharem de mim e confiarem no meu trabalho. Aos meus sobrinhos, Artur, Sofia, Lucas e Miguel, por trazerem alegria a tantos momentos.

Agradeço de coração à Raisa, Larissa, Helena, Ana, Ítalo, Matos e Wan, pela amizade, incentivo e compreensão durante as muitas ausências.

Aos meus amigos Ana, Pedrinho, Gledson, Marilia, Renan, Wendy e Aline Vale, agradeço por compartilharem momentos maravilhosos, por tornarem minha estadia em Fortaleza agradável.

Ao meu orientador, Prof. Ernesto Ferreira Nobre Júnior, agradeço profundamente por acreditar na minha capacidade, sempre me motivando e incentivando o avanço das atividades. Sua orientação, ensinamentos, paciência e compreensão durante os momentos de ausência foram essenciais para a conclusão deste trabalho.

To my co-advisor, Prof. Mujib Rahman, and researcher Senthan Mathavan, I thank for the technical support and monitoring in the development of research articles.

ABSTRACT

The implementation of new technologies in pavement management systems (PMS) is crucial, particularly through computer vision and machine learning to develop automatic pavement evaluation systems. This research presents a deep convolutional neural network (CNN) model capable of classifying various pavement types, including concrete, asphalt, interlocking, cobblestone/stone, and unpaved roads, achieving nearly 100% accuracy. Additionally, the study explores image-based distress diagnosis using multi-label classification (MLC) models with ResNet-50 architecture, achieving a 96% accuracy rate and a 91% F1-score, effectively identifying pavement distress from network-level video surveys. This approach reduces the need for expensive sensors, offering a cost-effective alternative for comprehensive road condition assessments. Continuous and extensive examination of pavement surfaces is vital for PMS effectiveness, and automated assessments using CNNs offer a feasible solution by quickly and accurately identifying defects, thereby reducing manual evaluation costs. The research proposes a Simplified Pavement State Index (SPS) using 2D images from low-cost cameras, demonstrating high similarity in prioritizing segments for intervention compared to the traditional Global Severity Index (IGG). Automating data collection through CNNs provides increased efficiency, enabling continuous monitoring and timely maintenance. Despite the initial investment and challenges in ensuring system robustness, the use of action cameras and smartphones presents a low-cost alternative supporting network-level PMS initiative. This study emphasizes the importance of collaboration between academia and highway management agencies for continuous improvement and underscores the potential of automated assessments to revolutionize pavement management, providing rapid, and comprehensive evaluations, ultimately aiding in effective maintenance and extending pavement service life.

Keywords: Pavement Management System; Convolutional Neural Network; Image Classification; Simplified Pavement State Index.

RESUMO

A implementação de novas tecnologias no sistema de gerência de pavimentos (SGP) é essencial, especialmente por meio da visão computacional e do aprendizado de máquina, para o desenvolvimento de sistemas automáticos de avaliação de pavimentos. Esta pesquisa apresenta um modelo profundo de rede neural convolucional (CNN) capaz de classificar diversos tipos de pavimentos – incluindo concreto, asfalto, blocos intertravados, paralelepípedos / pedra e vias não pavimentadas – alcançando uma precisão próxima de 100%. Adicionalmente, o estudo explora o diagnóstico de defeitos com base em imagens, utilizando modelos de classificação multi-rótulo (MLC) com arquitetura ResNet-50, obtendo uma precisão de 96% e um F1-score de 91%, identificando eficazmente deteriorações do pavimento a partir de levantamentos em vídeo em nível de rede. Essa abordagem reduz a necessidade de sensores caros, oferecendo uma alternativa econômica para avaliações abrangentes das condições das vias. O exame contínuo e extensivo das superfícies de pavimento é vital para a eficácia do SGP, e as avaliações automatizadas com CNNs representam uma solução viável ao identificar defeitos de forma rápida e precisa, reduzindo os custos com avaliações manuais. A pesquisa propõe um Índice Simplificado do Estado do Pavimento (SPS), utilizando imagens 2D de câmeras de baixo custo, demonstrando alta similaridade na priorização de segmentos para intervenção quando comparado ao tradicional Índice Global de Gravidade (IGG). A automação da coleta de dados por meio de CNNs proporciona maior eficiência, permitindo o monitoramento contínuo e a manutenção oportuna. Apesar do investimento inicial e dos desafios para garantir a robustez do sistema, o uso de câmeras de ação e smartphones apresenta uma alternativa de baixo custo que apoia as iniciativas do SGP em nível de rede. Este estudo enfatiza a importância da colaboração entre a academia e os órgãos de gestão rodoviária para a melhoria contínua e destaca o potencial das avaliações automatizadas para revolucionar a gestão de pavimentos, oferecendo avaliações rápidas e abrangentes, auxiliando na manutenção eficaz e prolongando a vida útil do pavimento.

Palavras-chave: Sistema de Gerência de Pavimentos; Rede Neural Convolucional; Classificação de Imagens; Índice Simplificado do Estado do Pavimento.

LIST OF FIGURES

Manuscript Paper 01

Manuscript Paper 02

Manuscript Paper 03

Manuscript Paper 04

LIST OF TABLES

Manuscript Paper 01

Manuscript Paper 02

Manuscript Paper 03

Manuscript Paper 04

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
ANN	Artificial Neural Network
ASCE	American Society of Civil Engineers
ASINVOS	Assistierendes und Interaktiv lernfähiges Videoinspektionssystem für Oberflächenstrukturen am Beispiel von Straßenbelägen und Rohrleitungen
ASTM	American Society for Testing and Materials
CDV	Corrected Deduct Value
CLR	Cyclical Learning Rates
CNN	Convolutional Neural Networks
CNT	National Transport Confederation
CV	Computer Vision
DCNN	Deep Convolutional Neural Network
DFP	Differential Flower Pollination
DL	Deep Learning
DNIT	*Departamento Nacional de Infraestrutura de Transportes* (National Department of Transport Infrastructure)
FLOP	Floating Point Operations per Second
FN	False Negatives
FNN	Feedforward Neural Network
FP	False Positives
FWD	Falling Weight Deflectometer
GAP	German Asphalt Pavement Distress
GDP	Gross Domestic Product
GFLOPS	Giga Floating Point Operations per Second
GIS	Geographical Information System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPUs	Graphics Processing Units
HD	High definition
IC	*Índice de Conservação* (Conservation Index)
ICM	*Índice de Condição da Manutenção* (Maintenance Condition Index)
ICM	Maintenance Condition Index

ICPF	*Índice de Condição dos Pavimentos Flexíveis e Semirrígidos* (Condition Nota for Flexible and Semi-Rigid Pavements)
IES	*Índice do Estado da Superfície* (Surface State Index)
IGG	*Índice de Gravidade Global* (Global Severity Index)
IGI	Individual Severity Index
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IP	*Índice de Pavimentação* (Paving Index)
IQL	Information Quality Level
IRI	International Roughness Index
LR	Learning Rate
LVC	*Levantamento Visual Contínuo* (Continuous Visual Survey)
M&R	Maintenance and Rehabilitation
ML	Machine Learning
MLP	Multi-layer Perceptron's
MNIST	Modified National Institute of Standards and Technology
MSE	Mean Squared Error
NB-CNN	Naïve Bayes Convolutional Neural Network
PCA	Principal Component Analysis
PCI	Pavement Condition Index
PMS	Pavement Management System
PRO	*Procedimento* (Procedure)
RCD	Road Crack Detection
R-CNN	Region-based Convolutional Neural Network
RGB	Red, Green, Blue
ROI	Region of Interest
SGD	Stochastic Gradient Descent
SIFT	Scale Invariant Feature Transform
SPS	Simplified Pavement State Index
SSD	Single Shot Detector
SSIM	Structural Similarity Index Measure
SVM	Support Vector Machines
TER	*Terminologia* (Terminology)
TN	True Negatives
TP	True Positive

TPU	Tensor Processing Units
U.S	United States
USACE	United States Army Corps of Engineers
VGG	Visual Geometry Group
YOLO	You Only Look Once

SUMMARY

1 INTRODUCTION

1.1 Contextualizing

According to the Transportation and Logistics Plan (CNT, 2018), it is estimated that it would take $112 billion to solve road infrastructure bottlenecks. The government spent only $9.2 billion on road infrastructure development and $27.9 billion on maintenance in 2021 (CNT, 2021), resulting in a decline in the quality of Brazilian highways over the past four years, with 66.0% rated as inadequate and 55.5% of assessed roads having pavement deficiencies (CNT: SENAT SEST, 2022). By and large, only the National Transport Confederation (CNT) and the highways contracted to the private initiative produce annual reports on pavement quality in accordance with contractual requirements (ANTT, 2018).

However, this problem is not unique to developing countries. According to projections by the American Society of Civil Engineers (ASCE), the U.S. economy will lose about $10 trillion in GDP by 2039 because of deteriorating and underinvested U.S. infrastructure. Over 40% of U.S. roads are in poor or mediocre condition. The U.S. has underfunded its road maintenance for years, resulting in a $786 billion backlog of road and bridge capital requests. Most of the backlog ($435 billion) is for rehabilitation of existing roads. As roads age and deteriorate, it is estimated that more than 62% of spending has gone toward maintaining the road network (ASCE, 2021).

Information that emphasizes the importance of implementing infrastructure management maintains service quality standards and extends service life with limited resources (Hartmann; Dewulf, 2009). The Pavement Management System (PMS) focuses on supporting management decisions based on long-term strategies. The structure of the PMS is composed of pavement inventory, pavement assessment, deterioration prediction models, priority analysis, and planning of actions and resources needed in the short, medium, and long term (Shahin, 2002).

In some states and the majority of cities, the problem is much worse (Causim, 2001). Conducted surveys of managers of small and medium-sized Brazilian city halls and concluded that cities need and are interested in a pavement management system. Human and financial resources are limited to pavement assessment. The lack of basic data on the road network makes the implementation of infrastructure management infeasible and prevents the definition of maintenance and rehabilitation plans.

The determination of the pavement condition is essential for the preparation of the pavement maintenance plan, with the possibility of manual or automatic approach. The manual approach has fallen out of favor at the network level due to its inherent subjectivity, high financial cost, time commitment, and the amount of work that is no longer practical due to the volume and frequency required. Consequently, automated approaches to pavement assessment have been developed in the last decade, and with technological advances, there is an increasing tendency to use automated surveys for pavement inventory with sophisticated equipment such as sensors, high performance cameras, and Global Navigation Satellite System (GNSS) (Li *et al.*, 2010).

To cope with the huge amounts of data, analysis techniques have also been constantly evolving. Researchers have developed a simple regression model for a sophisticated analysis tool such as fuzzy logic, neural network, and so on. When used properly, they can provide fast and accurate results (Gopalakrishnan *et al.*, 2017).

The problem addressed in this thesis is the automation of data collection for the inventory of road management using various methods and techniques of computer vision and deep learning. Emphasis is placed on simplifying the survey to enable widespread use and technological sophistication in developing models for identifying and building the pavement database to enable implementation of the network level PMS.

1.2 Research problem and issues

1.2.1 Research problem

The low applicability of a pavement management system (PMS) due to the inefficiency of the pavement evaluations currently in use results in high highway maintenance and rehabilitation (M&R) costs.

1.2.2 Research issues

Based on the research gaps identified in the literature review, this research will be conducted based on the following:

    a) What is the effect on the Pavement Management System (PMS) when transitioning from manual to automated collection of pavement condition data?

b) What are the essential advancements in the automated data collecting procedure that can enhance the efficacy of PMSs?

c) What is the minimum required input data for an automated PMS to be implemented at the network level?

d) What types of input data is required for an automatic collection of pavement condition to enable a network-level Pavement Management System (PMS)?

1.3 Research objectives

The overall goal of the research is to develop convolutional neural networks (CNNs) approach to automatically extract information from the pavement surface state for a network-level automated pavement management system (PMS).

1.3.1 Specific objectives

The detailed objectives of this research are:

1) Assessing the advantages and disadvantages of automating the collection of data on the state of the pavement surface;

2) Identification of the methodology and requirements essential for enhancing efficiency in collecting data;

3) Establishing the minimal database structure required to develop an automated network-level PMS (Performance Management System);

4) Identification of image processing techniques that enhance accuracy of automated data collection models;

5) Proposal of index for a low-cost network-level pavement diagnostic system for PMS.

1.4 Considerations about method and thesis organization

To achieve the previously defined objectives and present the results with better fluidity, the thesis was sectioned into six chapters following the structure:

– CHAPTER 1: INTRODUCTION – This chapter overviews the research context and defines the research problem. The research objectives and a summary of the programming plan's structure are also presented.

– CHAPTER 2: LITERATURE REVIEW – This chapter will comprehensively review the theories related to the topics addressed in the thesis. The review will be divided into three sub-topics, each focusing on a specific area of interest. The first sub-topic will delve into the pavement management system, providing a detailed overview of its essential components and functions. The second sub-topic will explore the theoretical framework of deep learning and computer vision. Finally, the last sub-topic will showcase the latest research on applying computer vision techniques and deep learning in PMS, providing valuable insights into the current state of the art. By the end of this chapter, readers will have a solid understanding of the theoretical foundations underpinning the thesis and the innovative approaches developed to address the challenges in PMS.

– CHAPTER 3: METHOD This chapter presents an overview of the proposed road diagnostic system framework discussed in the research. Concisely outlining the procedures employed for creating the two convolutional neural network (CNN) classification models and the pavement surface state index. The chapters pertaining to the papers provide a comprehensive explanation of the methodology employed at each stage.

– CHAPTER 4: Manuscript Paper 01 – Classification of Pavement Surface Types Using Deep Convolutional Neural Networks: This chapter delves into the study of automating the initial inventory of highways by identifying the type of pavement surface. This identification will guide the selection of the appropriate pavement evaluation system.

– CHAPTER 5: Manuscript Paper 02 – A Multi-Label Classifier for Pavement Images with Multiple Defects Using Deep Learning: This chapter presents the method proposed that utilizes deep learning techniques to classify pavement images with multiple defects, identifying defects such as potholes, cracks, bleeding, and patches in asphalt pavement. The goal is to infer the pavement condition for the PMS at the Network level while keeping the computational cost low.

– CHAPTER 6: Manuscript Paper 03 – Proposal of Simplified Pavement State Index (SPS) to Pavement Quality Low-Cost Diagnostics: This chapter will focus on unifying the previous chapters, initially identifying the pavement surface type, applying the defect classification model for the asphalt pavements, and then it proposes the Simplified Pavement State Index (SPS) on the road case studies.

Also, comparing the results of SPS with standardized Brazil's Global Severity Index (acronym in Portuguese is IGG – *Índice de Gravidade Global*).

– CHAPTER 7: Manuscript Paper 04 – Economical Automated Pavement Diagnostic System to Network Pavement Management Systems (PMS): This chapter describes the integration of results from the three preceding papers. The model first separates road segments with asphalt pavements. It then utilizes a second model to classify defects in the pavement images. Using this data, the Simplified Pavement State Index (SPS) is computed for the test section. Finally, a pavement quality map is generated based on the SPS.

– CHAPTER 8: CONCLUSIONS – This chapter will present general conclusions obtained after analyzing the neural networks developed regarding the accuracy performance and the applicability of the models and system in real-world situations. In addition, offer suggestions are offered for future research in this field.

This thesis has been conducted under the supervision of Prof. Dr. Ernesto Ferreira Nobre Júnior at the Federal University of Ceara and Prof. Ph.D Mujib Rahman at the Aston University. Except were otherwise indicated, this thesis is my own original work. Most of the results in this thesis have been submitted to or published at international conferences and journals or are currently being prepared for publication. Some of these results have been achieved in collaboration with other researchers, as indicated.

These results include peer-reviewed journal paper – "Comparing different deep learning architectures as vision-based multi-label classifiers for identification of multiple distresses on asphalt pavement" (Espíndola *et al.*, 2023a) – and peer reviewed conference papers – "Low-cost automated pavement diagnostic system to network pavement management systems" (Espíndola *et al.*, 2023b); "Pothole and patch detection on asphalt pavement using deep convolutional neural network" (Espíndola; Freitas; Nobre Júnior, 2021); and "Pavement surface type classification based on deep learning to the automatic pavement evaluation system" (Espíndola; Nobre Júnior; Silva Júnior, 2021).

2 LITERATURE REVIEW

2.1 Pavement Management System (PMS)

According to Haas *et al.* (1994), Pavement Management System (PMS) consists of a broad and coordinated set of activities associated with planning, design, construction, maintenance, evaluation, and research of pavements.

The general idea of a PMS is to increase the efficiency of decisions taken by road administrators, expand their scope, provide feedback on the consequences of these decisions and ensure the consistency of decisions taken at different levels within the organization responsible for highways (Haas; Hudson; Zaniewski, 1994). For Fernandes Junior (2001), a PMS aims to obtain the best return on invested resources, providing users with safe, comfortable, and economical pavements. It should also make it possible to improve pavement conditions and reduce maintenance and rehabilitation costs and vehicle operating costs.

It currently constitutes an essential tool for administrators to outline the most effective way of applying available public resources on highways that need recovery at different levels of intervention to respond to users' needs within a strategic plan that guarantees the reach of a more significant number of kilometers recovered (IPR; DNIT, 2011). PMS are typically conducted in two levels. The first is the network-level, determines when to intervene in each section and in what priority order to optimize budget expenditures and maximize the total performance of the pavement network (Haas; Hudson; Falls, 2015). The second is the project-level on the selected sites identified from the network-level, typically determining the severity of distress for designing maintenance and rehabilitation options (Figure 1).

Figure 1 – Pavement Management System's components
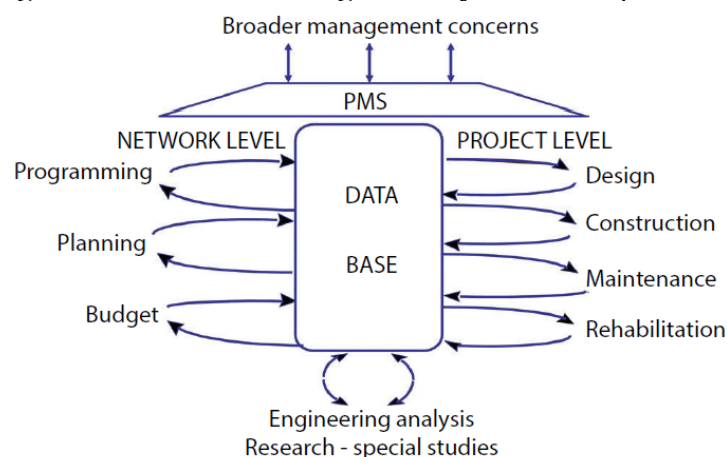
Source: Haas, Hudson and Falls (2015).

The key to the success of the PMS is the quality and detail of the information inventoried on the highway. Project and network-level activities require data that defines the material properties, loads, environment, behavior, distress, and actual performance. The data must be stored in a central database (Figure 2) and be accessible to the entire pavement management process agents. It allows the back analysis of actions and develops more accurate models (Haas; Hudson; Falls, 2015).

Figure 2 – Performance and cost model's diagram
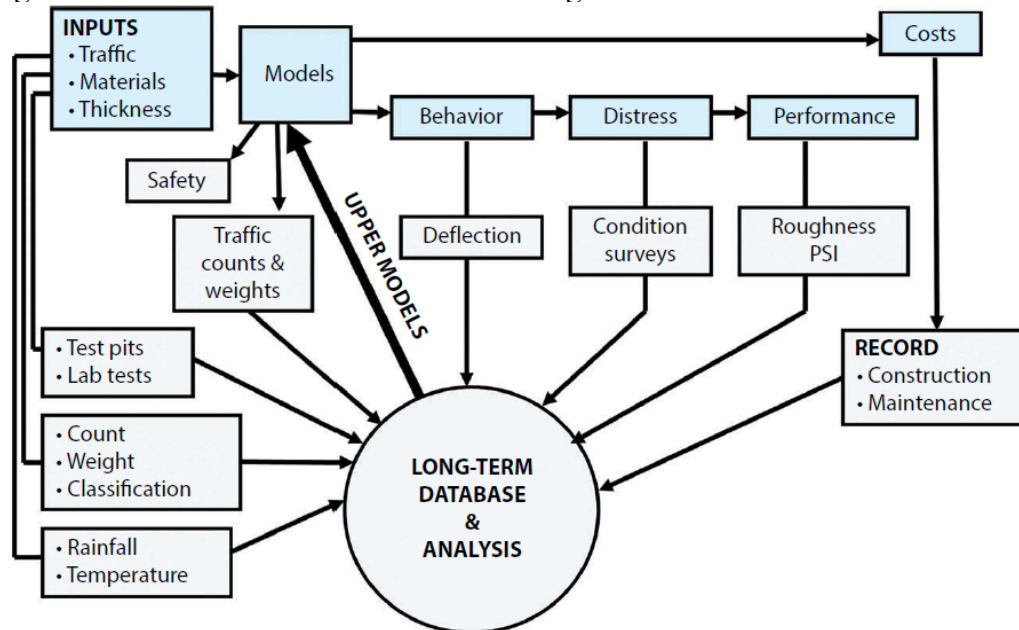
Source: Haas, Hudson and Falls (2015).

2.1.1 Primary activities pavement management system

Table 1 displays the applications and outputs generated by the PMS network and project. The network level is primarily concerned with comprehensive and generic management to managing and coordinating actions and resources, while the project level is focused on addressing specific road-related issues.

Table 1 – Network- and project-level uses

Aspect	Network-level	Project-level
Uses	- Planning - Programming - Budgeting - Pavement management system treatment triggers, identification of candidate projects, life cycle cost analysis - Network-level condition reporting - Mechanistic-Empirical Pavement Design Guide (MEPDG) calibration	- Project scope - Refine pavement management system treatment recommendations - MEPDG calibration

Source: Flintsch and McGhee (2009).

Table 2 depicts the procedures for implementing a structured Pavement Management System capable of delivering technical information for decision-making on priority sites, required activities, and financial resource estimation. The diagnosis and data-collecting phase is the PMS's first stage and foundation, whether at the project or network level.

Table 2 – Summary of activities and decisions within a complete pavement management structure

Basic Blocks of Activities	Network Management Level (Administrative and Technical Decision)	Project Management Level (Technical Management Decisions)
A. Data	1. Sectioning 2. Data a. Field inventory (roughness, surface distress, friction, deflection, geometries) b. Other (traffic, unit costs) 3. Data Processing	1. Detailed data structural, materials, traffic, climate, and unit costs 2. Subsectioning 3. Data Processing
B. Criteria	1. Minimum serviceability, friction, structural adequacy, max. distress 2. Maximum user costs, maintenance costs 3. Maximum program costs 4. Selection criteria (max. of benefits and max. cost-effectiveness)	1. Maximum as-built roughness; minimum structural adequacy and friction 2. Maximum project costs 3. Maximum traffic interruption 4. Selection criteria (such as least total costs)
C. Analyses	1. Network needs (now) 2. Perf. Predictions and future needs 3. Maint. And rehab. Alternatives 4. Technical and economic eval. 5. Priority analysis 6. Eval, of alternative budget levels	1. Within-project alternatives 2. Testing and technical analyses (performance and distress predictions) 3. Life-cycle economic analyses
D. Selection	1. Final priority program of capital projects 2. Final maintenance program	1. Best within-project alternative (rehab. Or new pavement) 2. Maintenance treatments for various sections of networks
E. Implementation	1. Schedule, contracts 2. Program monitoring 3. Budget and financial planning updates	1. Construction activities, contract control, and as-built records 2. Maintenance activities, Maint. Management, and records

Source: Haas, Hudson and Falls (2015).

2.1.2 Data collection

According to Haas, Hudson and Zaniewski (1994), each acquired data has some specific objectives network, and project levels. As can be observed in the surface distresses in Table 3 indicates the most common application of performance data for PMS. In addition, they present managerial actions for the network level and the project level with more immediate and practice applications.

Table 3 – Typical Uses of Pavement Management Performance Data

Data Item	Network Level	Project Level
Roughness	a) Describe present status b) Predict future status (deterioration curves of roughness vs. time or loads) c) Basis for priority analysis and programming	a) Quality assurance (as-built quality of new surface) b) Create deterioration curves c) Estimate overlay quantities
Surface distress	a) Describe present status b) Predict future status (deterioration curves) c) Identify current and future needs d) Maintenance priority programming e) Determine effectiveness of alternative treatments	a) Selection of maintenance treatment b) Identify needed spot improvements c) Develop maintenance quantity estimates d) Determine effectiveness of alternative treatments
Surface friction	a) Describe present status b) Predict future status c) Priority programming d) Determine effectiveness of alternative treatments	a) Identify spot or section rehabilitation requirements b) Determine effectiveness of alternative treatments
Deflection	a) Describe present status b) Predict future status (deterioration curves) c) Identify structural inadequacies d) Priority programming of rehabilitation e) Determine seasonal load restrictions	a) Input to overlay design b) Determine as-built structural adequacy c) Estimate remaining service life d) Estimate remaining load restrictions
Layer material properties	a) Estimate section-to-section variability b) Develop basis for improved design standards	a) Input to overlay design b) Provide as-built records

Source: Haas, Hudson and Zaniewski (1994).

Table 4 lists the components of data collection for each PMS level. As may also be seen, data collection at the network level can be done at operational road speed, as opposed to the project level, which suggests that surveys be done walking or at low speed.

Table 4 – Network- and project-level data collection

Aspect	Network-level	Project-level
Data Items Typically Collected	- IRI - Rut depth - Faulting - Cracking - Punchouts - Patching - Joint condition - Raveling - Bleeding - Surface texture	- Detailed crack mapping and other distresses; - Structural capacity (e.g., falling weight deflectometer [FWD]); - Joint load transfer - Base/soils characterization (e.g., ground penetrating radar, cores, trenches);
Other Items Collected Concurrently	- Video - GPS coordinates - Geometrics (e.g., curve, grade, elevation, cross slope) - Other assets (e.g., bridges, signals) - Events (e.g., construction zones, railroad crossings)	- Drainage conditions; - Appurtenances (e.g., sign and guardrail location and condition); - Geometrics (e.g., curve, grade, elevation, cross slope);
Speed	Typically highway speeds	Walking or slower speeds

Source: Flintsch and McGhee (2009).

Because the study focuses on PMS analysis at the network level, there will be more detail for this level of PMS. Determining the pavement type is critical during data collection

since the data collected can differ, particularly the surface defects. Table 5 details the distresses raised at the network-level phase by pavement type.

Table 5 – Network-level surface deterioration by pavement type

Pavement Type	Network-Level Data Items
Asphalt	- Roughness - Rut depth - Transverse cracking - Fatigue (wheelpath or load-related) cracking - Non-load related (block, edge, or construction joint) cracking - Shoving or distortion - Potholes and/or patching - Bleeding - Raveling - Polishing
Composite (asphalt over concrete)	- All distresses listed for asphalt pavements - Reflective cracking
Jointed Concrete	- Roughness - Faulting - Slab cracking (transverse and/or longitudinal) - Scaling – Polishing - Map cracking (or alkali-silica reactivity) - Durability cracking (D-cracking) - Joint spalling and/or pumping - Joint seal damage - Blowups - Patching
Continuously Reinforced Concrete	- Roughness - Punchouts and/or patching - Longitudinal cracking
Gravel	- Potholes - Washboarding - Loose aggregate or dust

Source: Pierce, McGovern and Zimmerman (2013).

Table 6 shows a classification of the Information Quality Level (IQL) proposed by Paterson and Scullion (1990) that can be employed in defining a specific PMS model. IQL-I and IQL-II are recommended for project-level use, whereas IQL-III and IQL-IV can be utilized for network-level use.

Table 6 – Classification of data collection for PMS by IQL

(continue)

Information Quality Level	Brief Description	Application
IQL-I	Highly detailed	- Research; - Operations; - Advanced Projects; - Diagnostics.
IQL-II	Detailed	- Projects (Preparation); - Programming; - Advanced Planning.

Table 6 – Classification of data collection for PMS by IQL

(conclusion)

Information Quality Level	Brief Description	Application
IQL-III	Summary of details and categorization of values	- Programming; - Planning; - Basic project.
IQL-IV	Very summarized	- Statistic; - Low-traffic Highway Project; - Simple linear planning and programming.

Source: Based on Paterson and Scullion (1990).

The ASTM - E 1777 - 96 (1990) standard specifies the priority recommendations (level of importance) for data requirements for roads and highways. Surface distress and roughness are the two critical items in the performance-related inventory, as shown in Table 7. And depending on the road hierarchy at the network level, the high priority is surface distress. Roughness surveys can now be conducted using low-cost devices such as smartphone applications such as Roadroid (Gamage; Pasindu; Bandara, 2016) and SmartIRI (Magalhães, 2019). So, this study focuses on the other high-priority factor during data collection: surface distress.

Table 7 – Priority Guidelines (level of importance) of Data Needs: Roads and Highways

Data Categories	Network Level				Project Level			
	Major		Minor		Major		Minor	
	High Traffic	Low Traffic	High Traffic	Low Traffic	High Traffic	Low Traffic	High Traffic	Low Traffic
Performance-Related								
Roughness	H	H	M	M	H	H	H	H
Surface distress	H	H	H	H	H	H	H	H
Surface friction[A]	M	M	M	L	H	L	M	L
Deflection	M	L	M	L	H	H	H	M
Layer material properties	L	L	L	L	H	M	M	L
History-Related								
Construction	H	H	M	L	H	H	M	L
Maintenance	H	M	M	L	H	M	M	L
Traffic	H	M	M	L	H	M	M	L
Accident	H	M	H	M	H	H	H	M
Cost-Related								
Construction	H	H	H	H	H	H	H	H
Maintenance	H	H	H	M	H	H	H	M
User	H	M	M	L	H	M	M	L
Policy-Related								
Budget	H	H	H	H	H	H	H	H
Available alternatives	H	H	H	M	H	H	H	M
Service level standard	H	M	M	L	H	M	M	L
Geometry-Related								
Section dimensions	H	H	H	H	H	H	H	H
Curvature	H	M	M	L	H	M	M	L
Cross slope	M	L	M	L	H	M	M	M
Vertical curvature	M	L	L	L	M	L	L	L
Shoulder/curb	H	M	H	M	H	H	H	M
Environment-Related								
Drainage	H	M	H	M	H	M	H	M
Climate	H	M	M	L	H	M	M	L

Source: ASTM E 1777 - 96 (1990).

Advancements in technology have led to the creation of several methods for data collection, ranging from basic manual surveys to more intricate techniques like lasers able to

collect 3D data (Sholevar; Golroo; Esfahani, 2022). Figure 3 illustrates various methods of obtaining data for evaluating the quality of pavement.

Figure 3 – Data collection techniques
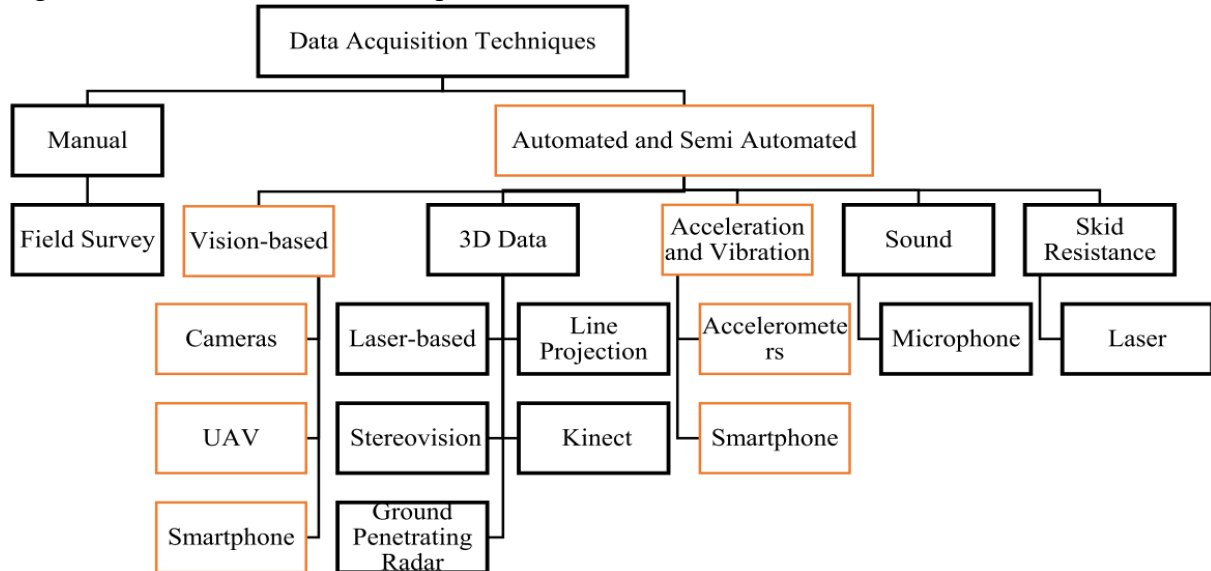
Source: Sholevar, Golroo and Esfahani (2022).

Ragnoli, De Blasiis and Di Benedetto (2018) conducted a literature study on the required data types for extracting information on the different defects present in asphalt pavement. Table 8 displays data type required per the defects, specifying whether the information should be in a two-dimensional (2D) or three-dimensional (3D) format.

Table 8 – Data type per distress

#	Distress	Type of Information
Cracking	Fatigue Cracking	2D | 3D
	Block Cracking	2D | 3D
	Edge Cracking	2D | 3D
	Longitudinal And Transverse Cracking	2D | 3D
	Joint Reflection Cracking	2D | 3D
	Slippage Cracking	2D
Visco Plastic deformations	Bumps And Sags	3D
	Rutting	3D
	Corrugations	3D
	Depressions	3D
	Potholes	2D | 3D
	Swell	2D | 3D
	Lane / Shoulder Drop Off	3D
	Shoving	2D | 3D
Surface defects	Bleeding	2D
	Polished Aggregate	2D
	Raveling	2D
Others	Patching-Utility Cut Patching	2D
	Manhole	2D

Source: Ragnoli, De Blasiis and Di Benedetto (2018).

In order to meet the evolving information requirements of Road Agencies engaged in PMS implementation, the survey vehicles are equipped with a multisensor platform. Typically, these vehicles utilize a GPS-based positioning system integrated with camera, laser, and other sensors to capture georeferenced data and accurately describe the road environment. To categorize the various detection technologies used for acquiring a PMS data set, the following criteria have been considered, as documented in Table 9.

Table 9 – Technological solutions classification

Technology	Information	Precision	Accuracy	Spatial Resolution	Productivity	Cost	Automation
Camera	2D	Medium	Medium	3-6 mm	High	Low	High
Linear scan camera	2D	High	High	2 mm	Medium	Medium	Medium
3D laser imaging	3D	High	High	1 mm	High	High	Medium
TLS	3D	Very High	High	3-6 mm	Medium	High	Low

Source: Ragnoli, De Blasiis and Di Benedetto (2018).

To meet the evolving information requirements of Road Agencies engaged in PMS implementation, the survey vehicles are equipped with a multisensor platform. Typically, these vehicles utilize a GPS-based positioning system integrated with camera, laser, and other sensors to capture georeferenced data and accurately describe the road environment. To categorize the various detection technologies used for acquiring a PMS data set, the following criteria have been considered, as documented in Table 9.

- Precision: indicating the statistical variability.
- Accuracy: indicating systematic error as cause of difference from the true value.
- Spatial resolution: indicating the smallest change it can detect; productivity, indicating the rate of production in time.
- Cost of implementation.
- Automation, implementability in automated process.

Each device has been categorized on a scale from "Low" to "Very High" based on the parameters given above (Ragnoli; De Blasiis; Di Benedetto, 2018). While cameras may possess desirable attributes such as high productivity, economic cost, and automation for network-level inspections, however they do not offer the same level of precision as more expensive equipment.

Figure 4 provides a summary of the review's findings by Ragnoli, De Blasiis and Di Benedetto (2018). Each distress has been assessed as either "influent" (white block) or "not influent" (black block) on comfort and safety. Afterwards, the suitable technology for gathering information has been assessed based on the geometric characteristics and severity level of each

problem. This evaluation is done using a symbolic scale, where "-" indicates that the technology is not suitable, "+" indicates that it is suitable for the survey, and "++" indicates that it is completely suitable.

Figure 4 – Distresses–Technologies relation

Technology	Fatigue cracking			Block cracking			Edge cracking			Longitudinal and transverse cracking			Joint reflection cracking			Slippage cracking			Bumps and sags		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Camera		+	++		+	++			++		+	++			++			++	-	-	+
Line scan		++	++		++	++			++		++	++			++			++	-	-	+
Laser triangulation		++	++		++	++			++		++	++			++			++	-	+	++
Laser scanner		+	++		+	++			++		+	++			++			++	+	++	++

Technology	Rutting			Corrugations			Depressions			Potholes			Swell			Lane / shoulder drop off			Shoving		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Camera	-	-	+	-	-	-		-	+	+	++	++			+		-	+		+	++
Line scan	-	-	+	-	-	-		-	+	+	++	++			+		-	+		+	++
Laser triangulation	-	+	++	-	+	++		+	+	++	++	++			+		+	+		+	++
Laser scanner	+	++	++	+	++	++		++	++	++	++	++			++		+	++		+	++

Technology	Bleeding			Polished aggregate			Raveling			Patching-utility cut patching			Railroad crossing *			Manhole		
	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Camera	+	++	++	+	+	++	+	+	++			++	++	++	++	++	++	++
Line scan	+	++	++	+	++	++	+	++	++			++	++	++	++	++	++	++
Laser triangulation	+	++	++	+	++	++	+	++	++			++	++	++	++	++	++	++
Laser scanner	-	+	+	-	+	+	-	+	+			++	+	++	++	++	++	++

Source: Ragnoli, De Blasiis and Di Benedetto (2018).

The next subtopic will describe some methods for evaluating pavement and their corresponding asphalt pavement quality indexes. The primary evaluation method utilized globally is the Pavement Condition Index (PCI), while the main methods employed exclusively in Brazil by DNIT include the Pavement Condition Index (PCI) and the main methods used by DNIT that are used exclusively in Brazil, Continuous Visual Survey (Continuous Visual Survey [LVC]), paving index (Paving Index [IP]) inside of the Maintenance Condition Index (ICM) and Global Severity Index (IGG).

2.1.3 Pavement evaluation

2.1.3.1 Pavement Condition Index (PCI)

The Pavement Condition Index (PCI) created by the United States Army Corps of Engineers (USACE) is the basis for the pavement distress condition evaluation. The use of PCI for airport pavement, roadways, and parking lots has been legally approved as standard

procedure by several organizations throughout the world. American Society for Testing and Materials (ASTM) issued PCI standards for D5340 to airfield and D6433 to road (Shahin, 2002).

The PCI is a numerical index that ranges from 0 for a failing pavement to 100 for a perfect pavement. PCI is calculated using the findings of a visual condition assessment in which the type, severity, and quantity of distress are determined. The PCI condition survey distress data provides insight into the sources of distress and if they are related to load or climate (Shahin, 2002).

The "deduct values" were included as a weighting factor to reflect how each combination of distress type, severity level, and distress density affects pavement conditions. To calculate the PCI, the sum of the deduct values is rectified depending on the number and value of the deducts, and the corrected sum is subtracted from 100 (Shahin, 2002).

The PCI of a pavement section is calculated by first dividing it into inspection units known as sample units. It decides how many sample units to inspect and which to explore. A critical piece of information is that the number of units to inspect is minimal and depends on the level of PMS or falls under special considerations (airfield, parking lots and unsurfaced roads.). Table 10 shows an example of the number of units to inspect as transportation agencies recommend at the network level. To acquire a meaningful assessment at the lowest cost, the sample units chosen should be reflective (rather than random) of the overall state of the section (Shahin, 2002).

Table 10 – Example of Network Level

No. of Sample Units in Section (N)	No. of Units to be Inspected (n)
1 to 5	1
6 to 10	2
11 to 15	3
16 to 40	4
over 40	10%
	(round up to next whole sample unit)

Source: Shahin (2002).

A sample unit is inspected by measuring the distress type and severity and recording the flexible pavement survey sheet data. An example asphalt-surfaced airport sample unit condition survey sheet is shown in Figure 5. Each sample unit has its own data sheet. Each pavement defect type has its own measurement process; for example, the number 48 (longitudinal and transverse cracking, low severity) in Figure 5 is measured in linear feet, thus 10 represents 10 ft of low-severity cracking.

Figure 5 – Example Airfield AC Sample Unit Condition Survey Sheet

AIRFIELD ASPHALT PAVEMENT CONDITION SURVEY DATA SHEET FOR SAMPLE UNIT		SKETCH:

BRANCH __R1230__ SECTION __B01__ SAMPLE UNIT __008__
SURVEYED BY __MYS__ DATE __MAR/15/92__ SAMPLE AREA __5000 sf__

41. Alligator Cracking	45. Depression	49. Oil Spillage	53. Rutting
42. Bleeding	46. Jet Blast	50. Patching	54. Shoving from PCC
43. Block Cracking	47. Jt. Reflection (PCC)	51. Polished Aggregate	55. Slippage Cracking
44. Corrugation	48. Long. & Trans. Cracking	52. Raveling/Weathering	56. Swell

DISTRESS SEVERITY	QUANTITY										TOTAL	DENSITY %	DEDUCT VALUE
48 L	10	20	17								47	0.94	4.8
48 M	7	9									16	0.32	6.7
41 L	53										53	1.06	21.0
45 L	10	5									15	0.3	1.6
53 L	20	45	10								75	1.5	17.1
53 M	25										25	0.5	20.1

Source: Shahin (2002).

Determine the average PCI for a pavement segment by calculating the PCI for each sample unit using deduct values. The PCI calculation for a sample unit is shown in Figure 6. Step 1: Determine the deduct values; Step 2: Determine the maximum number of deductions allowable. Step 3: Calculate the maximum corrected deduct value (CDV), and Step 4: Subtract the maximum CDV from 100.

Figure 6 – PCI Calculation Steps for a Sample Unit

Step 1. Determine Deduct Values

Step 2. Determine Maximum Allowable Number of Deducts (m)

$$m_I = 1.00 + \frac{9}{98}(100 - HDV_i)$$

Step 3. Determine Maximum Corrected Deduct Value

q = Number of Deducts Greater Than 2 Points

Corrected deduct value curves for asphalt-surfaced pavements

Step 4. Calculate PCI
PCI = 100 – Maximum CDV
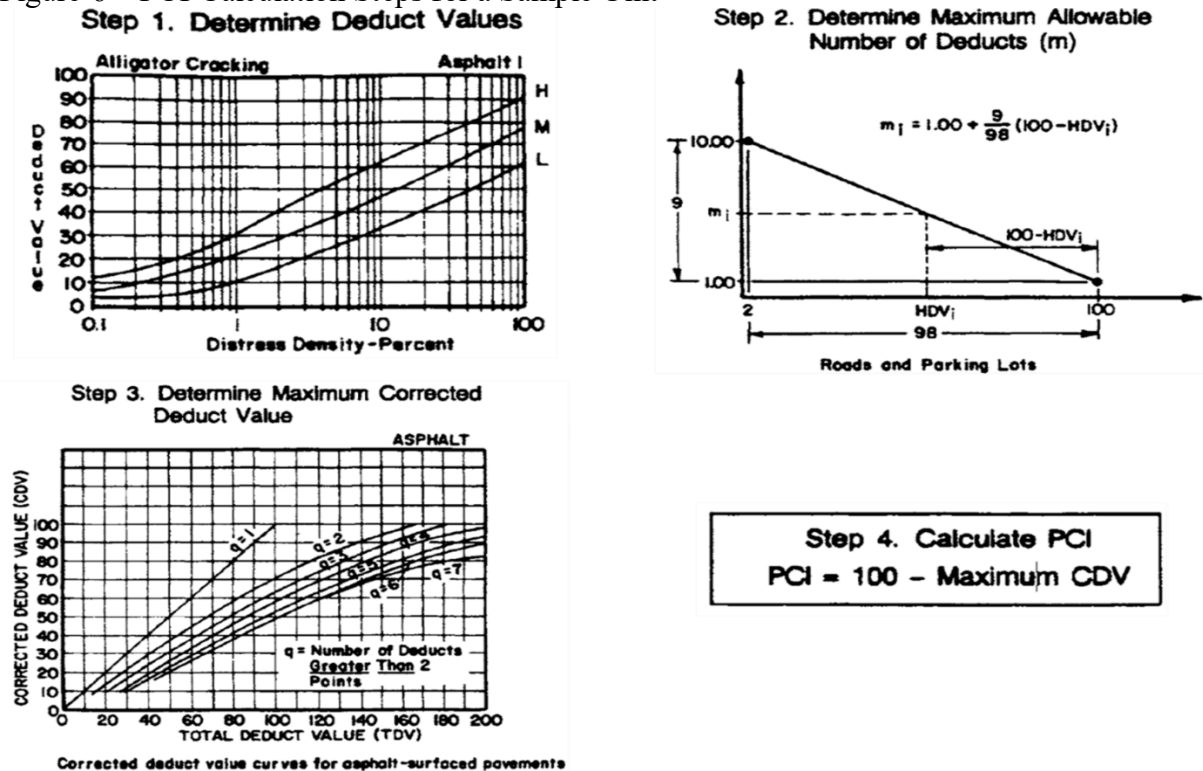
Source: Shahin (2002).

Even though the sample number in the PCI calculation is not particularly large, data collection needs to be manual, with a reasonably high degree of detail and a more difficult calculation technique. This system for measuring pavement conditions is not widely utilized in Brazil. For indicating the standards and procedures for evaluating the quality of the pavement, the responsible agency, the National Department of Transport Infrastructure (acronym in Portuguese is DNIT – *Departamento Nacional de Infraestrutura de Transportes*), uses methodologies with lower levels of detail in data collection and lower calculation complexity.

In general, DNIT manages its highways using information from the International Roughness Index (IRI), Maintenance Condition Index (acronym in Portuguese is ICM – *Índice de Condição da Manutenção*) of highways, which is composed of the Paving Index (acronym in Portuguese is IP – *Índice de Pavimentação*) and Conservation Index (acronym in Portuguese is IC – *Índice de Conservação*), Surface State Index (acronym in Portuguese is IES – *Índice do Estado da Superfície*) through the Continuous Visual Survey procedure (acronym in Portuguese is LVC – *Levantamento Visual Contínuo*) and lastly the Global Severity Index (acronym in Portuguese is IGG – *Índice de Gravidade Global*). The following are the ways to determine the quality of Brazilian pavements, except IRI.

2.1.3.2 Continuous Visual Survey (LVC)

The evaluation methodology based on the Continuous Visual Survey (acronym in Portuguese is LVC – *Levantamento Visual Contínuo*) (IPR; DNIT, 2003c) calculates three deterioration indices to determine the concept of pavement quality. The first index to be calculated is the Condition Nota for Flexible and Semi-Rigid Pavements (acronym in Portuguese is ICPF – *Índice de Condição dos Pavimentos Flexíveis e Semirrígidos*), obtained through the average of the indexes present in the Survey Form and conclusions of the defects observed by the evaluator.

Table 11 exemplifies obtaining the ICPF through the concepts that must be observed when viewing the pavement in the field and in the recordings.

Table 11 – ICPF values

State	Description	ICPF
Excellent	ONLY NEEDS ROUTINE MAINTENANCE	5-4
Good	SEAL TREATMENT – Superficial Raveling, not very severe cracks in not very extensive areas	4-3
Regular	SHAPE CORRECTION (PATCH) OR OVERLAY – Cracked pavement, with "potholes" and infrequent patches and with longitudinal or transverse irregularities.	3-2
Bad	SHAPE CORRECTION AND OVERLAY – Generalized defects with previous corrections in localized areas – superficial or deep patches.	2-1
Terrible	RECONSTRUCTION – Generalized defects with previous corrections to the full extent. Degradation of the asphalt layer and other layers – water infiltration and decompression of the base	1-0

Source: DNIT 008/2003 (IPR; DNIT, 2003c).

The second is Expedited Global Severity Index (acronym in Portuguese is IGGE – *Índice de Gravidade Global Expedito*) is calculated using the following Equation 1:

$$IGGE = (P_t \times F_t) + (P_{oap} \times F_{oap}) + (P_{pr} \times F_{pr}) \qquad (1)$$

Where:

$(P_t \times F_t) \rightarrow$ Multiplying the weight and frequency of the set of cracks, respectively;

$(P_{oap} \times F_{oap}) \rightarrow$ Multiplying the weight and frequency of the deformation set;

$(P_{pr} \times F_{pr}) \rightarrow$ Multiplying the weight and frequency of the set of potholes and patches.

The values of the weights and frequencies of the defects are obtained through Table 12 and Table 13.

Table 12 – Weight values

Severity	Pt	Poap	Ppr
3	0.65	1.00	1.00
2	0.45	0.70	0.80
1	0.30	0.60	0.70

Source: DNIT 008/2003 (IPR; DNIT, 2003c).

Table 13 – Frequency values

Frequency	Pothole (P) and Patch (R)* Fpr Factor		Other defects (cracks and deformation) Ft and Foap Factors	
	Quantity/Km	Severity	%	Severity
A - High**	≥ 5	3	≥ 50	3
M - Medium	2-5	2	50-10	2
B - Low***	≤ 2	1	≤ 10	1

Source: DNIT 008/2003 (IPR; DNIT, 2003c).
*Patch has the initials (R), because in Portuguese, it is called *Remendo* (R).
**High has the initials (A), because in Portuguese, it is called *Alto* (A)
***Low has the initials (B), because in Portuguese, it is called *Baixo* (B)

The Surface State Index of the pavement (IES) is used as a qualitative reference, as well as the others, and the state of the condition of the asphalt layer observed in each segment, ranging from "EXCELLENT" to "TERRIBLE". The IES is obtained through relations dependent on the ICPF and IGGE values, as shown in Table 14.

Table 14 – IES Values

Description	IES	Code	State
IGGE $\leq$ 20 and ICPF > 3.5	0	A	Excellent
IGGE $\leq$ 20 and ICPF $\leq$ 3.5	1	B	Good
20 $\leq$ IGGE <= 40 and ICPF > 3.5	2		
20 $\leq$ IGGE <= 40 and ICPF $\leq$ 3.5	3	C	Regular
40 $\leq$ IGGE <= 60 and ICPF > 2.5	4		
40 $\leq$ IGGE <= 60 and ICPF $\leq$ 2.5	5	D	Bad
60 $\leq$ IGGE <= 90 and ICPF > 2.5	7		
60 $\leq$ IGGE <= 90 and ICPF $\leq$ 2.5	8	E	Terrible
IGGE>90	10		

Source: DNIT 008/2003 (IPR; DNIT, 2003c).

2.1.3.3 Paving Index (IP) of the Maintenance Condition Index (ICM)

An index used in Brazil to diagnose the surface condition is the paving index (acronym in Portuguese is IP – *Índice de Pavimentação*) inside of the Maintenance Condition Index (acronym in Portuguese is ICM – *Índice de Condição da Manutenção*) (DNIT, 2017). The ICM consists of two sub-indexes, the pavement index (IP) which considers the frequency of occurrence of the pothole, patch, and alligator crack (Table 15) and conservation index (IC) which considers the quality and conservation of mowing, drainage and signalling (Table 16).

Table 15 – Occurrence Frequency of distressed

Distress	Low	Medium	High	Unity
Pothole	< 2	3-5	> 5	Quantity/km
Patch	< 2	3-5	> 5	Quantity/km
Alligator Crack	< 10%	10%-50%	> 50%	% of area

Source: DNIT (2017).

Table 16 – Conservation Ranges

Conservation Elements	Good	Regular	Bad
Mowing	Vegetation with a maximum height of 30 cm	Vegetation above 30 cm, but does not affect the visibility of vertical signs	High vegetation that affects the visibility of road signs
Drainage	Integral and painted surface drainage device	Surface drainage device with localized breaks and unpainted	Surface drainage device is broken or missing
Signaling	Vertical and horizontal elements visible and in good condition	Vertical and horizontal elements partially missing or frayed	Vertical and horizontal elements missing or frayed

Source: DNIT (2017).

However, as the research is focused on paving, only the calculation of the paving index and the quality ranges of this index will be shown. Equation 2 shows the IP calculation, and the inputs came from Table 17.

$$IP = P_{Pothole} \cdot 50 + P_{Patch} \cdot 30 + P_{Crack} \cdot 20 \tag{2}$$

Where:

$P_{Pothole}$ – Pothole parameter

P_{Patch} – Patch parameter

P_{Crack} – Crack parameter

Table 17 – Values of the Distressed Parameters

Distress	Low	Medium	High
$P_{Pothole}$	0.25	0.50	1.00
P_{Patch}	0.25	0.50	1.00
P_{Crack}	0.25	0.50	1.00

Source: DNIT (2017).

The paving index (IP) ranges that represent each quality condition of the pavement surface are shown in Table 18.

Table 18 – Pavement Superficial Condition according to Paving Index (IP)

Pavement Condition	IP Range
Good	IP < 34
Regular	34 <= IP < 63
Bad	63 <= IP < 91
Terrible	IP <= 91

Source: elaborated by the author.

2.1.3.4 Global Severity Index (IGG)

Global Severity Index (acronym in Portuguese is IGG – *Índice de Gravidade Global*) is the evaluation of the pavement based on the DNIT norms, DNIT 005/2003-TER (IPR; DNIT, 2003a) and DNIT 006/2003-PRO (IPR; DNIT, 2003b) which deal with the classification of defects and the systematic survey of defects, respectively.

The IPR and DNIT (2003a) defines the technical terms used in defects that occur in pavements, standardizing the language and codification adopted to identify defects, as shown in Table 19 and Table 20.

Table 19 – Summary table of cracks defects – coding and classification

Cracks				Code	Crack Class		
Fissure				FI	-	-	-
Cracks in the asphalt layer caused by excessive permanent deformation and/or due to fatigue	Isolated Cracks	Transversal	Short	TTC	FC-1	FC-2	FC-3
			Long	TTL	FC-1	FC-2	FC-3
		Longitudinal	Short	TLC	FC-1	FC-2	FC-3
			Long	TLL	FC-1	FC-2	FC-3
	Connected Cracks	Alligator	No erosion on crack edges	J	-	FC-2	-
			With erosion on crack edges	JE	-	-	FC-3
Cracks in the asphalt layer not caused by fatigue	Isolated Cracks	Due to thermal shrinkage or dissection of the base (soil-cement) or asphalt layer		TRR	FC-1	FC-2	FC-3
	Connected Cracks	Block crack	No erosion on crack edges	TB	-	FC-2	-
			With erosion on crack edges	TBE	-	-	FC-3

Source: DNIT 005/2003 (IPR; DNIT, 2003a).

Table 20 – Summary table of other defects – coding and classification

Other defects				Code
Deformation	Plastic	Local	Due to the plastic creep of one or more layers of the pavement or subgrade	ALP
		Wheel Rut		ATP
	Consolidation	Local	Due to differential consolidation occurring in pavement or subgrade layers	ALC
		Wheel Rut		ATC
Shoving/Corrugation – Transversal shoving caused by instability of the bituminous mixture constituting the asphalt layer or the base				O
Slippage (asphalt layer)				E
Bleeding of the bituminous binder in the asphalt layer				EX
Severe raveling on the surface of the asphalt layer				D
Potholes arising from breakdown of asphalt layer and sometimes from lower layers				P
Patch		Superficial Patch		RS
		Deep Patch		RP

Source: DNIT 005/2003 (IPR; DNIT, 2003a).

The DNIT 006/2003-PRO norm allows calculating a pavement quality indicator based on an inventory of pavement distress. The IGG is determined on a sampling basis for samples with a predetermined area, spacing between them of 20 m, and when switching between tracks rather than for the total road area. The evaluation is specified in each sample by a 6-meter-long region, 3 meters before and 3 meters after the road stakes, as illustrated in Figure 7 (IPR; DNIT, 2003b).

Figure 7 – Demarcation of areas to register defects
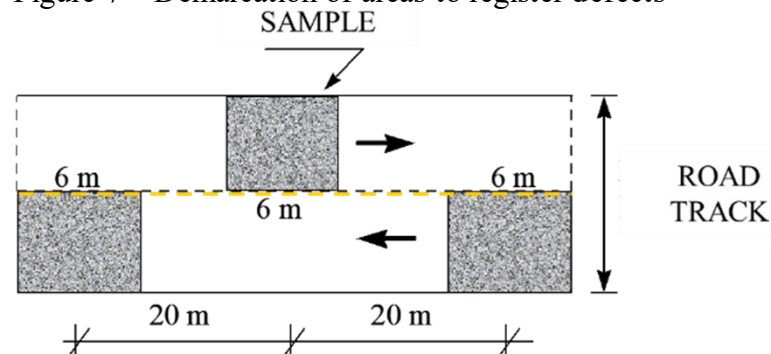
Source: Bernucci *et al.* (2022).

Defects are inventoried within the samples by filling out the form in Figure 8. Defects are grouped into 8 categories according to the defect criticality weighting factors.

Figure 8 – Asphalt Layer Defect Inventory Form

Sample	OK	Isolated cracks						Connected cracks				Other Defects												Depth of Wheel Rut	
		FC-1						FC-2		FC-3		Deformation				Shoving / Corrugation	Pothole	Slippage	Bleeding	Raveling	Pacth	Internal	External		
		FI	TTC	TTL	TLC	TLL	TRR	J	TB	JE	TBE	ALP	ATP	ALC	ATC	O	P	E	EX	D	R	TRI (mm)	TRE (mm)
1	No	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	1	1	-	-
2	No	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-
3	No	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-
4	Yes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	Yes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	No	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-
7	No	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-
8	No	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
9	No	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	No	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-
Total		5						1		1		2				3		0	6		2	0	0

Source: elaborated by the author.

After registering the defects, the weighting factors (fp) are assigned according to Table 21 for each type of distress. The weighted value is a function of the severity of the defects.

Table 21 – Value of the weighting factor to calculate the IGG

Category	Defects	Weighting Factor
1	Isolated cracks	0.2
2	Connected cracks without erosion	0.5
3	Connected cracks with erosion	0.8
4	Deformation	0.9
5	Corrugation, Shoving, Pothole and Slippage	1.0
6	Bleeding	0.5
7	Raveling	0.3
8	Patch	0.6

Source: DNIT 005/2003 (IPR; DNIT, 2003a).

For this methodology, when occurrences of categories 1, 2, and 3 are found in the same sample, only those in category 3 are considered to calculate the relative frequency in percentage and Individual Severity Index (IGI); similarly, when occurrences of categories 1 and 2 are verified in the same sample, only those in category 2 are considered. Always select the most severity.

The absolute and relative frequencies of the inventoried occurrences in the worksheet must be determined to process the collected data (calculating the individual and global severity index). The calculations are performed in the Equations 3, 4, and 5:

$$fr = \frac{fa*100}{n} \tag{3}$$

Where:

fr – relative frequency;

fa – absolute frequency;

n – number of inventoried samples.

$$IGI = fr * fp \tag{4}$$

Where:

fr – relative frequency;

fp – weighting factor;

IGI – individual severity index.

$$IGG = \Sigma\ IGI \tag{5}$$

Where:

IGI – individual severity index;

IGG – global severity index.

Table 22 presents the asphalt layer quality expressed in the state as excellent, good, regular, bad, and terrible according to the IGG values.

Table 22 – State of degradation of the asphalt layer according to the IGG

State	IGG
Excellent	0 – 20
Good	21 – 40
Regular	41 – 80
Bad	81 – 160
Terrible	>160

Source: DNIT 005/2003 (IPR; DNIT, 2003a).

The DNIT 006/2003-PRO norm presents a Worksheet (Table 23) to be filled in with information on the number of samples for each defect category, the IGIs and IGG calculations and the final state of the asphalt layer.

Table 23 – Global Severity Index (IGG) Calculation Worksheet

Category	Defects	Absolute Frequency	Absolute Frequency Considered	Relative Frequency	Weighting Factor	Individual Severity Index (IGI)
1	FI, TTC, TTL, TLC, TLL, TRR	5	4	40	0,2	8
2	(FC – 2) J, TB	1	1	10	0,5	5
3	(FC – 3) JE, TBE	1	1	10	0,8	8
4	ALP, ATP, ALC, ATC	2	╳	20	0,9	18
5	O, P, E	3	╳	30	1	30
6	EX	0	╳	0	0,5	0
7	D	6	╳	60	0,3	18
8	R	2	╳	20	0,6	12
9	The arithmetic mean of the mean values of the depths measured in mm in the TRI and	TRE= 0	TRI= 0	F= 0	1A ( )	0
					1B ( )	0
10	The arithmetic mean of the variances of the depths measured in both wheel ruts	TREv= 0	TRIv= 0	FV= 0	2A ( )	0
					2B ( )	0
	TOTAL NUMBER OF SAMPLES	10	Sum of IGI = IGG =			99
	1A) IGI = F x 4/3 when F ≤ 30	2A) IGI = FV when FV ≤ 50		State		Bad
	1B) IGI = 40 when F > 30	2B) IGI = 50 when FV > 50				

Source: elaborated by the author.

Modernizing the approach of surveying and analyzing pavement quality with current technologies while ensuring feasibility for agencies or cities with limited technical and financial resources remains a challenge. The main challenge is to make the pavement management system (PMS) universally applicable to cities of all sizes. Implementing the PMS at the network level using affordable equipment for data collection and analysis, along with techniques like computer vision and deep learning for defect identification and evaluation, would be a significant advancement in the PMS domain. Sub-topic 2.2 explains the operations and objectives of computer vision and deep learning approaches essential for comprehending the development of the models discussed in the research results.

2.2 Computer Vision and Deep Learning

Computer vision is a field that focuses on developing methods for acquiring, processing, analyzing, and extracting useful information from images or sequences of images. It plays a crucial role in various applications, including object detection, localization, and understanding (Anthony; Kusnadi, 2021).

Humans use their experience of the world along with the information provided by the visual system to analyze and sometimes infer background story of the scene they are viewing. Computers interpret images as numerical data, which necessitates distinct analyses and processing of the image's information while requiring the same outcomes as the human visual system. The endeavor presents a considerable challenge for the computers in

comprehending the numerical matrix due to their lack of knowledge regarding the surrounding context. The primary objective of computer vision is to endow computers with visual capabilities comparable to those possessed by the human visual system (Ikram, 2019).

Artificial Intelligence (AI) describes the use of computational techniques to perform tasks characteristic of human intelligence. Machine Learning (ML) is a subfield of AI that enables computers to automatically learn by being exposed to large amounts of data. Deep Learning (DL) is a specific form of ML that uses multilayered artificial neural networks to make predictions directly from data input. Convolutional Neural Networks (CNNs or ConvNets) are a specialized form of multi-layer neural networks designed to directly identify visual patterns from image pixels with minimal preprocessing (Lin *et al.*, 2021). Figure 9 illustrates the relationship between AI, ML, and DL and CNN.

Figure 9 – Interplay among artificial intelligence (AI), machine learning (ML), deep learning (DL), Convolutional Neural Networks (CNN) and computer vision
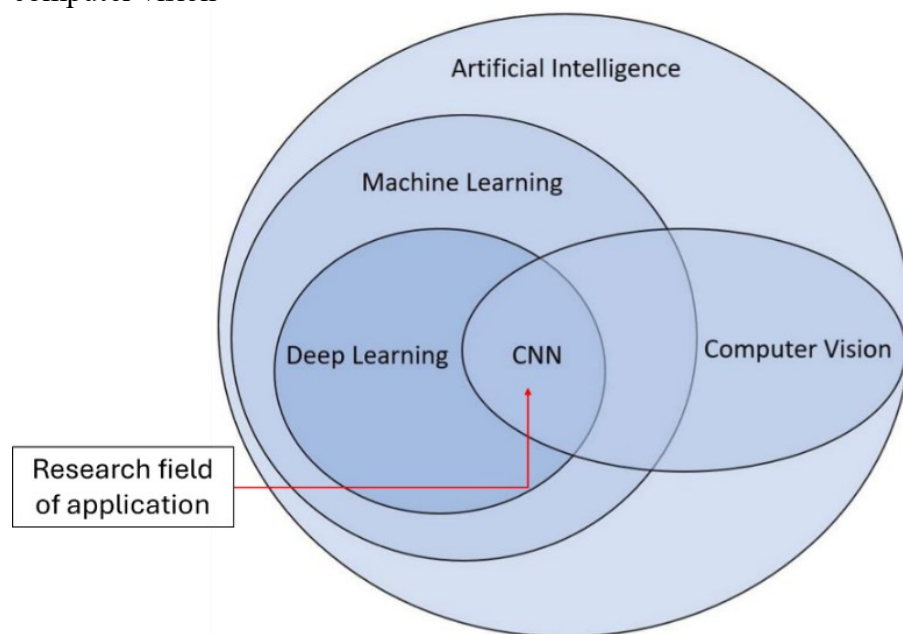
Source: adapted from Beltzung *et al.* (2023).

The classic computer vision (CV) techniques can be categorized into three domains: value, space, and frequency. In the value domain, classic computer vision techniques focus on processing and analyzing pixel values directly. These techniques involve operations such as thresholding, intensity transformations, histogram equalization, and image segmentation. In the space domain, classic computer vision techniques involve analyzing images based on their spatial characteristics. This includes techniques like edge detection, corner detection, image filtering, and template matching. In the frequency domain, classic computer vision techniques

involve analyzing images based on their frequency content. This includes techniques like Fourier analysis, image compression using discrete cosine transform, and filtering in the frequency domain using techniques like convolution (Brosnan; Sun, 2004).

This area pertains to a sector of artificial intelligence (AI) that aims to replicate human intelligence (Gonzalez; Woods, 2010). Machine capability to interpret and comprehend images and videos. Computer vision algorithms are commonly used for tasks such as facial identification, object detection, medical image analysis, and other applications (Szelisk, 2010). There are numerous applications for this technology, including education, athletics, medicine, agriculture, engineering, and astronomy. Investigations have been conducted in the subsequent domains:

- analyze the information included in images related to car and pedestrian identification (Papageorgiou; Poggio, 2000);
- Automated categorization of defects in railroad ties using visual data (Kulkarni; Parikh; Abbott, 2017);
- Remote sensing data is analyzed for geographical information systems (Jung, 2004);
- extraction of graphical and textual information from document images (Esposito; Malerba; Lisi, 2000);
- face recognition (White *et al.*, 2015);
- identifying digits and handwritten characters (Dash; Nayak, 2013);
- research on behavior (Garg *et al.*, 2013);
- autonomous vehicle guidance systems (Huval *et al.*, 2015);
- automatic detection and analysis of curb ramps in Google Street View images (Hara; Kobayashi; Abe, 2016);
- medical science (microscopy, endoscopy, angiography, magnetic resonance, and ultrasound; cardiovascular imaging, retinal blood vessels, and nuclear medicine) (Antonio; Alava, 2015).

A machine learning system has inputs, outputs, and a model that establishes a mapping between the inputs and outputs. The main goal is to acquire a model that attains the utmost accuracy and the lowest error rate when evaluated using the information provided. Machine learning can be classified into three distinct types based on the inputs, outputs, and characteristics of the problem being tackled. The three main types of machine learning are supervised learning, unsupervised learning, and reinforcement learning (Ikram, 2019). Some regularly employed methods in machine learning include convolutional neural networks

(CNN), k-means clustering, and support vector machines. Recent applications encompass tasks such as object detection, object classification, and the extraction of pertinent information from images, graphic documents, and videos (Khan; Al-Habsi, 2020).

Since the initial proposal of CNN by LeCun, Bottou, Bengio, *et al.* (1998), there has been advancement in overcoming the three restrictions that were present at that time. CNN's popularity is growing because of the general accessibility of datasets, Graphics Processing Units (GPUs), and DL techniques (algorithms) (Kulkarni; Parikh; Abbott, 2017).

Figure 10 depicts the fundamental structure that serves as the foundation for multiple prevalent artificial neural network (ANN) architectures. A simple three-layered feedforward neural network (FNN) comprised of an input layer, a hidden layer and an output layer. Typically, the input layer of a neural network receives a multidimensional vector, which is then distributed to the hidden layers. The hidden layers in a neural network analyze the information from the previous layer and evaluate how a random change inside itself affects the final output. This process is known as learning. The term "deep learning" refers to the practice of stacking numerous hidden layers on top of each other (O'Shea; Nash, 2015).

Figure 10 – A simple three-layered feedforward neural network (FNN)
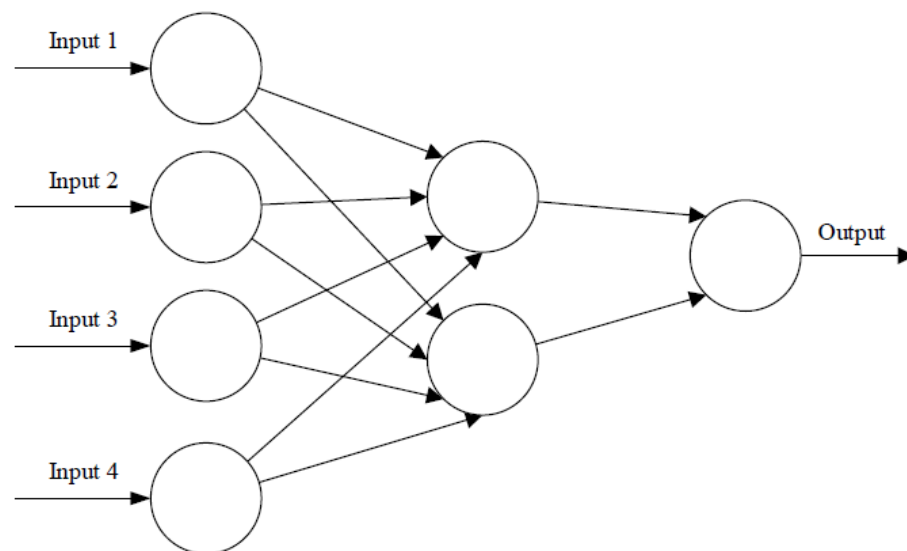
Source: O'Shea and Nash (2015).

Supervised learning involves a dataset that includes both input values and their associated output values. The objective is to determine a mathematical model that accurately maps the inputs to the outputs, conceptualized as a function. Each training example consists of input and output data, which are represented as matrices or vectors. The objective is to enhance

the capacity to generate accurate predictions for additional inputs by integrating dependencies and connections between inputs and outputs in the model (Russell; Norvig, 2022).

Unsupervised learning focuses on data that just includes inputs without any corresponding outputs for each individual sample. The objective is to detect patterns, structures, and commonalities in the data. This is typically accomplished via clustering, in which each training sample is assigned to one of several pre-established clusters based on similarities (Barlow, 1989).

Reinforcement learning does not have access to labelled training data, but it relies on a reward signal to determine the correctness of actions or predictions. Agents engage with environments, progressively improving their performance through a process of experimentation and getting feedback based on their activities. The objective is to optimize the total reward by exploring various actions and evaluating the benefits obtained. During this iterative process, agents acquire knowledge from the environment, which improves their understanding and ability to make decisions (Kaelbling; Littman; Moore, 1996).

Deep Learning (DL) allows computer vision (CV) engineers to achieve higher accuracy in tasks such as imagine classification, segmentation, and object detection (Figure 11) compared to standard CV techniques. Deep learning (DL) offers greater flexibility compared to CV methods, as CNN models and frameworks may be retrained using a custom dataset for each unique application, but CV algorithms are typically more limited to specific domains (O'Mahony *et al.*, 2019).

Figure 11 – Classification, object detection, and instance segmentation
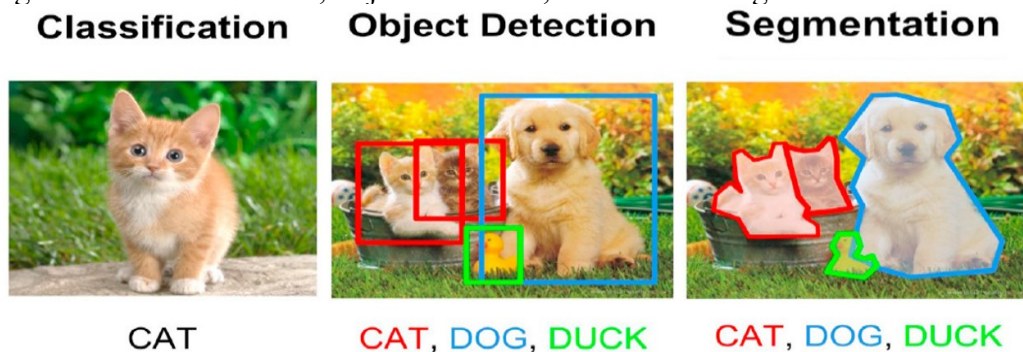
Source: Khan and Al-Habsi (2020).

Prior to the advent of deep learning, a process known as feature extraction was conducted to accomplish tasks like picture classification. During the deployment phase, these definitions are scanned for in other pictures. An image is classed as holding a certain object, such as a chair or a horse, if it shares a substantial number of traits with another bag of words.

The challenge with this conventional method lies in the need to determine the crucial aspects in each specific image. As the number of classes to categorize increases, the process of extracting features gets increasingly burdensome. The task of determining the most appropriate features to characterize various object classes is left to the discretion of the CV engineer, who must rely on their judgment and engage in a lengthy process of trial and error. In addition, every feature definition necessitates the management of a multitude of parameters, all of which must be meticulously adjusted by the computer vision engineer (O'Mahony *et al.*, 2019).

DL pioneered the idea of end-to-end learning, where a machine is provided with a dataset of photos that have been tagged with the classes of objects present in each image (Koehn, 1994). A DL model is trained on the provided data, allowing neural networks to identify the underlying patterns in different image classes. The model then automatically determines the most descriptive and prominent features for each specific item class. Due to the advanced techniques in computer vision (CV), the workflow of CV engineers has significantly transformed. The traditional practice of manually extracting features has been replaced by the skill of navigating through complex deep learning architectures, as shown in Figure 12.

Figure 12 – A comparison between two methodologies: (a) classical machine learning, and (b) deep learning
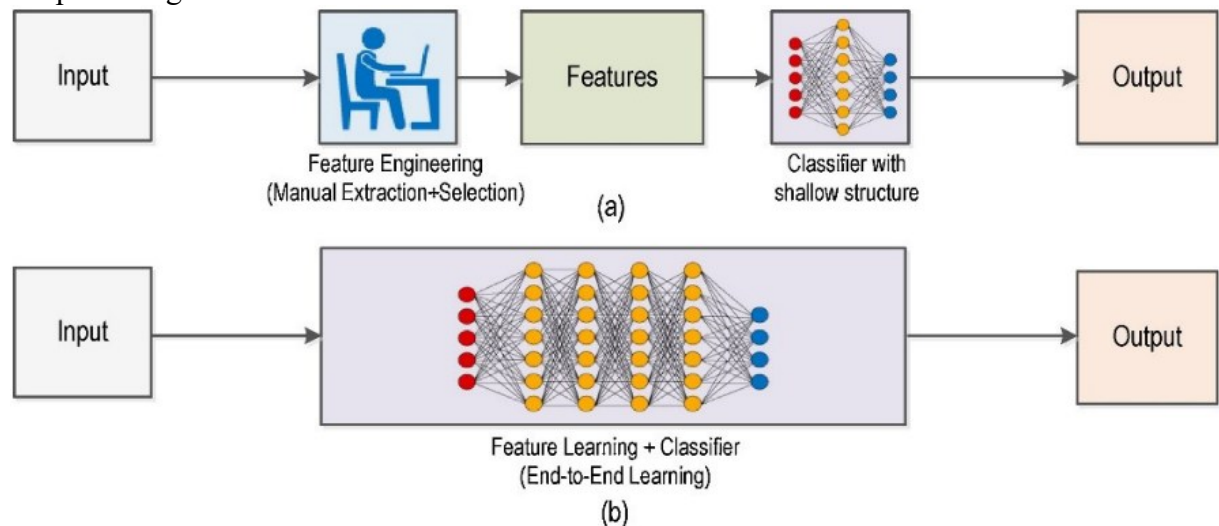
Source: Wang *et al.* (2018).

Deep learning (DL) might be excessive in certain cases, as classical computer vision (CV) approaches can often provide a more efficient solution to a problem with fewer lines of code than DL. Algorithms such as SIFT, as well as basic color thresholding and pixel counting algorithms, lack class specificity, meaning they are highly broad and yield consistent results regardless of the image. On the other hand, the features acquired from a deep neural network

are tailored to the specific characteristics of the dataset used for training. If the training dataset is not well designed, these features are unlikely to yield good results when applied to images that are distinct from the training set. Hence, the Scale Invariant Feature Transform (SIFT) and other algorithms are frequently employed for tasks like image stitching and 3D mesh reconstruction that do not necessitate specialized class knowledge. Training big datasets has demonstrated the feasibility of doing these tasks. However, this endeavor demands significant research resources and is not feasible for a specific application (O'Mahony *et al.*, 2019). Figure 13 depicts the domains, action, and paradigms of deep learning and classical computer vision.

Figure 13 – Computer vision domains, steps, and paradigms

Paradigm	Domain	Step/Action in the Processing Pipeline			
		I.Filtering and Pre-Processing	II.Conditioning and Simplifying	III.Labelling and Description	IV.Interpretation and Description
		Enhance the image	Simplify an image	Describe an image	Reason about an image
Classic Computer Vision	1.Value	Windowing and scale transform	Thresholding, arithmetic and logic operations with pixels	Histogram analysis	
	2.Space	Despeckling, mathematical morphology operations, convolutional and diffusion filters	Convolutional-based edge detection (Marr, Canny, Sobel, etc), saliency detection (Shaa'shua, etc), region-growing segmentation (watershed, Mumford-Shah, CSC, etc), skeletonization, split & merge	Spatial classifiers, classic ANNs (backpropagation, Boltzmann, Kohonen), statistic classifiers, classic machine learning (SVM, IBL, random forests, etc), rule-based classifiers, feature detection (SIFT, SURF, etc)	Spatial data structures (octrees, quadtrees), face-based models (DLFL, etc), boundary models, graph-based and spatial grammars, semantic networks
	3.Frequency	FFT- and wavelet-based filters (low-, band- and high-pass, etc)	FFT- and wavelet-based operations for frequency elimination, grouping and separation	FFT- and wavelet-based classification methods, Gabor filters	
Deep Learning	Convolutional Neural Networks	AlexNet, VGG, ResNet, Inception, YOLO, SSD, Mask-RCNN, etc			

Source: Matias *et al.* (2021).

Computer vision, when combined with deep learning applications, particularly using CNN models and frameworks, has made significant progress in automating the acquisition, processing, analysis, and extraction of valuable information from images or image sequences. It has a vital function in certain applications, such as image classification, segmentation, and object detection (O'Mahony *et al.*, 2019).

2.2.1 Computer Vision: Image Analytics Tasks

This section introduces the three primary uses of image analytics: image classification, image segmentation, and object detection.

2.2.1.1 Classification

The main goal of computer vision is to determine the existence of a specific object in an image, using a predefined set of objects that are being considered. Image classification involves analyzing an image that contains an object and determining which category it belongs to among the classes being studied. The objective of the model is to accurately predict the appropriate label for the given input image, where each class is allocated a distinct label (Ikram, 2019). Figure 14 displays of items classified inside the images.

Figure 14 – Image classification

Source: Krizhevsky, Sutskever and Hinton (2012).

2.2.1.2 Object Detection

Object detection is the process of accurately highlighting the rectangular areas that include all the objects in each scene, together with their respective labels. Object detection and

object localization are two distinct concepts. Object detection involves identifying and recognizing all items present in each scene. On the other hand, object localization focuses on precisely determining its location (Ikram, 2019). Figure 15 presents examples of objects detection.

Figure 15 – Object Detection
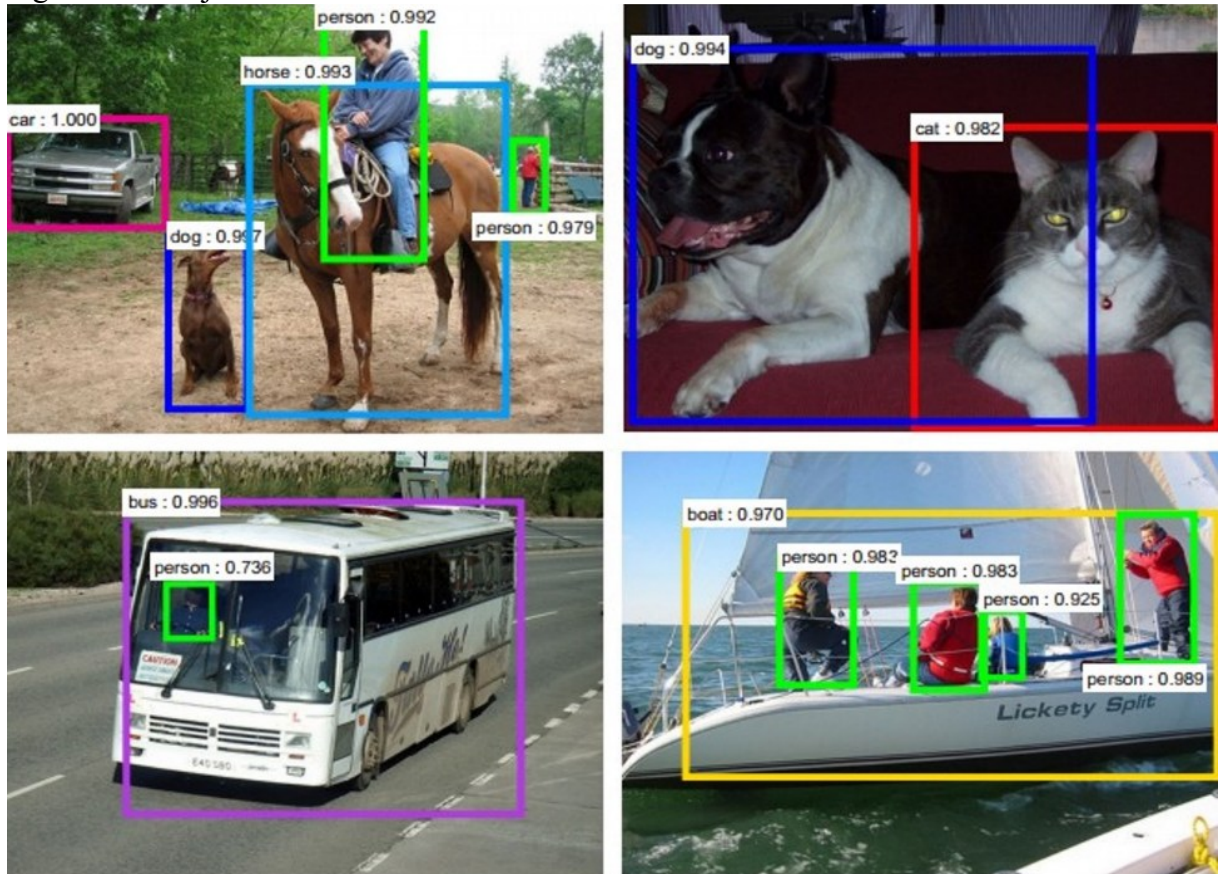
Source: Ren *et al.* (2017).

2.2.1.3 Segmentation

Image segmentation is the procedure of categorizing the pixels of an image into distinct groups according to the objects present in the image. Semantic segmentation is the procedure of attributing a specific label to every individual pixel in an image, ensuring that all pixels belonging to the same category are assigned the same label. If there are numerous things belonging to a specific category in the image, each object will be designated with an identical label. Instance segmentation is the procedure of assigning a specific class label to every pixel in an image, treating each individual item as a distinct class (Ikram, 2019). Figure 16 presents examples of image segmentation.

Figure 16 – Image Instance Segmentation
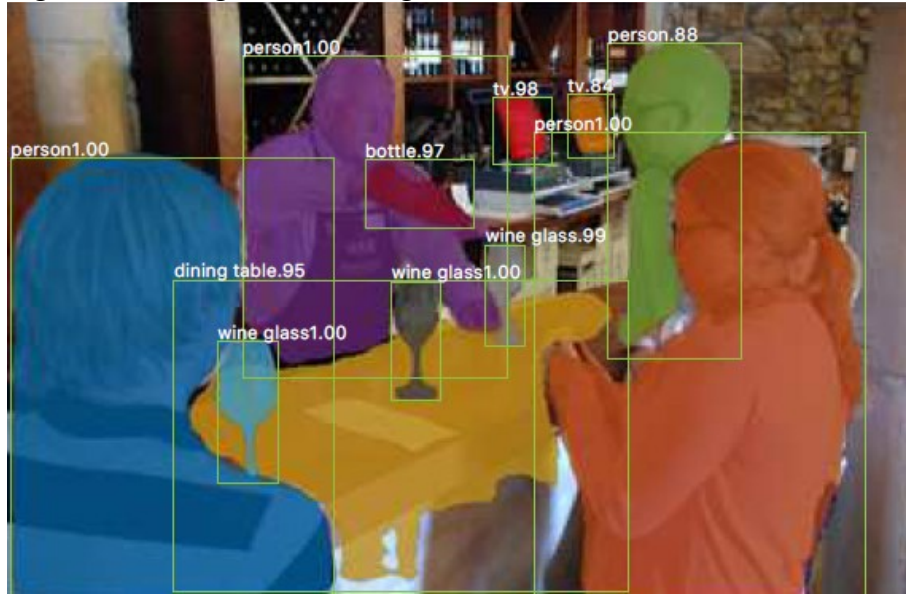
Source: He *et al.* (2017).

Deep learning typically achieves a performance in the range of billions of floating-point operations per second (FLOPS). When analyzing FLOPs in structures and architectures for various computer vision tasks, it is evident that categorization structures require significantly fewer processing resources. As an illustration, the computational capacity of classification architectures like VGG16 is 15.3 Giga Floating Point Operations per Second (GFLOPS) (Simonyan; Zisserman, 2014), yet ResNet50 only has 3.8 GFLOPS (He *et al.*, 2016). Conversely, the YOLOv3 design, frequently employed for object detection, possesses a computational capacity of 65 GFLOPS (Long *et al.*, 2020), while the U-Net architecture, utilized for picture segmentation, has a computational capacity of 30.98 GFLOPS (Ronneberger; Fischer; Brox, 2015). Therefore, in this study, the method of multi-class classification was utilized due to its significantly lower FLOPS and, as later proved, its favorable performance.

2.2.2 Convolutional Neural Networks

The objective of this topic is to provide a comprehensive overview of the structure, mathematical foundations, and applications of Convolutional Neural Networks (CNNs). This text will discuss the origin and development of these networks, providing a thorough explanation of their primary structures and groundbreaking methods. Moreover, this discussion will explore the versatility of CNNs in several domains, including image classification, object detection, and segmentation.

Convolutional Neural Networks (CNNs or ConvNets) are a specific form of artificial neural network that was designed to directly recognize visual patterns from image pixels with minimal preprocessing. CNNs have transformed conventional image processing pipelines, which usually entail sequential filtering, simplification, feature extraction, and classification, by executing these tasks simultaneously and so yielding improved outcomes. These networks have shown exceptional achievement in tasks related to computer vision and image identification by acquiring hierarchical features from input data, such as images. They do this by utilizing the underlying spatial structure and local connectedness of the data. The central component of a Convolutional Neural Network consists of convolutional layers, which utilize a collection of adaptable filters to process the input image (Zheng, 2018).

The Convolutional Neural Network (CNN) has gained prominence because to its superior accuracy in image recognition tasks, which may be attributed to two key aspects of its design: the utilization of three-dimensional inputs and the application of convolutions. Utilizing 3D input enhances the realism of the model by incorporating width, height, and depth dimensions, which represent color channels like RGB, Figure 17. On the other hand, convolutions improve the efficiency of the model by significantly reducing the number of operations and parameters required compared to a fully connected layer (Li, 2024).

Figure 17 – Left: A regular 3-layer Neural Network. Right: A ConvNet arranges its neurons in three dimensions (width, height, depth), as visualized in one of the layers
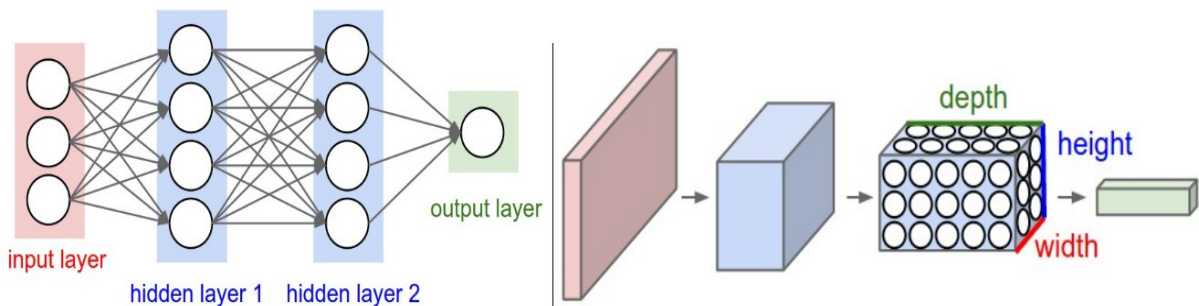
Source: Li (2024).

Each layer of the Convolutional Neural Network (CNN) can be understood as a distinct feature that the network extracts from the image (Figure 18). The initial layers of the CNN primarily capture fundamental properties like edges, contours, and geometric shapes (Li, 2024).

Figure 18 – Representation from the first five layers of a CNN

A Convolutional Neural Network (CNN) consists of three primary types of neural layers: convolutional layers, pooling layers, and fully connected layers. Every layer has a distinct function, Figure 19 shows an example of a simple schematic representation of a basic CNN. Each layer of a Convolutional Neural Network (CNN) modifies the input volume to produce an output volume of activated neurons. This process continues until the final fully connected layers are reached, which ultimately results in a mapping of the input data to a one-dimensional feature vector (Voulodimos *et al.*, 2018).

Figure 19 – Schematic diagram of a basic convolutional neural network (CNN) architecture
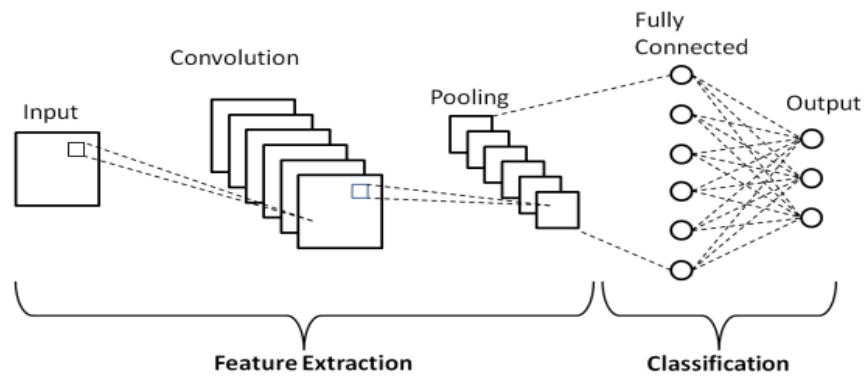
Source: Phung and Rhee (2019).

The layers are partitioned into two components: feature extraction and classification. Feature extraction comprises an input layer, a convolution layer, and a pooling layer, whereas classification includes a fully connected layer and an output layer. The input layer establishes a predetermined size for the input images, which can be adjusted if necessary. Next, the image undergoes convolution with several learnt kernels utilizing shared weights in the convolution layer. Subsequently, the pooling layer diminishes the dimensions of the image while endeavoring to preserve the encapsulated information. The results of the process of extracting features are referred to as feature maps. The classification process integrates the retrieved features into the fully connected layers. There is one output neuron for each object category in the output layer. The output of the classification part is the classification result (Phung; Rhee, 2019).

2.2.2.1 Convolutional Layer

Convolution is a specialized linear operation employed for feature extraction. It involves applying a small array of numbers, known as a kernel, across the input, which is

represented as an array of numbers, termed a tensor. At each location in the tensor, an element-wise product between the kernel elements and the corresponding input tensor elements is calculated, and the results are summed to obtain the output value in the corresponding position of the output tensor, referred to as a feature map. This process is repeated by applying multiple kernels, creating an arbitrary number of feature maps that represent distinct characteristics of the input tensors. These different kernels can be considered as distinct feature extractors. Convolutional layers, which are the core components of Convolutional Neural Networks (CNNs), apply these filters to the input to generate feature maps. The filters slide over the original image, performing convolution operations that detect local patterns such as edges, textures, and basic shapes (Yamashita *et al.*, 2018). The filter described in Figure 20 is simplistic for the main purpose to illustrate the mathematical operations involved in the convolution process.

Figure 20 – An example of convolution operation with a kernel size of 3 × 3, no padding, and a stride of 1.
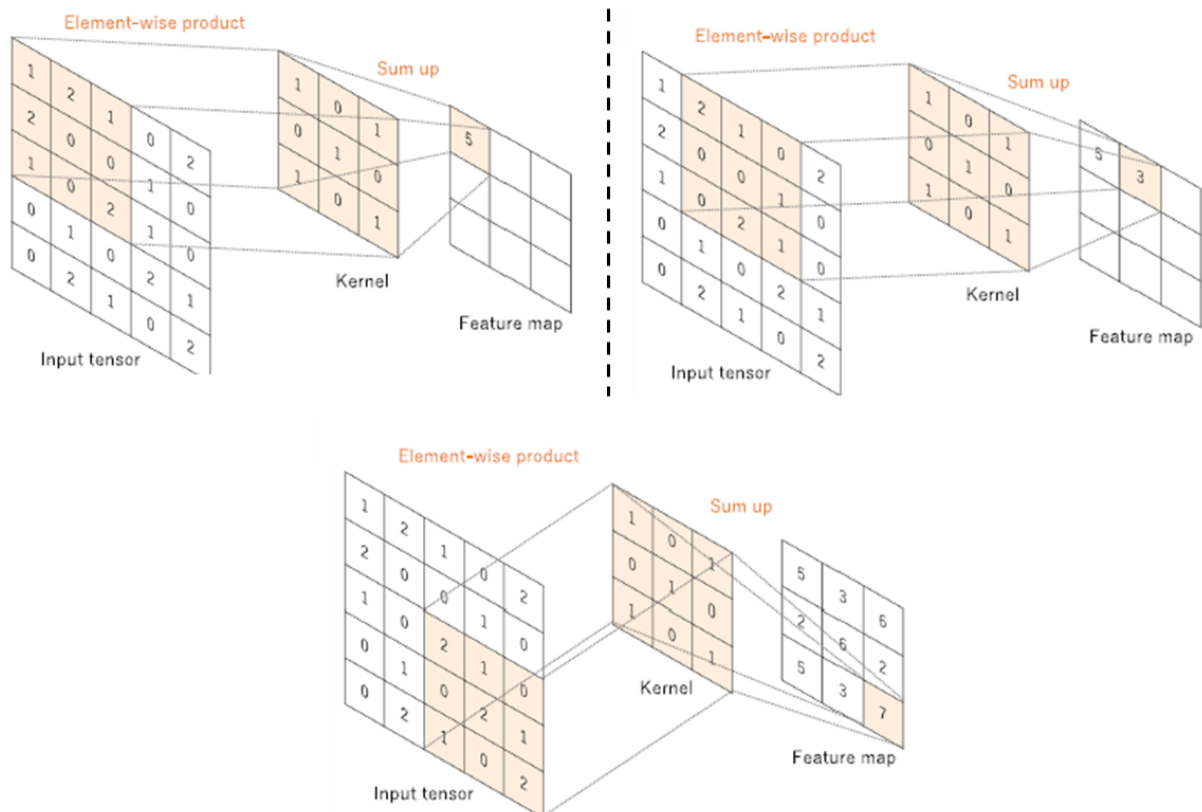
Source: Yamashita *et al.* (2018).

In contrast, Figure 21 presents examples of actual visualizations of the filters from the first convolutional layer of a trained network. However, the underlying argument remains the same: the filters in the initial layer of the Convolutional Neural Network convolve across

the input image and become "activated" when they detect the specific feature they are designed to recognize within the input volume.

Figure 21 – Filter representation from the first layers of a CNN
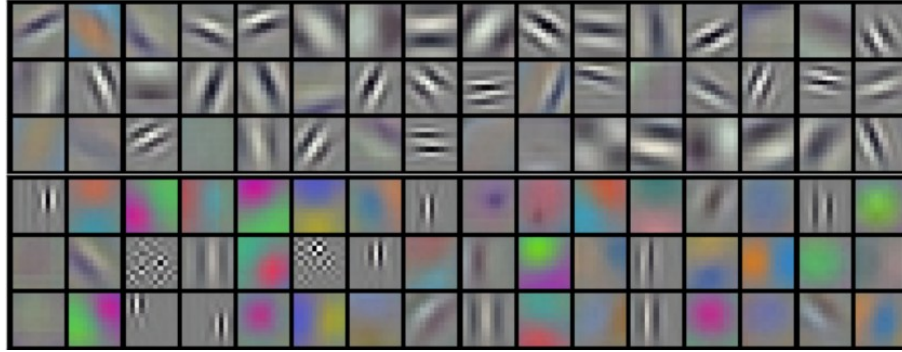
Source: Krizhevsky, Sutskever and Hinton (2012).

2.2.2.2 Pooling Layers

Pooling layers are an integral component of Convolutional Neural Networks, responsible for reducing the spatial dimensions (width × height) of the input volume while preserving the depth dimension. This process, known as subsampling or down sampling, involves applying a filter with a specific stride to each subregion of the input volume, utilizing pooling functions such as max pooling and average pooling (Voulodimos *et al.*, 2018). As demonstrated by Scherer, Muller and Behnke (2010), max pooling can contribute to faster convergence, improved invariant feature selection, and enhanced generalization capabilities. The primary purpose of pooling layers is to downsample the input activation volume, which helps to decrease the number of parameters and computational requirements of the network, as well as mitigate overfitting. Although this size reduction results in some information loss, it is a beneficial operation for the network as it decreases computational overhead and enhances the model's generalization abilities, allowing it to learn more robust and abstract feature representations (Ikram, 2019).

Max Pooling involves selecting the maximum element from each window in the feature map. Consequently, the output after the max-pooling layer is a feature map that encapsulates the most dominant features identified in the preceding feature map. On the other hand, Average Pooling calculates the average of the elements within the region of the feature map covered by the filter. It essentially computes the meaning of the features from the feature map. Figure 22 shows example for the max-pooling and the average-pooling (O'Mahony *et al.*, 2020-).

Figure 22 – Example for the max-pooling and the average-pooling with a filter size of 2×2 and a stride of 2×2

(a) Max-Pooling   **(b)** Average-Pooling
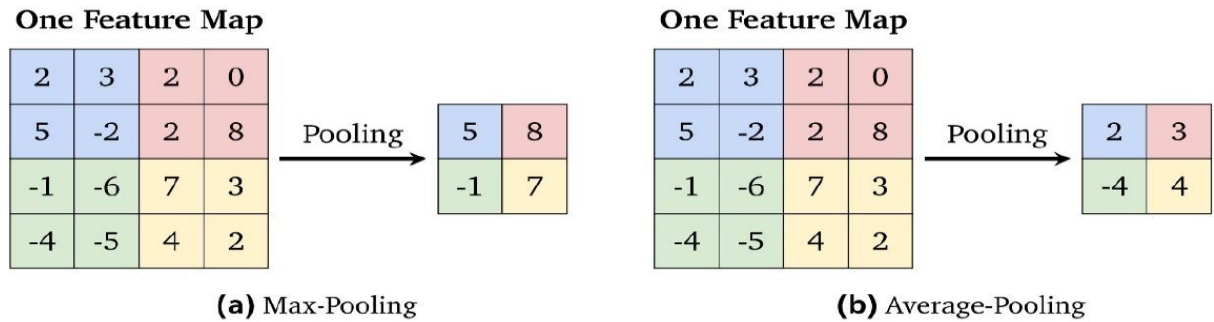
Source: adapted from Li (2024).

2.2.2.3 Fully Connected Layers

        The final stages of a convolutional neural network often involve fully connected layers, which are similar to the layers in a standard neural network. In these fully connected layers, each neuron has full connections to all activations in the preceding layer, and their activations are computed through matrix multiplication and a bias offset. The fully connected layers transform the 2D feature maps generated by the earlier convolutional and pooling layers into a 1D feature vector (Voulodimos *et al.*, 2018). This feature vector can then be used for tasks such as classification, where it is fed forward into a set of output categories (Krizhevsky; Sutskever; Hinton, 2012), or further processing, where it serves as a high-level representation of the input data (Girshick *et al.*, 2014). The flattened matrix goes through a fully connected layer to classify the images, Figure 23.

Figure 23 – The fundamental structure of a Convolutional Neural Network (CNN)
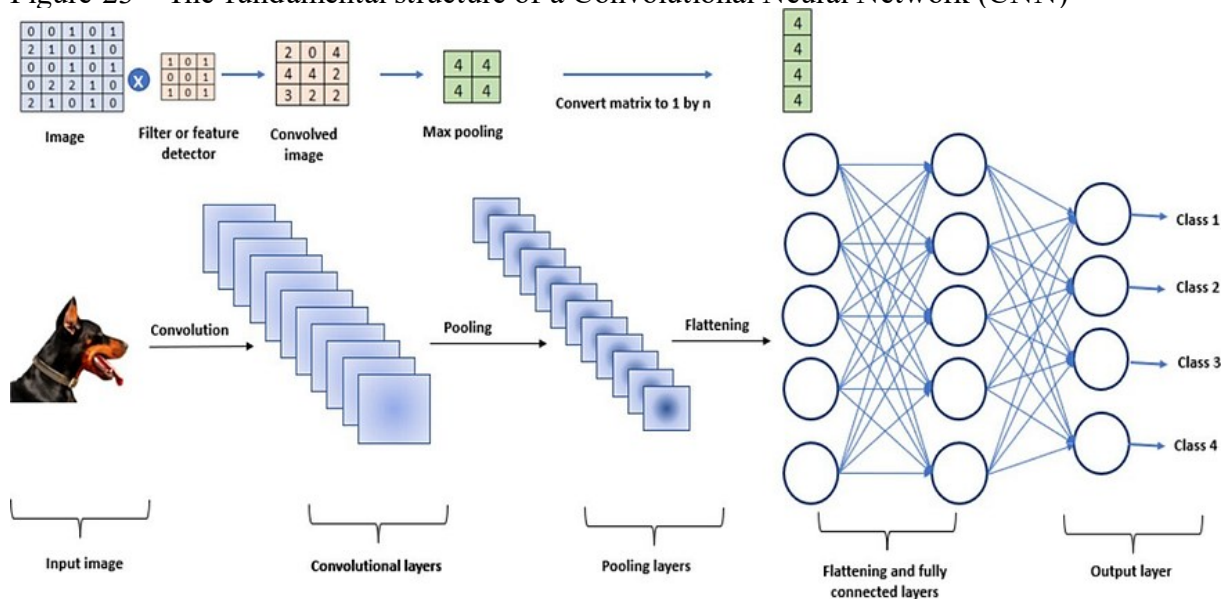
Source: Apeagyei, Ademolake and Adom-Asamoah (2023).

2.2.3 Mathematical foundations of deep learning

Deep learning is a specialized branch of machine learning that employs artificial neural networks consisting of many layers. Deep neural networks. This topic explores the mathematical principles that support deep learning, with a specific focus on the multi-layer neural networks that form the cornerstone of this field.

The structure of a multi-layer neural network is composed of three types of main layers, Figure 24. The input layer is responsible for receiving the input data and transmitting it to the subsequent layers. The hidden layers of the network perform nonlinear transformations on the incoming inputs, enabling the network to learn and represent intricate relationships in the data. The output layer is responsible for generating the ultimate output of the network, which represents the results obtained from the processing performed by the preceding layers (Goodfellow; Bengio; Courville, 2016).

Figure 24 – Deep Neural Network Architecture
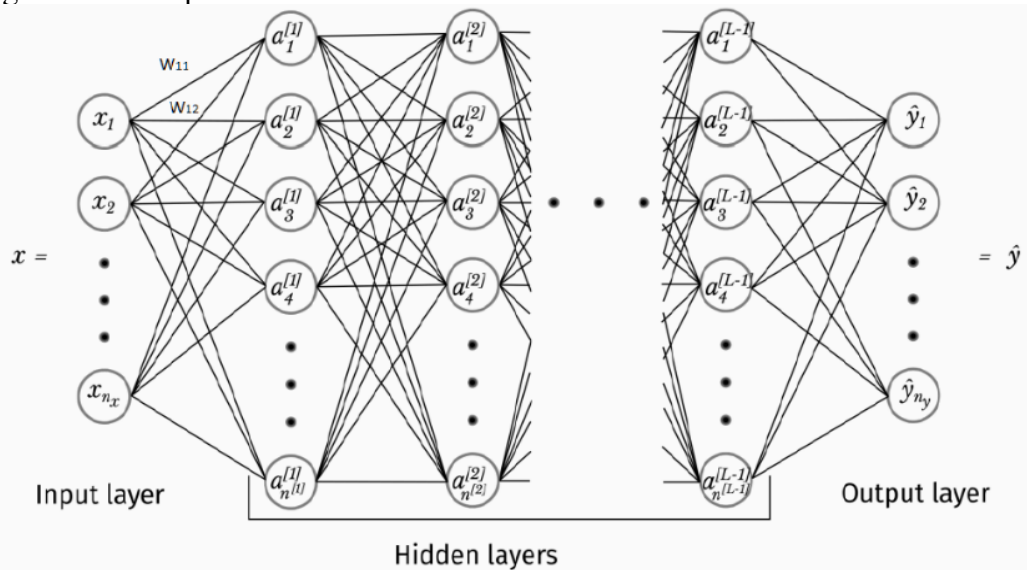
Source: Ikram (2019).

In the context of neural networks, each neuron performs a weighted input operation before applying the activation function. This operation begins by multiplying each input x_i by a corresponding weight w_i. The weights are values that the network adjusts during training to minimize prediction error. After multiplication, a bias term b is added, which allows the network to adjust the activation function to better fit the data. The mathematical formula for the weighted input z in a neuron is given by Equation 6:

$$z = \sum_{i=1}^{n} w_i x_i + b \qquad (6)$$

Where *n* is the number of inputs. This process can be viewed as a dot product between the input vector and the weight vector, with the addition of the bias. This linear combination of inputs allows the network to model linear relationships between the input variables (Goodfellow; Bengio; Courville, 2016).

After calculating the weighted input *z*, the next step is to apply the activation function *σ(z)*. The activation function is crucial because it introduces non-linearity into the network, allowing it to learn and represent complex relationships in the data. Without this non-linearity, the network would be equivalent to a simple linear regression, regardless of the number of layers. There are several popular activation functions, each with its characteristics (Table 24) (Nielsen, 2015).

Table 24 – Activation function

(continue)
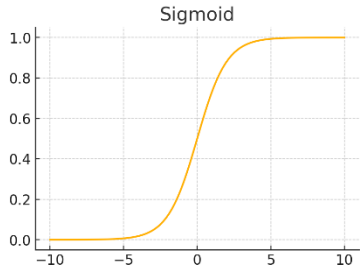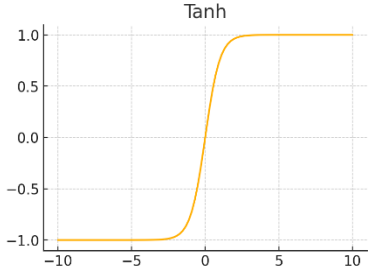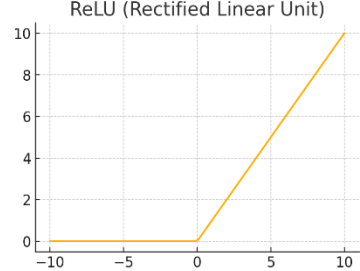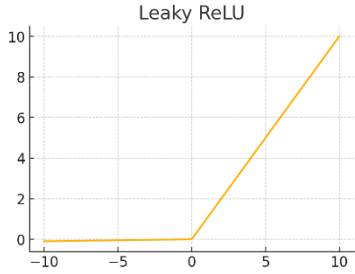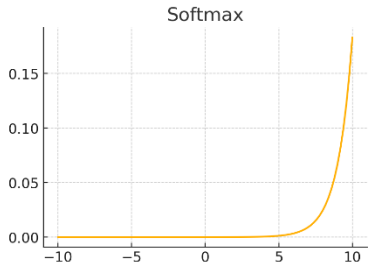
Activation Function	Plot	Mathematical Formula	Characteristics
Sigmoid		$\sigma(x) = \dfrac{1}{1 + e^{-x}}$	Smooth, non-linear, values between 0 and 1, mainly used in binary classification outputs.
Tanh		$\tanh(x) = \dfrac{e^x + e^{-x}}{e^x + e^{-x}}$	Smooth, non-linear, values between -1 and 1, used in recurrent networks.
ReLU (Rectified Linear Unit)		$\text{ReLU}(x) = \max(0, x)$	Non-linear, computationally efficient, helps avoid vanishing gradients, widely used in deep networks.

Table 24 – Activation function

Activation Function	Plot	Mathematical Formula	Characteristics
Leaky ReLU		$\text{Leaky ReLU}(x) = \max(0.01x, x)$	Variant of ReLU, allows a small gradient when x < 0, helps prevent dead neurons.
Softmax		$\sigma(x_i) = \dfrac{e^{x_i}}{\sum_j e^{x_j}}$	Converts a vector of values into a probability distribution, used in the output layer of multi-class classifiers.

Source: elaborated by the author.

These activation functions enable the neural network to capture and model complex relationships in the data (Nielsen, 2015).

2.2.3.1 Forward propagation

Forward propagation each node performs a weighted sum of the inputs from previous nodes and applies an activation function over the result to get the final activation value. The following formula gives activations (Ikram, 2019).

This process begins at the input layer, where the initial data $x^{(0)}$ is fed into the network. Each subsequent layer l in the network transforms the input from the previous layer using a series of weighted sums and activation functions. For each layer, the weighted input $z^{(l)}$ is calculated as Equation 7:

$$z_i^{[l]} = \sum_{j=1}^{n^{[l-1]}} w_{ji}^{[l]} a_i^{[l-1]} + b_i^{[l]} \qquad (7)$$

Where $W^{(l)}$ represents the weights, $a^{(l-1)}$ is the activation from the previous layer, and $b^{(l)}$ is the bias. The activation $a^{(l)}$ is then computed by applying an activation function σ/sigmaσ to the weighted input (Equation 8):

$$a_i^{[l]} = \sigma\left(z_i^{[l]}\right) \tag{8}$$

This process is repeated for each layer until the final layer, where the network produces its output $a^{(L)}$. Forward propagation allows the network to transform raw input data into meaningful predictions or classifications by leveraging the learned weights and biases throughout its layers. The efficiency and effectiveness of this transformation depend on the architecture of the network and the quality of the training process (Goodfellow; Bengio; Courville, 2016).

2.2.3.2 Loss Functions

The loss function, sometimes referred to as the cost function or objective function, plays a crucial role in the training of neural networks. The term "quantifies" refers to the process of measuring or determining the extent of something. In this context, it is used to describe the calculation of the difference between the projected outputs produced by the network and the actual target values (Yamashita *et al.*, 2018). Quantification is important as it provides a metric to assess the model's performance, whether it is positive or negative. By reducing this loss, it directs the network to generate more precise predictions. The selection of the loss function is contingent upon the nature of the task at hand, such as whether it involves regression, binary classification, or multi-class classification (Goodfellow; Bengio; Courville, 2016).

Mean Squared Error (MSE): For regression tasks, where the goal is to predict continuous values, the Mean Squared Error (MSE) is a commonly used loss function. MSE calculates the average of the squares of the errors, where the error is the difference between the predicted value $\hat{y}_i$ and the actual value y_i. The formula for MSE shows Equation 9:

$$L(\hat{y}, y) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{9}$$

Where n is the number of data points. The squaring of the errors ensures that both positive and negative errors contribute positively to the loss, and it penalizes larger errors more severely. Minimizing MSE encourages the model to produce predictions that are closer to the actual values, reducing the overall error (Goodfellow; Bengio; Courville, 2016).

Cross-Entropy Loss: For classification tasks, particularly binary classification, the Cross-Entropy Loss, also known as Logarithmic Loss, is widely used. It measures the

performance of a classification model whose output is a probability value between 0 and 1. The Cross-Entropy Loss increases as the predicted probability diverges from the actual label. The formula for Cross-Entropy Loss in binary classification is Equation 10:

$$L(\hat{y}, y) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{10}$$

In this formula, y_i represents the actual binary label, and $\hat{y}_i$ is the predicted probability of the positive class. This loss function penalizes predictions that are confident but wrong more than predictions that are less confident, thus encouraging the model to output probabilities that reflect uncertainty when it is unsure (Bishop, 2006-).

Categorical Cross-Entropy Loss: For multi-class classification problems, where the task is to classify inputs into one of several classes, the Categorical Cross-Entropy Loss is used. This loss function is an extension of the binary cross-entropy to multiple classes. It is calculated as Equation 11:

$$L(\hat{y}, y) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{c=1}^{C} y_{ic} \log(\hat{y}_{ic}) \tag{11}$$

Where C is the number of classes, y_{ic} is a binary indicator (0 or 1) if class label ccc is the correct classification for observation i, and $\hat{y}_{ic}$ is the predicted probability of observation i, being in class c. This loss function ensures that the predicted probability distribution is close to the true distribution (Goodfellow; Bengio; Courville, 2016).

2.2.3.3 Backward propagation

Backward propagation, sometimes referred to as backpropagation, a fundamental method for training neural networks, aims to adjust the weights between the nodes to minimize the error function. This algorithm operates by transmitting the errors retroactively from the output layer to the input layer, allowing the network to modify its parameters to reduce overall error. The process involves two main stages: the forward pass, where input data is propagated through the network to compute the output, and the backward pass, where gradients of the loss function concerning each parameter are calculated and used to adjust the parameters, thereby achieving error minimization (Hecht-Nielsen, 1992).

During the backward pass, the backpropagation algorithm calculates the gradients of the loss function with respect to each weight and bias in the network. This is done using the chain rule of calculus, which allows the gradient of a composite function to be computed as the product of the gradients of its constituent functions. For a weight $w_{ij}^{(l)}$ in layer l, the gradient of the loss L with respect to this weight is given by Equation 12:

$$\frac{\partial L}{\partial w_{ij}^{(l)}} = \frac{\partial L}{\partial a_j^{(l)}} \cdot \frac{\partial a_j^{(l)}}{\partial z_j^{(l)}} \cdot \frac{\partial z_j^{(l)}}{\partial w_j^{(l)}} \tag{12}$$

Where $a_j^{(l)}$ is the activation of neuron j in layer l, $z_j^{(l)}$ is the weighted input to neuron j in layer l, and $w_{ij}^{(l)}$ is the weight connecting neuron i in layer l-1 to neuron j in layer l. By applying the chain rule in this manner, the algorithm efficiently computes the gradients for all weights and biases in the network (Nielsen, 2015).

2.2.3.4 Gradient Descent

After calculating the gradients, the subsequent stage in backpropagation involves adjusting the weights and biases to minimize the loss. Gradient descent is commonly employed to do this, which is an optimization procedure that updates each parameter by moving in the opposite direction of the gradient of the loss function with respect to that parameter. The update rule for a weight $wij(l)$ is as Equation 13:

$$w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta \frac{\partial L}{\partial w_{ij}^{(l)}} \tag{13}$$

Where η is the learning rate, a hyperparameter that controls the size of the steps taken during the optimization process. The learning rate must be chosen carefully; too high a learning rate can cause the training process to diverge, while too low a learning rate can result in slow convergence. The same update rule is applied to the biases in the network. By iteratively updating the weights and biases using the gradients, the network gradually learns to produce outputs that are closer to the target values (Goodfellow; Bengio; Courville, 2016).

Gradient descent implementation requires iterative parameter updates. There are three types of Gradient Descent: Batch Gradient Descent, Stochastic Gradient Descent, and Mini-Batch Gradient Descent, as depicted in Figure 25.

Figure 25 – Types of Gradient Descent
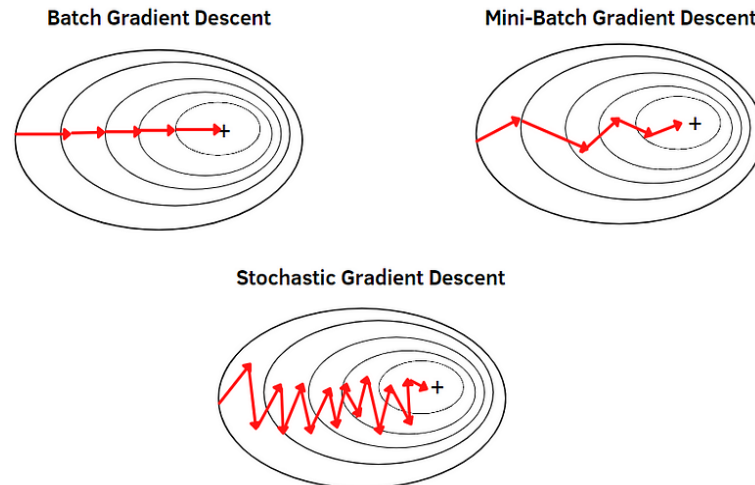
Source: Nguyen (2023).

Batch Gradient Descent: Batch gradient descent, often referred to as vanilla gradient descent, involves computing the gradient of the loss function using the entire training dataset at each iteration. Specifically, it calculates the average of the gradients for all training examples before updating the model's parameters. This method ensures stability during the training process, as it leverages the full dataset to make informed updates (Goodfellow; Bengio; Courville, 2016). However, batch gradient descent can be computationally expensive, especially when dealing with large datasets, as it requires processing the entire dataset before each update. Additionally, the method may lead to slower convergence in the presence of noisy or redundant data due to the averaging effect, which can smooth out the gradient's trajectory (Bishop, 2006-).

Stochastic Gradient Descent (SGD): Stochastic Gradient Descent (SGD) adopts a different strategy by updating the model parameters for each training example individually. Instead of computing the gradient based on the entire dataset, SGD calculates the gradient using only one randomly selected training example at a time (Goodfellow; Bengio; Courville, 2016). This approach significantly increases the update frequency, making SGD faster than batch gradient descent in terms of per-iteration updates. SGD's ability to quickly adapt to changing patterns in the data is advantageous, particularly in dynamic environments. However, the method can exhibit more oscillations in the loss function due to the high variance introduced

by individual samples. This noise can lead to longer convergence times, as the updates may frequently overshoot the optimal solution (Nielsen, 2015).

Mini-Batch Gradient Descent: Mini-Batch gradient descent strikes a balance between batch gradient descent and stochastic gradient descent. It computes the gradient using a small subset, or mini batch, of training examples. This approach combines the computational efficiency of SGD with the stability of batch gradient descent (Goodfellow; Bengio; Courville, 2016). By utilizing mini batches, the algorithm reduces the noise inherent in SGD while maintaining a more frequent update cycle than batch gradient descent. This results in a more accurate estimate of the true gradient, facilitating faster and more stable convergence (Bishop, 2006-). Mini-batch gradient descent is particularly effective in parallel and distributed computing environments, as it allows for efficient processing of training data in manageable chunks (Nielsen, 2015).

Gradient descent is an essential approach used to minimize a cost function and determine the optimal values for model parameters. The learning rate (LR) is important in defining the rate at which the algorithm reaches the minimum of the cost function. The learning rate essentially determines the magnitude of the step made during each iteration of the gradient descent process (Goodfellow; Bengio; Courville, 2016). The influence of the learning rate on two situations: a learning rate that is too small and a learning rate that is too large, Figure 26.

Figure 26 – Gradient descent – The learning rate
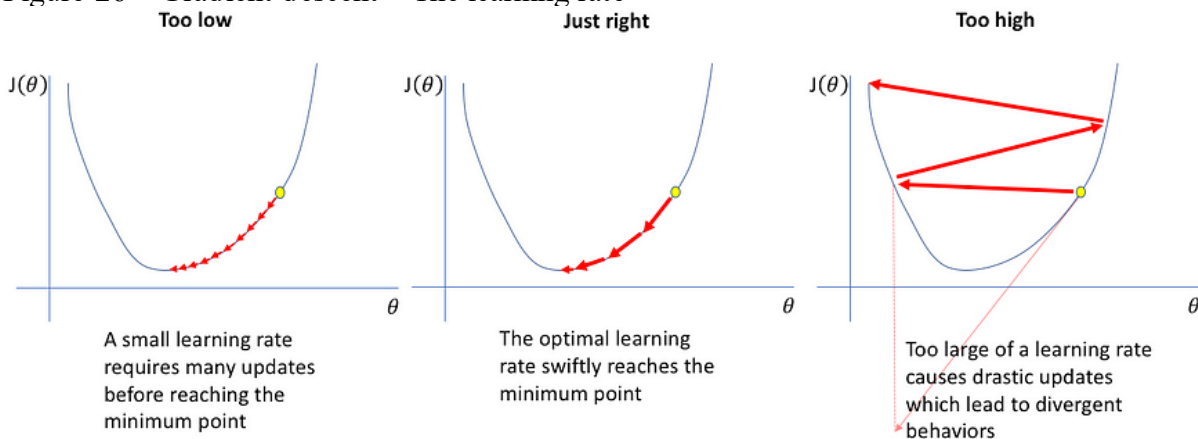
Source: Nguyen (2023).

When the learning rate is set to a very small value, the algorithm proceeds by minuscule steps towards the minimum of the cost function. These tiny steps can render the convergence process exceedingly slow. For instance, envision progressing gradually and cautiously towards a certain destination–it would necessitate a considerable duration to achieve

your objective. Similarly, when using a low learning rate, the gradient descent algorithm requires a large number of iterations to get close to the minimum, resulting in delayed convergence. The sluggish convergence in actual applications can be inefficient, particularly when working with huge datasets or complex models (Bishop, 2006-).

On the other hand, if the learning rate is set to a very large value, gradient descent can overshoot the minimum and fail to converge. When the learning rate is high, the algorithm takes significant strides towards the minimum, but it may consistently go beyond it, resulting in a rise in the cost function instead of a drop. This might result in divergence, a situation where the algorithm is unable to discover the most optimal solution and continuously goes further away from the minimum This tendency not only hinders the model from reaching convergence, but it can also lead to unstable training processes, characterized by erratic oscillations of the model parameters without settling on a solution (Nielsen, 2015).

Ensuring an optimal learning rate is essential for the efficient and effective training of machine learning models. An appropriately chosen learning rate can significantly enhance the convergence speed and ensure that the gradient descent algorithm finds the optimal solution. Methods such as learning rate schedules and adaptive learning rates (such as AdaGrad, RMSprop, and Adam) have been created to dynamically modify the learning rate while training, thus enhancing the stability and effectiveness of gradient descent algorithms (Goodfellow; Bengio; Courville, 2016).

Although batch gradient descent is generally reliable and consistent, it may occasionally become trapped in a local minimum and fail to discover the global minimum of the cost function. This constraint emerges due to the utilization of the complete dataset in batch gradient descent for gradient computation. This approach has the ability to diminish subtle variations and cause the algorithm to converge towards the closest local minimum (Goodfellow; Bengio; Courville, 2016). On the other hand, stochastic gradient descent (SGD) and mini-batch gradient descent are more suited for avoiding local minima (Nielsen, 2015).

One often used method to tackle the problem of local minimum is to systematically decrease the learning rate or step size as gradient descent approaches the minimum (Figure 27). This method, referred to as learning rate annealing or decay, guarantees that the algorithm's steps gradually decrease in size as it gets closer to the minimum. By gradually decreasing the learning rate, the algorithm reduces the chances of surpassing the minimum value, hence improving its capacity to converge at or close to the global minimum (Bishop, 2006-).

Figure 27 – Multiple Local Minima
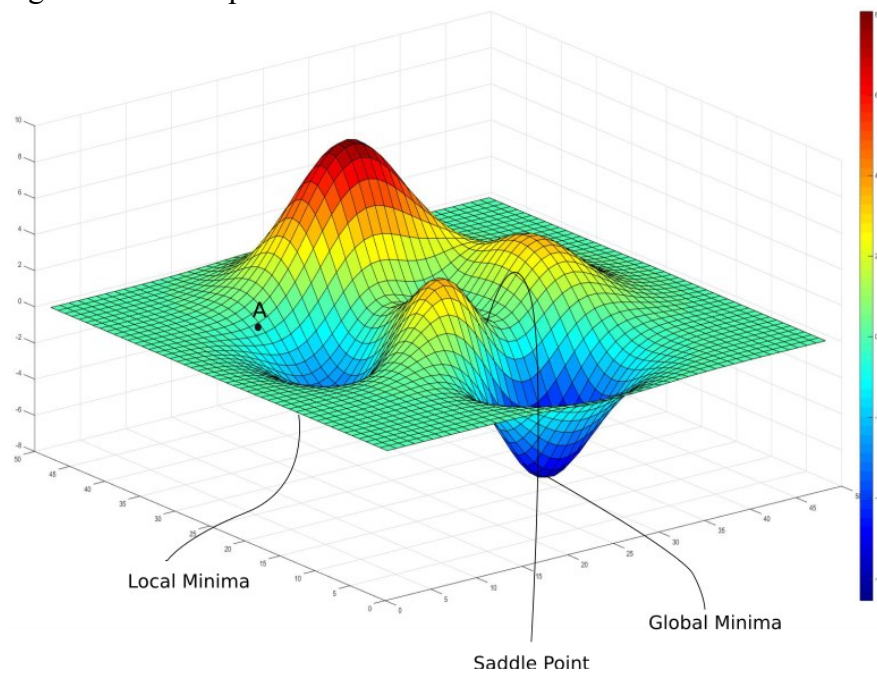
Source: Nguyen (2023).

2.2.4 Practical aspects of training a Convolutional Neural Network (CNN)

This subtopic focuses on the development of datasets for training, testing, and validation, in addition to approaches to prevent over-fitting and under-fitting. It covers methods such as data normalization, regularization, dropout, batch normalization, and data augmentation.

2.2.4.1 Training, validation and test

Training Set: Firstly, the model undergoes training using the training set. This dataset is only utilized for updating the weights of the model's layers, with the objective of minimizing the loss function and enhancing performance. The training set is essential for the model to acquire knowledge of patterns and relationships present in the data. Nevertheless, it is crucial to acknowledge that the model's performance on the training set does not provide a conclusive indication of its capacity to generalize to unfamiliar data. The primary objective is for the model to exhibit strong performance on the test data, demonstrating its capacity to generalize beyond the training instances (Goodfellow; Bengio; Courville, 2016). Typically, 60-80% of the dataset should be allocated for training. This set is used to train the model by adjusting the weights based on the loss function (NG, 2012).

Validation Set: The main objective of the validation set is to assess the performance of the model while it is being trained. This dataset is utilized for optimizing hyperparameters and for evaluating the model's performance to prevent overfitting. In contrast to the training set, the validation set is typically utilized to evaluate the model's ability to generalize to new data, however the model does not directly learn from this set. Through the utilization of the validation set, one may fine-tune the model architecture, hyperparameters, and other configurations to enhance performance. Using the test set for this purpose could result in the model becoming biased towards doing well on the test set, which can lead to overfitting and limited ability to generalize to completely new data (Bishop, 2006-). Allocate 10-20% of the dataset for validation. This set is used to tune hyperparameters and make decisions about model architecture. It helps monitor the model's performance during training and prevent overfitting (NG, 2012).

Test Set: The test set is exclusively utilized at the conclusion of the training procedure to acquire an authentic assessment of the model's accuracy and error rate. Accuracy is a measure of the proportion of successfully predicted examples, whereas the error rate is the proportion of incorrectly predicted examples. The test set should be used exclusively for the final evaluation to guarantee an impartial assessment of the model's performance. Employing the test set during the training process would undermine its objectivity, since the model may become too calibrated to this data, thereby impeding its ability to generalize to novel, unseen instances. Hence, the test set functions as the definitive standard for assessing the model's ability to generalize (Nielsen, 2015). Reserve 10-20% of the dataset for testing. This set is used only once after the model is fully trained to evaluate its performance on unseen data (NG, 2012).

2.2.4.2 Network Capacity, Over-fitting and under-fitting

Network capacity is a critical concept in model training, referring to a model's ability to approximate various functions. A model with excessively high capacity can overfit the training data, capturing noise and leading to poor generalization on new data. Conversely, a model with too low capacity may underfit, failing to capture the underlying patterns in the data. As Bishop (2006). explains, selecting the appropriate network architecture and regularization techniques is crucial for balancing these issues. Each model acts as a function approximator from inputs to outputs, and its capacity determines its ability to represent a broad range of functions. Insufficient capacity results in underfitting, where the model cannot fit the training data properly, leading to poor test performance. On the other hand, excessive capacity

results in overfitting, where the model memorizes training data noise, also leading to poor test performance (Goodfellow; Bengio; Courville, 2016). Low bias and low variance will give a balanced model, whereas high bias leads to underfitting, and high variance lead to overfitting, Figure 28.

Figure 28 – Bias vs. Variance
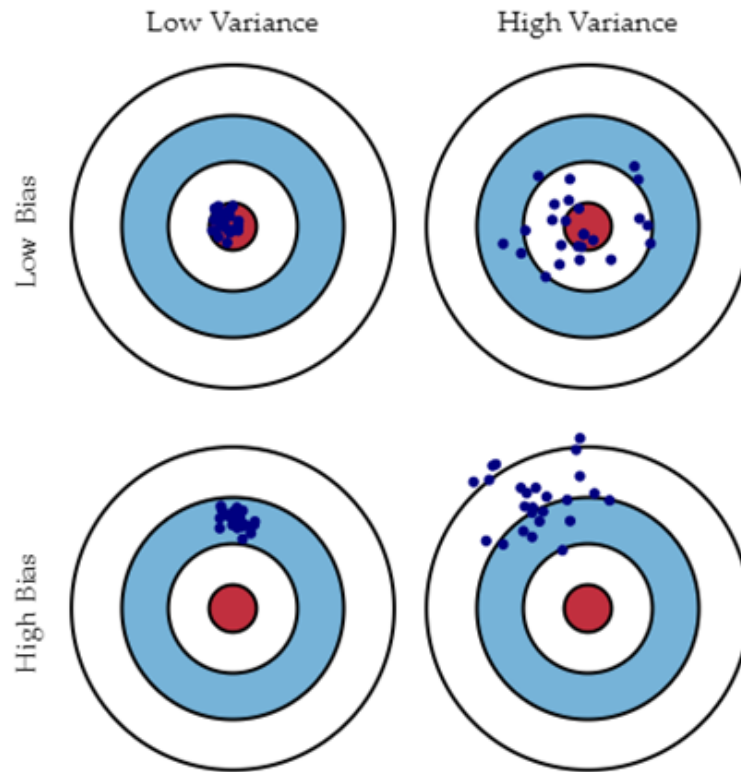
Source: Goodfellow, Bengio and Courville (2016).

Hypothesis space, which is the set of functions that the learning algorithm can choose from, plays a vital role in managing model capacity. For instance, in linear regression, the hypothesis space consists of all linear functions, which may not suffice for classifying nonlinear data. Expanding the hypothesis space to include quadratic or higher-degree polynomial functions increases the model's capacity. Generalization refers to a model's ability to perform well on unseen data, with generalization error measured over the test set. Underfitting occurs when the model fails to achieve a low error rate, often due to inadequate capacity. Overfitting is evident when there is a significant discrepancy between training and test error, indicating the model's capacity exceeds the task requirements. For optimal performance, the model's capacity should align with the task's true requirements, considering the amount of training data available, Figure 29 (Jabbar; Khan, 2014).

Figure 29 – Underfitting, overfitting

$$\theta_0 + \theta_1 x$$

High Bias
(Underfitting)

$$\theta_0 + \theta_1 x + \theta_2 x^2$$

Low Bias, Low Variance
(Goodfitting)

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

High Variance
(Overfitting)
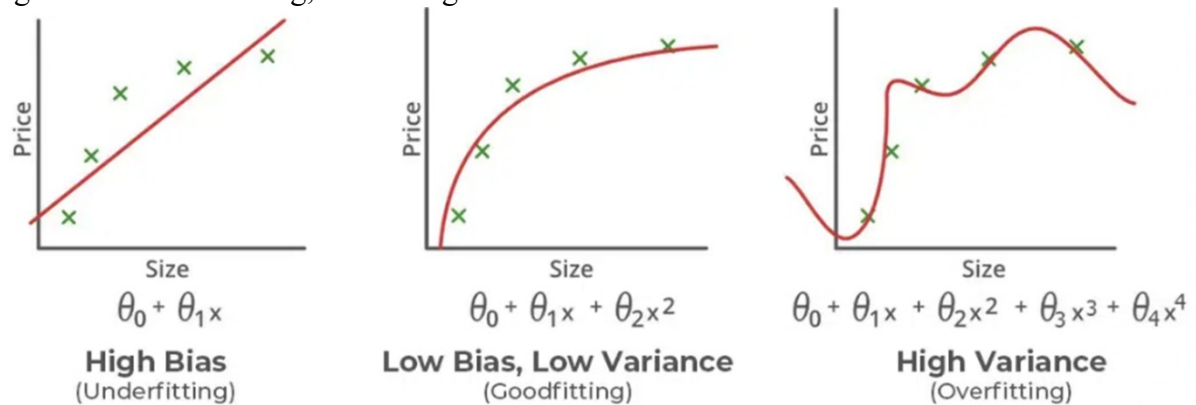
Source: NG (2012).

It is essential to recognize that while simpler functions tend to generalize better, exhibiting a smaller gap between training and test errors, it is still necessary to select a sufficiently complex hypothesis to achieve low training error. Generally, training error decreases and approaches its minimum possible value as model capacity increases, provided the error measure has a defined minimum. However, the generalization error typically follows a U-shaped curve relative to model capacity, initially decreasing and then increasing as the model begins to overfit the training data, this is illustrated in Figure 30 (Heaton, 2018).

Figure 30 – Capacity
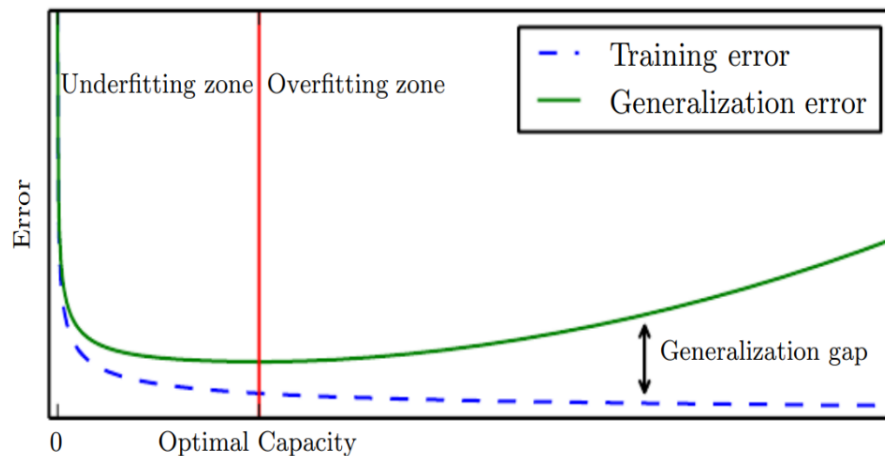
Source: Heaton (2018).

2.2.4.3 Data Scaling and Normalization

One of the significant challenges in handling datasets for machine learning is the discrepancy in feature scales. In many datasets, features are not scaled uniformly, which means some features can have much larger values compared to others, not necessarily indicating a

substantial change. For instance, a feature might be represented in millions, showing large changes between samples, while another feature might be in the order of hundreds, showing only minor changes. This variation in scales can lead to problems during model training as the learning algorithm might disproportionately focus on features with larger scales, potentially overlooking important patterns in smaller-scale features. To address this, feature scaling is employed to adjust the values of features to a common scale, which is crucial in many machine learning problems (Bishop, 2006).

In the context of image datasets, scaling concerns are somewhat mitigated as pixel values for all channels (RGB) typically range from 0 to 255, ensuring uniform scaling across features. However, this does not eliminate the need for normalization. Data normalization is the process of rescaling input features to a standard range, such as [0, 1] or [-1, 1]. This step is essential because it ensures that each input feature contributes equally to the learning process, preventing any feature from dominating the training process due to its scale. Normalization also enhances the convergence speed of gradient descent by smoothing the optimization landscape, which is supported by LeCun, Bottou, Orr, *et al.* (1998) who demonstrated that normalized inputs can significantly improve the efficiency of learning algorithms by ensuring a more stable and faster convergence.

Normalization can be particularly effective in improving the performance of Convolutional Neural Networks (CNNs). By transforming the pixel values to a standardized range, CNNs can achieve better generalization and more stable training. This standardization helps in managing the dynamic ranges of different images, leading to more robust feature extraction and ultimately improving the model's predictive accuracy (Ioffe; Szegedy, 2015). For instance, in CNNs, normalization layers, such as Batch Normalization, are used to standardize intermediate activations, further stabilizing and accelerating the training process.

2.2.4.4 Regularization

Regularization is a critical modification to the learning algorithm aimed at reducing the generalization error of a model without affecting its training error. The primary objective of regularization is to prefer simpler solutions within the hypothesis space, meaning that when multiple functions can be selected based on the model's capacity, the simpler function is chosen. This preference for simplicity helps to prevent overfitting by imposing constraints on the model's parameters, typically by favoring smaller weights. Overfitting occurs when a model learns noise and inaccuracies in the training data, which leads to poor performance on unseen

data. Regularization mitigates this by introducing an additional term to the loss function, which penalizes large weights, thus encouraging the model to learn a smoother function (Ng, 2004).

Regularization can be implemented in various forms, with L1 and L2 being the most common techniques. L1 regularization, also known as Lasso, adds the absolute values of the weights to the loss function, promoting sparsity by effectively setting some weights to zero. This results in simpler models that are less likely to overfit. L2 regularization, or Ridge, adds the squared values of the weights to the loss function, discouraging large weights by penalizing their magnitudes. The regularization term can be expressed as $\lambda R(w)$, where λ is the regularization constant that controls the strength of the penalty, and $R(w)$R(w) is the norm of the weights (either L1 or L2). The appropriate value for λ is crucial; if it is too high, the model may underfit by being overly constrained. The optimal value of λ is often determined through cross-validation (Goodfellow; Bengio; Courville, 2016).

Lasso Regularization (L1) – Equation 14:

$$\text{Cost function} = \text{Loss} + \lambda + \Sigma \, \|w\| \tag{14}$$

Where:
Loss = sum of squared residual;
λ = penalty;
w = slope of the curve.

Ridge Regularization (L2) – Equation 15:

$$\text{Cost function} = \text{Loss} + \lambda + \Sigma \, \|w\|^2 \tag{15}$$

Where:
Loss = sum of squared residual;
λ = penalty;
w = slope of the curve.

The choice of regularization method depends on the specific requirements of the task and the nature of the data. For instance, L1 regularization is particularly useful in feature

selection due to its tendency to produce sparse models, while L2 regularization is beneficial for controlling the overall complexity of the model by penalizing large weights (Table 25).

Table 25 – Comparison of L1 Regularization and L2 Regularization

L1 Regularization	L2 Regularization
Penalty is the absolute value of coefficients	Penalty is the square of the coefficients
Estimate median of the data	Estimate median of the data
Shrinks coefficients to zero	Shrinks coefficients equally
Can be used for dimension reduction and feature selection	Useful when we have collinear features

Source: elaborated by the author.

Implementing regularization requires careful tuning of hyperparameters. The regularization constant λ must be selected appropriately to balance the trade-off between bias and variance. A higher λ value increases bias, leading to underfitting, while a lower λ value decreases bias but increases variance, potentially causing overfitting. Cross-validation is commonly used to find the optimal λ value by evaluating the model's performance on different subsets of the training data (Ng, 2004).

2.2.4.5 Dropout

Dropout is a widely used regularization technique designed to mitigate overfitting in neural networks. It operates by randomly setting a fraction of the neurons to zero during each forward pass through the network. The probability of dropping a neuron is controlled by a hyperparameter, which leads to different subsets of neurons being inactive for each forward pass. This stochastic behavior can be viewed as creating a different neural network structure for each pass, effectively forming an ensemble of models within a single neural network. By preventing any single neuron from becoming overly influential, dropout encourages the network to develop more robust and generalizable features (Srivastava *et al.*, 2014).

This approach significantly reduces the likelihood of co-adaptation among neurons, where certain neurons might become overly dependent on the presence of others. By forcing the network to use different neurons to classify the same object during different forward passes, dropout enhances the network's ability to handle noise and variability in the data. Srivastava *et al.* (2014) provided empirical evidence demonstrating that dropout substantially reduces overfitting and improves performance across various benchmarks. The technique is particularly beneficial in deep neural networks, where the risk of overfitting is higher due to the large number of parameters involved (Figure 31).

Figure 31 – Dropout

(a) Standard Neural Net   (b) After applying dropout.
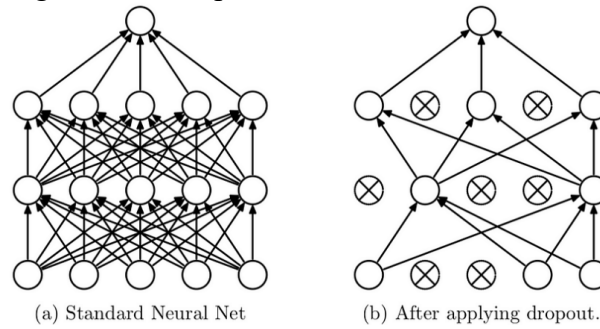
Source: Srivastava *et al.* (2014).

2.2.4.6 Batch normalization

The concept of batch normalization was introduced by researchers from Google in 2015. Batch normalization is a technique aimed at normalizing the activations of hidden layers by ensuring that these activations have a mean of zero and a unit variance during the forward pass through the network. For a mini-batch of m data samples, the empirical mean and variance are calculated independently for each dimension across the mini-batch. The new activation value for each node is then computed by subtracting the meaning of the batch values and dividing by the square root of the variance. This normalized value is subsequently scaled by a parameter γ and shifted by a parameter β, both of which are learnable parameters. These parameters allow the network to set new mean and variance for the activations, adapting to the data distribution (Ioffe; Szegedy, 2015).

Batch normalization has several significant effects on the training of neural networks. It effectively normalizes the inputs to the next layer, which helps in stabilizing the learning process and adds a regularization effect to the model. This stabilization allows for the use of higher learning rates, which accelerates the convergence of the network. However, when using batch normalization, the dropout technique should either be avoided or used with reduced strength to achieve optimal results. The regularization effect of batch normalization introduces some noise into the network, which can help to prevent overfitting. This noise acts similarly to the noise introduced by dropout, thus reducing the need for additional regularization techniques (Ioffe; Szegedy, 2015).

2.2.4.7 Data Augmentation

One of the most effective methods to improve the performance of a machine learning model is by training it on more data (Perez; Wang, 2017). However, acquiring

additional data is often impractical due to constraints such as cost, human effort, or other limitations. Data augmentation addresses this issue by generating new data samples from the existing ones, ensuring that the classifier remains invariant to various transformations and occlusions. It is crucial to ensure that these augmentation techniques do not alter the correct class of the data. In the context of image data, common augmentation methods include cropping, scaling, flipping, rotations, adding noise, and adjusting brightness and contrast. Figure 32 provides examples of data augmentation, while Table 26 displays the most prevalent data augmentation strategies.

Figure 32 – Data augmentation examples
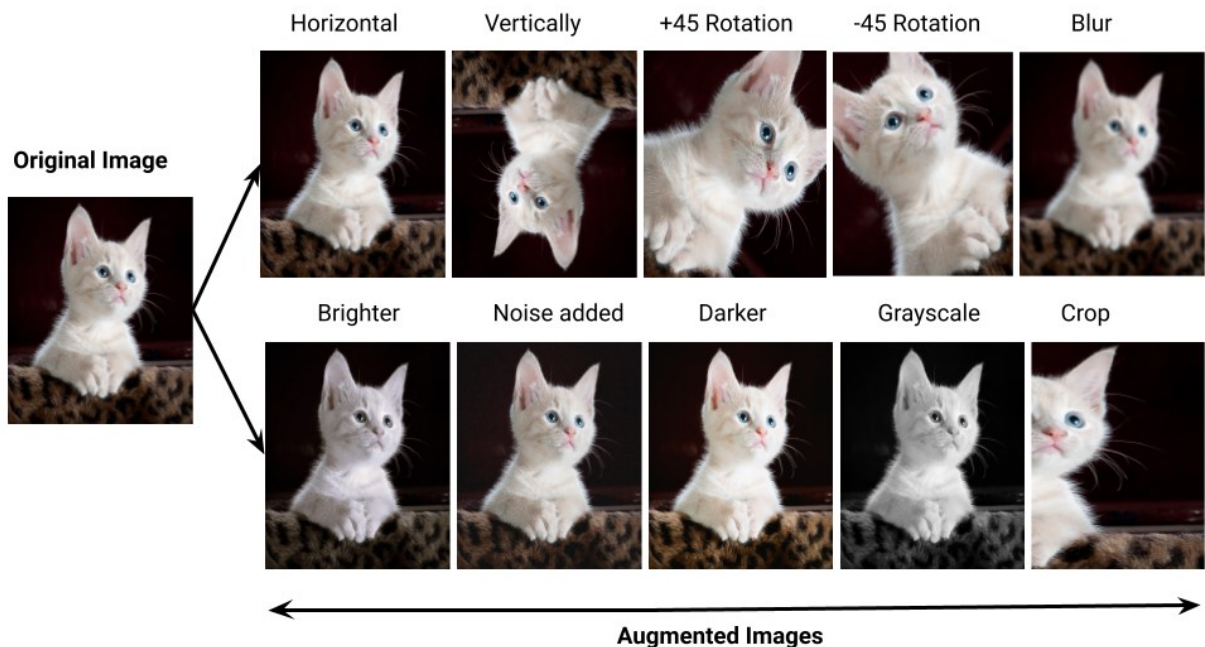
Source: UBIAI (2023).

Table 26 – Data augmentation techniques for image classification

Technique	Description	Impact
Cropping	Randomly cropping parts of the image	Improves robustness to localization variations
Flipping	Horizontal and vertical flips	Enhances orientation invariance
Rotation	Rotating images by random degrees	Increases robustness to object orientation
Brightness and Contrast Adjustment	Altering brightness and contrast	Makes model invariant to lighting conditions
Hue and Saturation Adjustment	Changing hue and saturation levels	Improves generalization to different environments
Gaussian Noise	Adding random noise	Enhances robustness to real-world imperfections
Scaling	Randomly scaling images	Increases invariance to object size
Aspect Ratio Variation	Adjusting the aspect ratio	Helps with shape distortions
Cutout	Masking out regions of the image	Forces focus on salient features
Mixup	Combining pairs of images and labels	Improves robustness and calibration

Source: elaborated by the author.

Data augmentation increases the size of the training dataset artificially by applying such transformations to the existing data. This approach helps the model generalize better by exposing it to a more diverse set of examples, thus improving its robustness and performance on unseen data. By learning from varied instances of the same data, the model becomes more resilient to variations in the data input. According to (Krizhevsky; Sutskever; Hinton, 2012), extensive use of data augmentation is essential for achieving state-of-the-art performance in many computer vision tasks. Their research demonstrated that data augmentation played a significant role in the success of deep convolutional neural networks (CNNs) in competitions such as ImageNet.

2.2.5 Transfer Learning and CNN Architectures for Image Classification

Transfer learning is a method in machine learning that involves using a pre-trained model from one task as the initial model for another task. This strategy utilizes the knowledge acquired from the initial work to enhance the performance and efficiency of the model on the subsequent task, particularly when the latter task has a limitation of data. Transfer learning is very efficient in deep learning, specifically in models such as Convolutional Neural Networks (CNNs), which necessitate substantial amounts of data and processing resources for training from the beginning. Transfer learning, through the process of fine-tuning a pre-trained model, enables quicker convergence and enhanced performance on new tasks with a smaller amount of data (Pan; Yang, 2010).

Transfer learning in practice involves freezing the initial layers of the pre-trained model, which capture general information, and subsequently retraining the final layers, which are more specialized to the task, using the new dataset. This technology is widely employed in the fields of computer vision and natural language processing. For instance, models that have been trained on extensive datasets like ImageNet can be customized for tasks such as medical picture classification or object recognition. This customization process results in a substantial reduction in training time and an enhancement in accuracy (Krizhevsky; Sutskever; Hinton, 2012). Figure 33 illustrates the disparity between errors rates of ImageNet classification models generated in previous challenges and the error rate that would be expected from a human being. The ResNet architecture beat human performance by 5.1% (Russakovsky *et al.*, 2015), making it the first to do so.

Figure 33 – State of the art ImageNet classification models
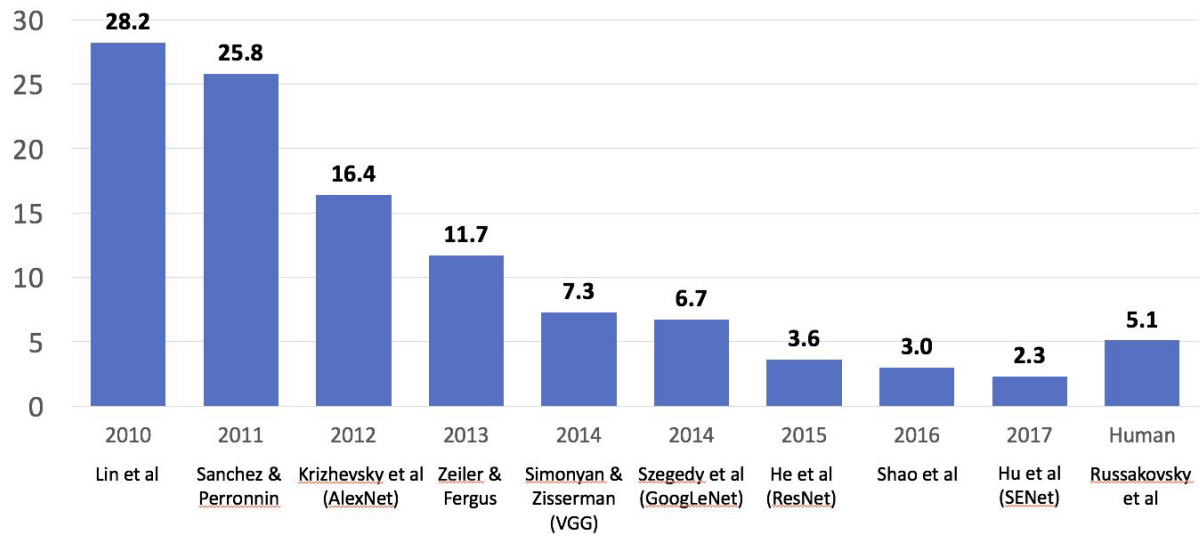
Source: Li, Gao and Li (2022).

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual image recognition competition that leverages a substantial image database, known as ImageNet, to train and evaluate deep learning models. The dataset utilized in the ILSVRC comprises approximately 1.2 million images for training, 50,000 images for validation, and 100,000 images for testing. This large-scale dataset provides a comprehensive benchmark for assessing the performance of various deep learning architectures in handling complex visual recognition tasks. The researchers (Canziani; Culurciello; Paszke, 2017) displayed a graph in Figure 34 that illustrates the performance comparison of different architectures designed for the ILSVRC.

Figure 34 – Top1 vs. CNN Architectures
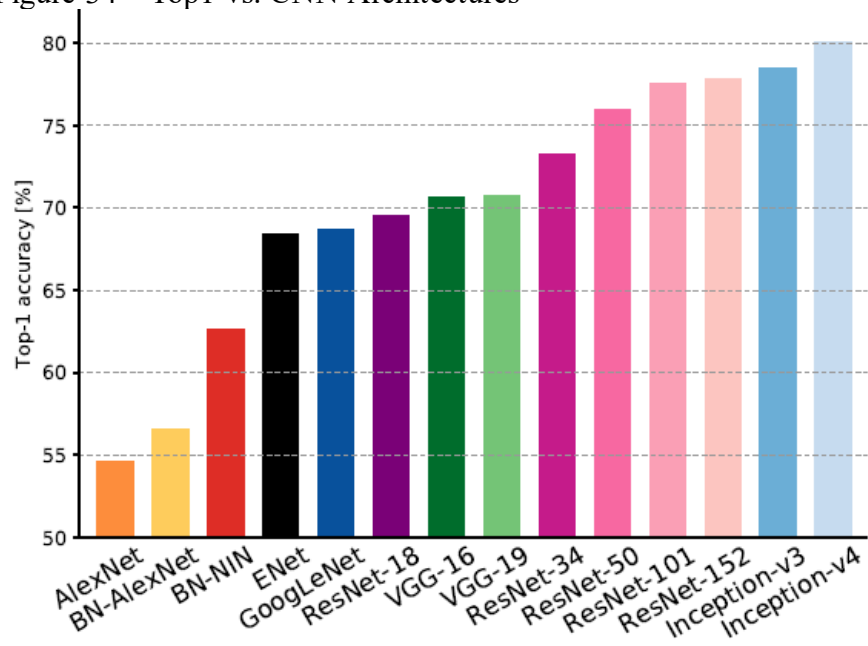
Source: Canziani, Culurciello and Paszke (2017).

2.2.5.1 LeNet

LeNet, created by Yann LeCun and his colleagues during the late 1980s and early 1990s, is a revolutionary architecture in convolutional neural networks (CNNs). Its primary purpose was to be used for problems involving the recognition of digits written by hand, specifically for datasets like MNIST. The LeNet design comprises two sets of convolutional and subsampling layers, which are then followed by fully linked layers. This architecture facilitated the effective and precise handling of image data by collecting spatial hierarchies. The main innovation of LeNet was the inclusion of convolutional layers, which significantly reduced the number of parameters compared to fully connected networks. This allowed for more scalable and efficient learning (LeCun; Bottou; Bengio, *et al.*, 1998). Although considered simple by today's standards, LeNet established fundamental ideas for contemporary deep learning architectures (Figure 35).

Figure 35 – LeNet architecture
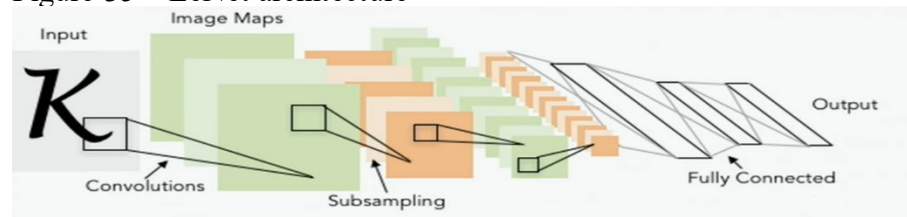
Source: Li, Gao and Li (2022).

2.2.5.2 AlexNet

AlexNet, introduced by Krizhevsky, Sutskever and Hinton (2012), marked a significant breakthrough in computer vision and deep learning. Winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) that year, AlexNet demonstrated the power of deep convolutional networks for large-scale image classification (Figure 36).

Figure 36 – AlexNet architecture
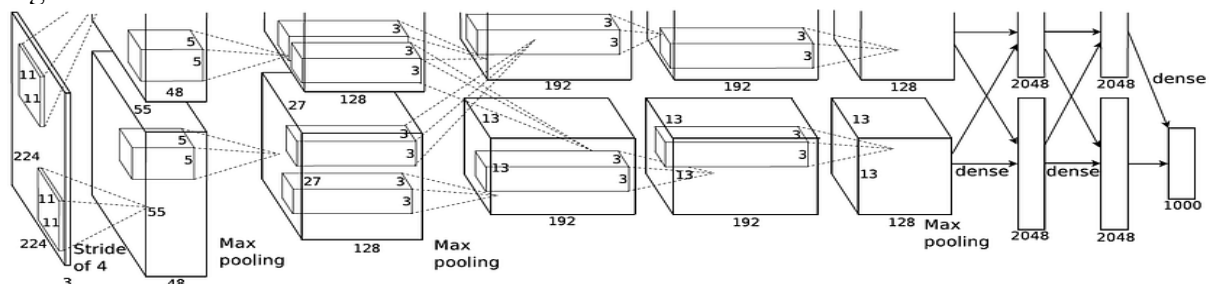
Source: Li, Gao and Li (2022).

The architecture consists of five convolutional layers followed by three fully connected layers, with ReLU activations applied after each layer to introduce non-linearity. Notably, AlexNet utilized dropout layers to mitigate overfitting and employed data augmentation techniques to enhance the robustness of the model. The introduction of GPU acceleration for training was another key innovation, allowing AlexNet to handle the extensive computations required by deep networks (Krizhevsky; Sutskever; Hinton, 2012).

2.2.5.3 VGG

The VGG architecture, developed by the Visual Geometry Group at Oxford, was introduced in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" by Simonyan and Zisserman (2014). VGGNet is characterized by its simplicity and uniform architecture, utilizing small (3x3) convolution filters throughout the network, stacked in increasing depth. The models VGG16 (Figure 37) and VGG19 refer to the number of weight layers, 16 and 19, respectively. The deep and homogeneous architecture of VGG allows for capturing complex features at different levels of abstraction, leading to impressive performance on various image recognition tasks. However, this depth comes at the cost of increased computational complexity and memory usage, posing challenges for deployment in resource-constrained environments (Simonyan; Zisserman, 2014).

Figure 37 – VGG-16 architecture
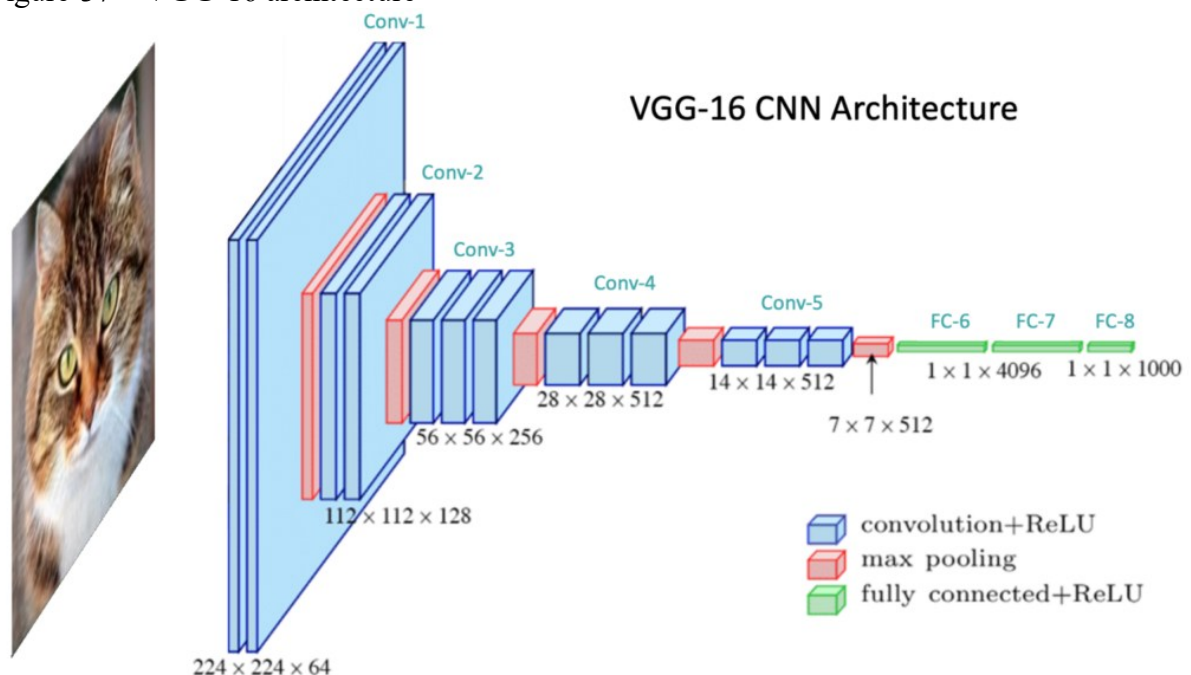
Source: Patel (2023).

2.2.5.4 GoogleNet / Inception

GoogleNet, also known as Inception, was introduced by Szegedy *et al.* (2015) and won the ILSVRC in 2014. The primary innovation of the Inception architecture is the Inception module, which allows for multiple convolutional operations to be performed in parallel within the same layer (Figure 38). This approach enables the network to capture multi-scale features and reduces computational costs compared to a standard deep network with the same number of parameters. The Inception v1 model, also known as GoogleNet, contains 22 layers, and subsequent versions (Inception v2, v3, and v4) introduced additional refinements such as batch normalization and factorized convolutions to further improve performance and efficiency (Szegedy *et al.*, 2015).

Figure 38 – Inception-v3 model's architecture
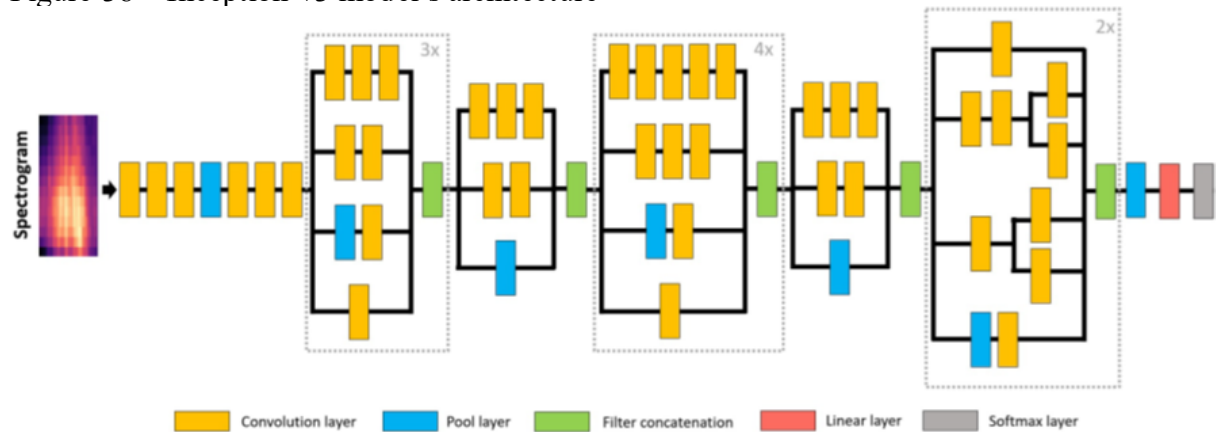
Source: Nogueira *et al.* (2022).

2.2.5.5 ResNet

ResNet, or Residual Networks, introduced by He *et al.* (2016), revolutionized deep learning by addressing the degradation problem observed in very deep networks. ResNet employs residual learning through skip connections, which directly pass input information to deeper layers, bypassing intermediate layers. This innovation allows for the training of extremely deep networks with hundreds or even thousands of layers without suffering from vanishing gradients. The ResNet family (Figure 39) includes various models such as ResNet-34, ResNet-50 (Figure 39), ResNet-101, and ResNet-152, which differ in the number of layers. ResNet models have achieved state-of-the-art performance on many benchmarks, significantly improving image recognition accuracy (He *et al.*, 2016).

Figure 39 – ResNet family

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112			7×7, 64, stride 2		
				3×3 max pool, stride 2		
conv2_x	56×56	$\begin{bmatrix} 3\times3,\ 64 \\ 3\times3,\ 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3,\ 64 \\ 3\times3,\ 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1,\ 64 \\ 3\times3,\ 64 \\ 1\times1,\ 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1,\ 64 \\ 3\times3,\ 64 \\ 1\times1,\ 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1,\ 64 \\ 3\times3,\ 64 \\ 1\times1,\ 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3,\ 128 \\ 3\times3,\ 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3,\ 128 \\ 3\times3,\ 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1,\ 128 \\ 3\times3,\ 128 \\ 1\times1,\ 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1,\ 128 \\ 3\times3,\ 128 \\ 1\times1,\ 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1,\ 128 \\ 3\times3,\ 128 \\ 1\times1,\ 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3,\ 256 \\ 3\times3,\ 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3,\ 256 \\ 3\times3,\ 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1,\ 256 \\ 3\times3,\ 256 \\ 1\times1,\ 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1,\ 256 \\ 3\times3,\ 256 \\ 1\times1,\ 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1,\ 256 \\ 3\times3,\ 256 \\ 1\times1,\ 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3,\ 512 \\ 3\times3,\ 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3,\ 512 \\ 3\times3,\ 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1,\ 512 \\ 3\times3,\ 512 \\ 1\times1,\ 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1,\ 512 \\ 3\times3,\ 512 \\ 1\times1,\ 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1,\ 512 \\ 3\times3,\ 512 \\ 1\times1,\ 2048 \end{bmatrix}\times3$
	1×1			average pool, 1000-d fc, softmax		
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Source: He *et al.* (2016).

Figure 40 – ResNet-50 architecture, (A) CNN structure, and (B) and (C) residual blocks, based on the GitHub code from Keras-team

Source: Karim (2019).

2.2.5.6 DenseNet

DenseNet, introduced by Huang *et al.* (2017), further builds on the concept of skip connections by connecting each layer to every other layer in a feed-forward fashion, Figure 41. This dense connectivity pattern ensures maximum information flow between layers, which leads to improved feature reuse and gradient propagation. DenseNet requires fewer parameters than traditional CNNs because it eliminates the need to relearn redundant feature maps. The architecture has demonstrated superior performance on various benchmarks while being more parameter efficient. The DenseNet family (Figure 42) includes DenseNet-121, DenseNet-169, DenseNet-201, and DenseNet-264, indicating the number of layers in each model (Huang *et al.*, 2017).

Figure 41 – DenseNet-169 architecture

Source: Huang *et al.* (2017).

Figure 42 – DenseNet Family for ImageNet

Layers	Output Size	DenseNet-121		DenseNet-169		DenseNet-201		DenseNet-264	
Convolution	112×112	7×7 conv, stride 2							
Pooling	56×56	3×3 max pool, stride 2							
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 6$
Transition Layer (1)	56×56	1×1 conv							
	28×28	2×2 average pool, stride 2							
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 12$
Transition Layer (2)	28×28	1×1 conv							
	14×14	2×2 average pool, stride 2							
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 64$
Transition Layer (3)	14×14	1×1 conv							
	7×7	2×2 average pool, stride 2							
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix}$	$\times 48$
Classification Layer	1×1	7×7 global average pool							
		1000D fully-connected, softmax							

Source: Huang *et al.* (2017).

2.2.5.7 Xception

Xception, proposed by Chollet (2017), stands for "Extreme Inception" and is an extension of the Inception architecture. Xception replaces the standard Inception modules with depthwise separable convolutions, which decompose a convolution into a depthwise convolution followed by a pointwise convolution, Figure 43. This modification results in a more efficient model with fewer parameters while maintaining high performance. Xception has shown strong results on various image classification benchmarks and has been adopted in numerous applications due to its balance of accuracy and computational efficiency (Chollet, 2017).

Figure 43 – Xception architecture
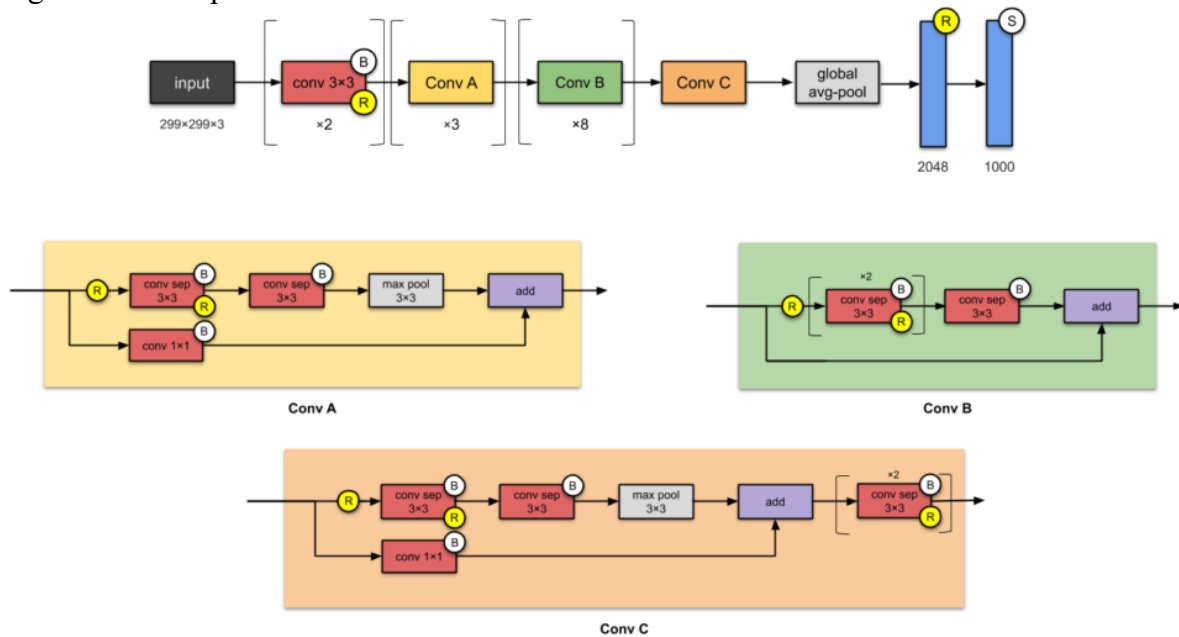
Source: AI Singapore (2019).

2.2.5.8 ResNeXt-50

ResNeXt, introduced by Xie *et al.* (2017), builds on the principles of ResNet by incorporating grouped convolutions to enhance learning capacity without significantly increasing computational complexity. The ResNeXt-50 model consists of 50 layers and uses a modular, cardinality-based approach, where increasing cardinality (the number of parallel paths in each block) improves model performance, Figure 44. This architecture achieves higher accuracy than ResNet with similar computational costs, making it an attractive option for various computer vision tasks (Xie *et al.*, 2017).

Figure 44 – ResNeXt architecture
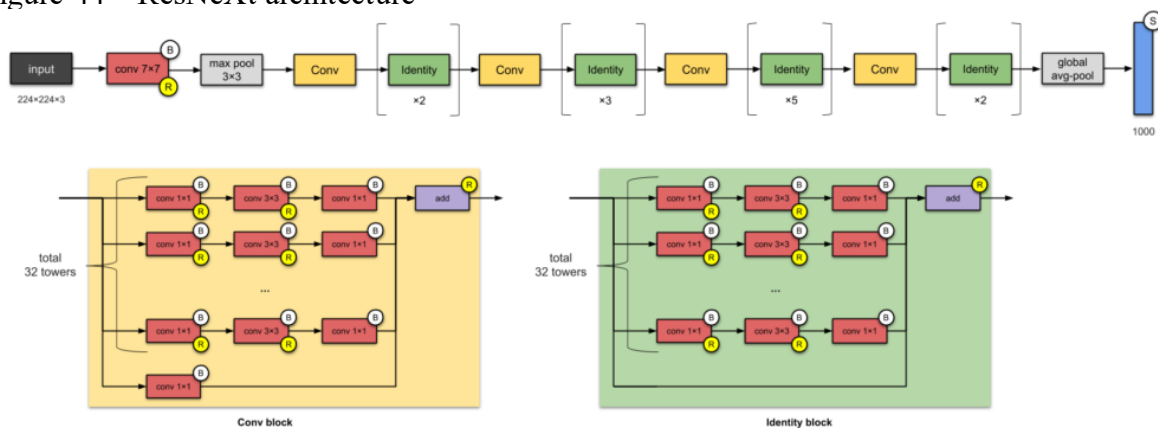
Source: AI Singapore (2019).

2.2.6 Performance and Efficiency of CNN Architectures

GFLOPS is a measure of the computational complexity of a neural network model. It indicates the number of billion floating-point operations required to perform inference on a single input sample. Understanding GFLOPS is crucial for assessing the performance and efficiency of Convolutional Neural Network (CNN) architectures.

Models with more GFLOPS generally necessitate greater processing resources, resulting in longer inference times. This is because each operation requires a certain amount of time, and as the number of operations increases, the amount of time needed to process each input also increases. In contrast, models with smaller GFLOPS can handle inputs more rapidly, rendering them appropriate for real-time applications that require prompt answers. Complex models with more GFLOPS tend to contain more parameters and layers, which allows them to capture sophisticated patterns in the data with greater accuracy and capacity. This can result in improved precision, particularly in intricate tasks that necessitate extensive feature extraction. Conversely, less complicated models with fewer GFLOPS may have restricted capability, leading to reduced accuracy on complex tasks, but they may still perform satisfactorily for simpler tasks (He *et al.*, 2016).

When it comes to efficiency, an elevated amount of GFLOPS results in increased energy consumption. This is an important factor to consider when deploying models in environments that have restricted power resources, such as mobile devices or edge computing scenarios. On the other hand, models that require fewer calculations use less energy, which makes them better suited for areas with limited power resources. In addition, models with large GFLOPS may necessitate the use of specialist hardware, such as GPUs or TPUs, to achieve optimal performance, hence raising the expenses and intricacy involved in deploying them. Models with lower GFLOPS can be executed on regular CPUs, which reduces the requirement for specialist hardware and simplifies deployment while still being cost-effective (Huang *et al.*, 2017; Szegedy *et al.*, 2015).

Canziani, Culurciello and Paszke (2017) presented a plot comparing the performance of architectures developed for the ILSVRC against the number of Floating Point Operations per Second (FLOPs), as illustrated in Figure 45. This graphical representation provides a clear visualization of the trade-offs between computational complexity and model accuracy, highlighting the efficiency and effectiveness of different deep learning models used in the challenge.

Figure 45 – Top-1 Accuracy vs. Operations
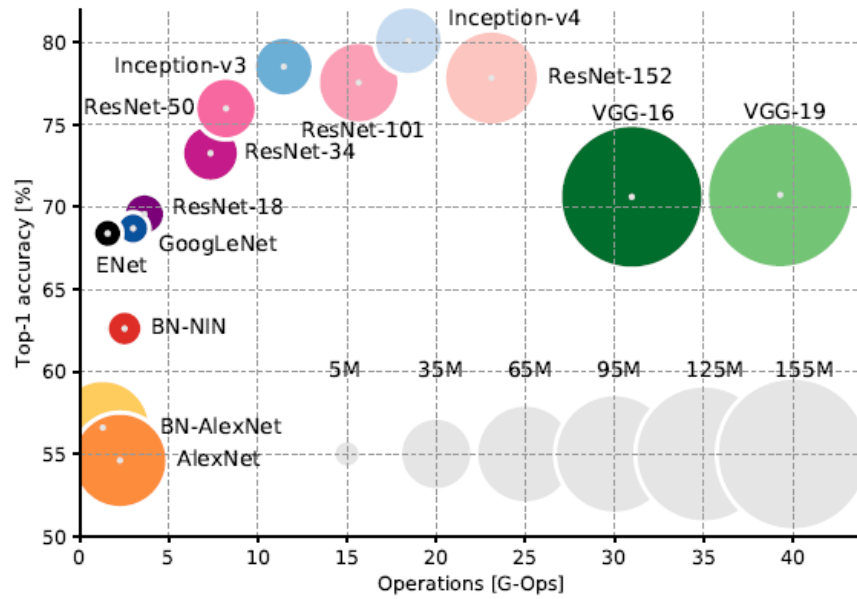
Source: Canziani, Culurciello and Paszke (2017).

Table 27 presents the GFLOPS values for architectures developed for classification, object detection, and segmentation tasks. It is evident that Object Detection and Segmentation tasks require significantly higher computational power, as indicated by the substantially greater GFLOPS values. These architectures can extract different types of information from the data, which accounts for their increased computational demands. The comparison underscores the complexity and computational intensity involved in performing object detection and segmentation, as opposed to standard classification tasks.

Table 27 – Top-1 Accuracy vs. Operations (GFLOPS) for Classification, Object Detection, and Segmentation Architectures

(continue)

Architecture	GFLOPS	Top-1 Accuracy (%)	Tasks Type	Reference
LeNet-5	0.0023	-	Classification	LeCun; Bottou; Bengio, *et al.* (1998)
AlexNet	1.45	56.5	Classification	Krizhevsky, Sutskever and Hinton (2012)
VGG16	15.3	71.5	Classification	Simonyan and Zisserman (2014)
VGG19	19.6	71.5	Classification	Simonyan and Zisserman (2014)
Inception v3	5.72	78.8	Classification	Szegedy *et al.* (2015)
Inception v4	12.27	80.1	Classification	Szegedy *et al.* (2017)
ResNet-34	3.6	73.3	Classification	He *et al.* (2016).
ResNet-50	3.8	76.4	Classification	He *et al.* (2016).
ResNet-101	7.6	77.0	Classification	He *et al.* (2016).
ResNet-152	11.3	78.3	Classification	He *et al.* (2016).
DenseNet-121	2.9	74.9	Classification	Huang *et al.* (2017).
DenseNet-169	3.4	76.2	Classification	Huang *et al.* (2017).
DenseNet-201	4.4	77.3	Classification	Huang *et al.* (2017).
DenseNet-264	5.7	77.9	Classification	Huang *et al.* (2017).
Xception	8.4	79.0	Classification	Chollet (2017).
ResNeXt-50	4.2	77.8	Classification	Xie *et al.* (2017)
YOLOv3	65.86	57.9	Object Detection	Redmon and Farhadi (2018)

Table 27 – Top-1 Accuracy vs. Operations (GFLOPS) for Classification, Object Detection, and Segmentation Architectures

(conclusion)

Architecture	GFLOPS	Top-1 Accuracy (%)	Tasks Type	Reference
YOLOv4	71.65	62.7	Object Detection	Bochkovskiy, Wang and Liao (2020)
Faster R-CNN	180	73.2	Object Detection	Ren *et al.* (2017)
Mask R-CNN	260	75.1	Object Detection	He *et al.* (2017)
SSD	31.39	74.3	Object Detection	Liu *et al.* (2016)
RetinaNet	199	72.4	Object Detection	Lin *et al.* (2020)
U-Net	30.98	67.9	Segmentation	Ronneberger, Fischer and Brox (2015)
DeepLabV3	33.2	70.1	Segmentation	Chen *et al.* (2018)
DeepLabV3+	45.02	72.5	Segmentation	Chen *et al.* (2018)
PSPNet	177	71.3	Segmentation	Zhao *et al.* (2017)

Source: elaborated by the author.

Using the data from Table 27, a comparative plot was generated to visualize the relationship between Top-1 Accuracy and the number of operations (GFLOPS) for various architectures across classification, object detection, and segmentation tasks, Figure 46. This comparative analysis provides critical insights into the trade-offs between computational complexity and model performance across different types of neural network architectures.

Figure 46 – Top-1 Accuracy vs. Operations (GFLOPS) for Classification, Object Detection, and Segmentation Architectures
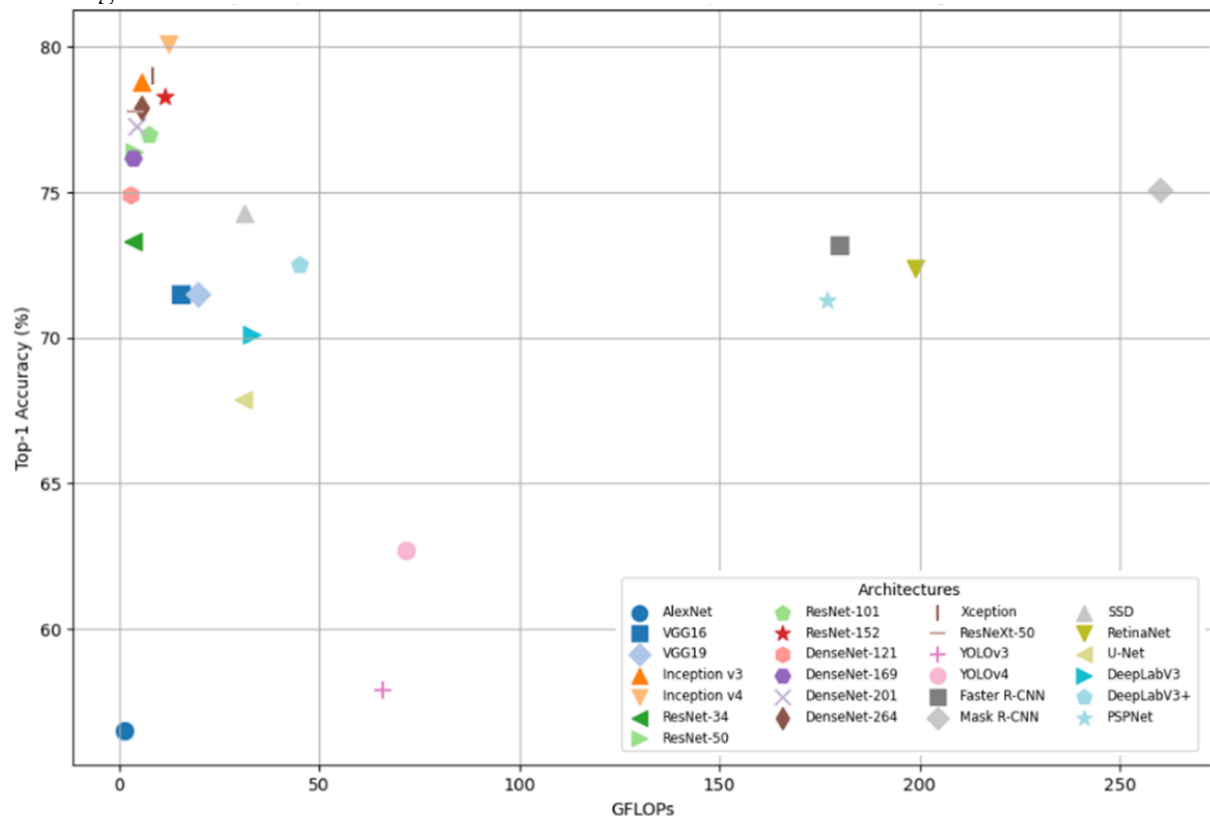
Source: elaborated by the author.

By plotting Top-1 Accuracy against GFLOPS, this comparative analysis highlights the advancements in neural network design and the ongoing efforts to balance computational efficiency with performance. Such insights are invaluable for researchers and practitioners in selecting appropriate models for their specific applications, ensuring optimal performance while managing computational resources.

2.3 State of the Art – Convolutional Neural Networks Applied to Pavement Management

Traditional Pavement Management Systems (PMS) face limitations due to costly data collection, reliance on specialized equipment, and budgetary constraints that lead to outdated assessments. Visual inspections remain subjective, and conventional distress indices often fail to capture localized issues, compromising evaluation accuracy. Predictive models struggle with complex interactions between traffic loads, climate, and materials, resulting in unreliable forecasts. Additionally, high initial costs and resistance to change hinder the adoption of new technologies, despite their long-term benefits (Tamagusko; Gomes Correia; Ferreira, 2024).

To address these challenges, machine learning techniques have been applied to classify pavement distress types and severity levels using image, sensor, or survey data, primarily through object detection or semantic segmentation. These algorithms enhance defect detection accuracy, allowing skilled labor to focus on safer, less repetitive tasks. For example, specialists conducting field data collection or analyzing pavement videos can work more efficiently and with reduced risk through advanced computer vision techniques (Tamagusko; Gomes Correia; Ferreira, 2024).

According to Figure 47, traditional machine learning algorithms, including support vector machines (SVM), artificial neural networks (ANN), convolutional neural networks (CNN), multi-layer perceptron's (MLP), logistic regression, and random forests, account for 50.7% of the total applications in pavement distress detection. In comparison, deep learning techniques, such as semantic segmentation and object detection algorithms, represent 42.2%, reinforcing the predominant role of AI-based methodologies in this domain (Zhang *et al.*, 2024).

Figure 47 – Percentage distributions with respect to pavement distress detection algorithms and targeted pavement distresses (2010–2023). (a) Pavement distress detection algorithms. (b) Targeted pavement distresses.
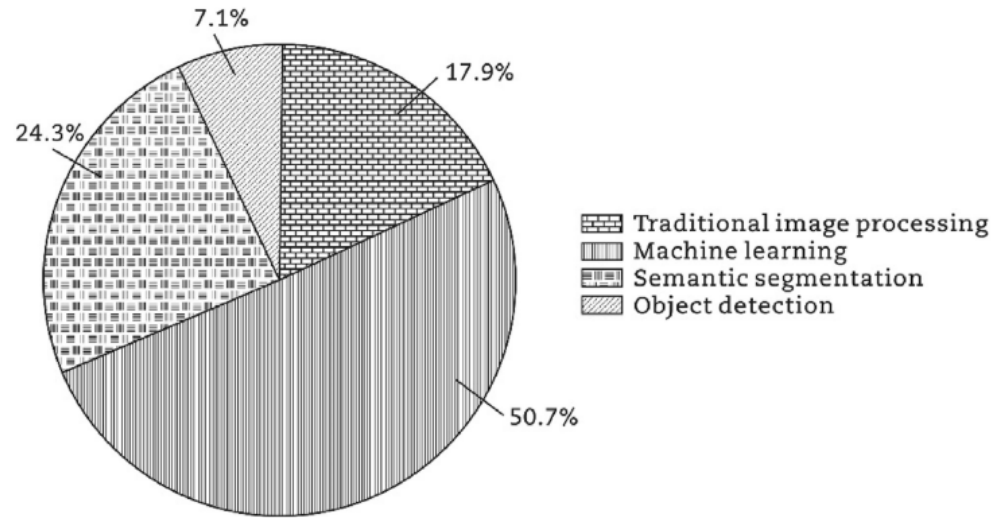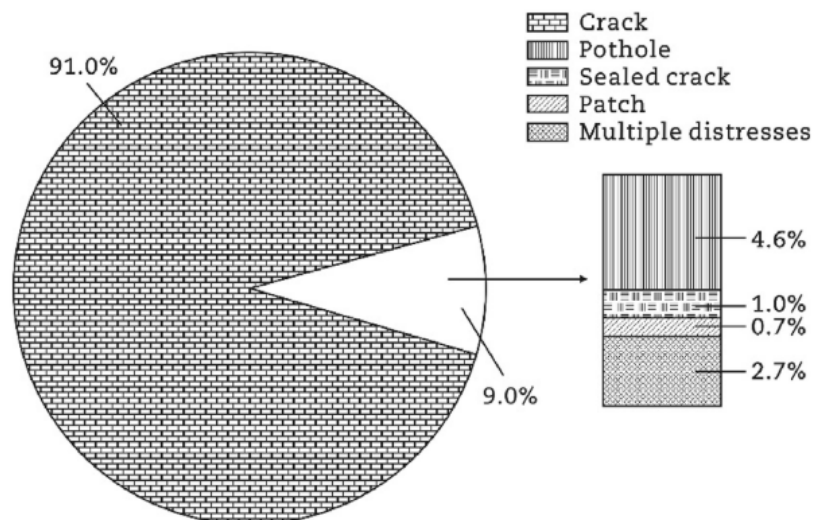
Source: Zhang *et al.* (2024).

Further analysis of detection algorithms across different pavement distress types reveals that machine learning techniques are primarily employed in crack detection and multi-distress recognition from 2010 to 2023. Deep learning, on the other hand, is more commonly utilized for the automated identification of potholes, patches, and sealed cracks, while traditional image processing techniques exhibit the lowest application rates. The increasing dominance of AI-driven algorithms underscores their growing effectiveness in pavement distress detection (Zhang *et al.*, 2024). Figure 48 shows the percentages of different detection algorithms for varying detection targets.

Figure 48 – Partitioning of pavement distress detection algorithms for different detection targets
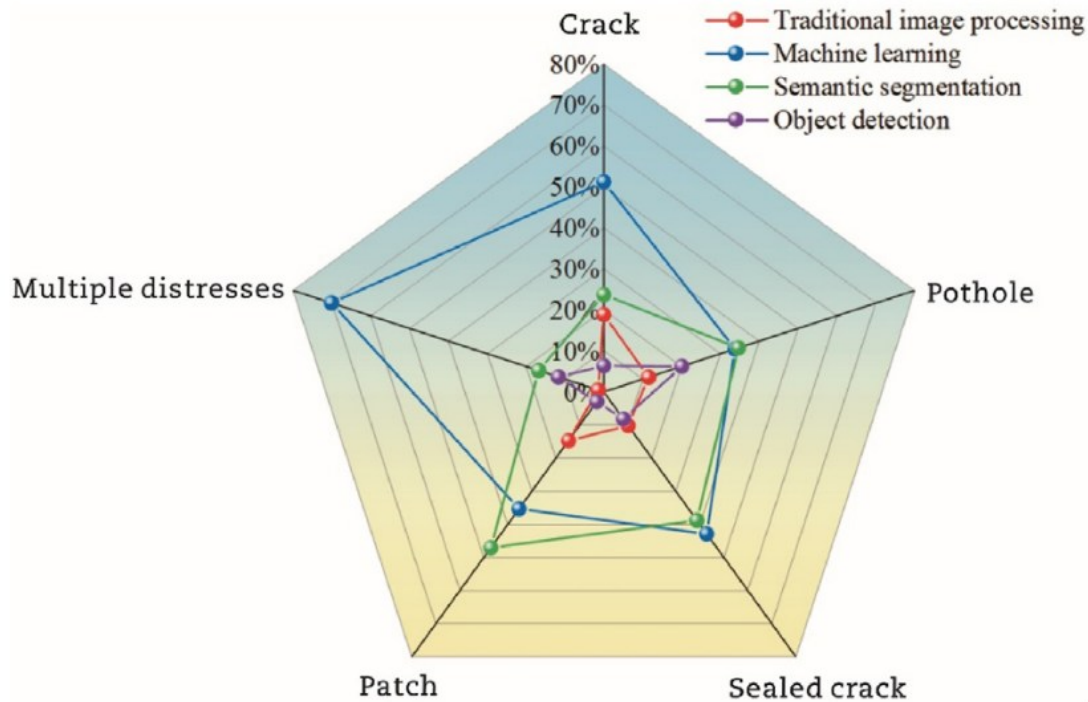
Source: Zhang *et al.* (2024).

The study also highlights that CNNs, SVMs, and traditional ANNs are the most frequently used AI methods for pavement distress recognition, each contributing over 15% of total applications. In contrast, transformer models constitute less than 1% of detection algorithms, reflecting their limited adoption in this field. Moreover, object detection approaches, including YOLO, SSD, and Faster R-CNN, are scarcely applied to the identification of sealed cracks and pavement patches, further emphasizing the prevailing reliance on conventional AI-based models in pavement assessment (Zhang *et al.*, 2024).

Machine learning and deep learning have significantly advanced pavement distress identification by improving classification, object detection, and segmentation techniques, particularly in data collection for diagnostic purposes. This subsection presents studies that have contributed to the development of state-of-the-art classification techniques in the field of pavement defect identification.

Zhang *et al.* (2016) proposed a CNN-based method for crack detection using smartphone-acquired pavement images. Their model, trained on 500 high-resolution images, achieved higher classification accuracy than traditional feature-based approaches, confirming the superiority of CNNs in precision and recall compared to Support Vector Machines (SVM) and Boosting techniques. The study demonstrated that deep learning models could significantly reduce misclassification rates, enhancing real-time crack detection efficiency.

Building on this foundation, Paulya *et al.* (2017) explored the performance of deep CNNs for crack detection, revealing that deeper networks (five convolutional layers) outperformed shallower architectures, achieving 91.3% accuracy and 92.0% recall. However, the study also highlighted generalization challenges when tested on images from different locations, indicating the need for domain adaptation techniques to improve transferability. Wang and Hu (2017) further advanced automated crack detection by integrating CNNs with Principal Component Analysis (PCA). Their approach segmented pavement images into grids of varying scales, allowing CNNs to detect cracks efficiently, achieving an F-measure of 94.7%. The study also applied PCA to classify cracks into longitudinal, transverse, and alligator types with accuracy rates of 97.2%, 97.6%, and 90.1%, respectively, confirming the robustness of grid-based segmentation techniques.

A major contribution in 2017 came from Eisenbach *et al.* (2017), who introduced the German Asphalt Pavement Distress (GAPs) dataset, a publicly available dataset for benchmarking deep learning models in pavement distress detection. Their evaluation of state-of-the-art CNN architectures, including ASINVOS and RCD, demonstrated that deep learning significantly outperformed traditional vision-based methods, achieving high accuracy in detecting cracks, potholes, inlaid patches, applied patches, and open joints. This study established a foundation for automated and scalable pavement condition assessment by standardizing evaluation benchmarks.

In 2018, several studies pushed the boundaries of deep learning for pavement assessment. Chen and Jahanshahi (2018) introduced NB-CNN, a fusion framework combining CNNs and a Naïve Bayes approach for microcrack detection in nuclear power plant inspections. Their method achieved a 98.3% hit rate with only 0.1 false positives per frame, significantly surpassing conventional vision-based approaches. The study demonstrated that integrating probabilistic fusion techniques with deep learning models could substantially reduce false positive rates while improving sensitivity.

Hoang, Nguyen and Tran (2018) integrated metaheuristic-optimized edge detection with CNNs, demonstrating that CNN models outperformed traditional edge detection algorithms. Their method is fine-tuned Sobel and Canny edge detection thresholds using Differential Flower Pollination (DFP), leading to a classification accuracy rate of 92.08%, compared to 79.99% for optimized Sobel and 76.69% for optimized Canny. This research underscored the potential of hybrid optimization techniques in refining pavement distress detection models.

Yusof *et al.* (2018) introduced a two-stage deep CNN model for crack detection and classification in asphalt pavement images. The first network detected cracks using image grids of 32×32 and 64×64 pixels, while the second network classified them as transverse or longitudinal cracks. The study found that the 32×32 grid scale provided superior performance, achieving 99.2% accuracy in crack detection and 98% and 97% accuracy in transverse and longitudinal crack classification, respectively. The results validated the effectiveness of multi-scale grid segmentation for enhanced crack classification.

An *et al.* (2018) explored the application of deep CNNs for pothole detection in smartphone-acquired pavement images. Their study tested multiple architectures, including Inception_ResNet_v2, ResNet_v2_152, and MobileNet_v1, using grayscale and color images. The results indicated that grayscale preprocessing significantly improved detection accuracy, with Inception_ResNet_v2 and ResNet_v2_152 achieving 97.5% accuracy and 100% recall. These findings highlighted the advantages of optimized feature extraction in real-time pothole detection.

Li *et al.* (2018) focused on 3D image patches to classify pavement cracks with CNNs. Their study trained four CNN architectures with varying receptive field sizes, with the best-performing model (7×7 receptive field) reaching 96.3% accuracy. The results highlighted the potential of 3D convolutional features in improving the precision of pavement distress classification.

Xia (2018) proposed a deep CNN approach for detecting and classifying road pavement damage using HD camera videos. The study introduced a weakly supervised labeling method, significantly reducing manual annotation efforts while maintaining high detection accuracy. The CNN-based object detection model achieved high precision in identifying alligator cracks, longitudinal cracks, transverse cracks, seam breaks, and potholes. Furthermore, the system quantified damage severity by extracting shape-based parameters, improving prioritization in pavement maintenance operations.

Pereira *et al.* (2018) introduced a CNN model optimized for classifying paved and unpaved roads. Their study utilized a dataset of 16,372 road images encompassing various surface conditions, including wet, muddy, dry, dusty, and shaded environments. The model, inspired by the VGG network, was optimized for binary classification and compared with traditional machine learning methods such as Random Forest and SVM. Experimental results showed that the proposed CNN model achieved an accuracy of 98.2%, with a precision of 98.0%, recall of 98.4%, and an F1-score of 98.2%, outperforming traditional approaches. The findings validated deep learning's effectiveness in automated road condition monitoring.

Further developments in 2019 broadened the applications of deep learning beyond crack detection. Aparna *et al.* (2019) explored thermal imaging-based pothole detection using ResNet architectures, reaching 97.08% accuracy. Their study leveraged temperature differentials between potholes and surrounding pavement to enhance detection performance, demonstrating that thermal imaging can overcome limitations of conventional optical techniques in adverse weather and low-light conditions.

Aslan *et al.* (2019) developed a CNN framework for classifying four types of cracks, achieving 76.2% accuracy. The study applied data augmentation techniques to compensate for limited training samples, highlighting the potential of synthetic data in improving pavement distress classification models.

In 2020, deep learning applications in pavement distress detection advanced through fusion techniques. Zhou and Song (2020) introduced a DCNN-based framework that integrated heterogeneous image data fusion, enhancing classification accuracy with an F-measure of 99.3%. Their approach combined raw intensity and range images captured by a vehicle-mounted laser imaging system, demonstrating that heterogeneous data fusion could improve robustness against low-contrast pavement surfaces and shallow cracks.

Ranjbar, Nejad and Zakeri (2020) combined transfer learning and wavelet transforms to improve crack detection, achieving up to 99% accuracy. Their method utilized pre-trained convolutional neural networks (AlexNet, GoogleNet, ResNet, and DenseNet) adapted for pavement crack classification, demonstrating that transfer learning could enhance model efficiency even with limited datasets.

Srinidhi and Devi (2020) compared Spectral Clustering with CNN-based approaches for pothole detection, with AlexNet achieving 100% accuracy. The study demonstrated that CNN-based models significantly outperformed traditional clustering algorithms, reinforcing the efficiency of deep learning for automated pothole detection.

More recent studies have explored multi-label classification and improved AI-based pavement analysis (Espíndola *et al.*, 2023a) evaluated VGG16, ResNet-34, and ResNet-50 for classifying multiple pavement distress types, with ResNet-50 achieving 97.6% accuracy and an F1-score of 92.8%. Sirhan, Bekhor and Sidess (2024) proposed a multilabel CNN model for simultaneous classification of multiple distress types, achieving an F1 score of 80.52% on a dataset of 42,520 images. These advancements highlight the growing role of deep learning in automating pavement assessment, improving detection accuracy, and enhancing scalability for practical implementation.

3 METHOD

The outcomes of the thesis will be categorized into four distinct sections, which will be aligned with the pre-established objectives and will correspond to papers. The proposed methodology comprises four distinct phases, with the first three stages being subdivided to address tasks, while the final phase involves combining all the stages.

The present doctoral thesis comprises of four papers. The text outlines four distinct areas of focus. Firstly, the identification of pavement surface types including asphalt, concrete, cobblestone, interlocking, and unpaved. Secondly, the classification of multiple defects within an image such as potholes, cracks, bleeding, raveling, and patches. Thirdly, a proposal for a pavement quality index that utilizes information extracted automatically from images. Lastly, the text discusses a pavement diagnostic system for network-level PMS, as illustrated in Figure 49.

Figure 49 – Propose structure of automatic PMS
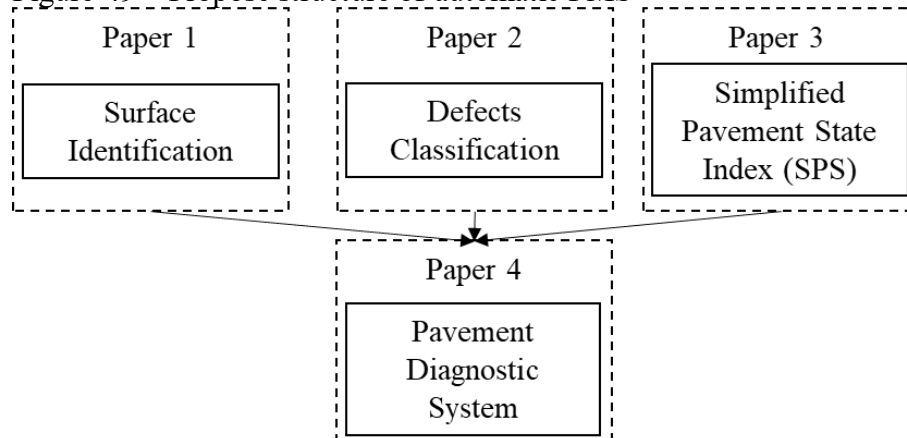
Source: elaborated by the author.

The objective of the group's research is to devise an automated approach that is cost-effective for the purpose of inventorying various road elements, including but not limited to pavement defects, pavement type, horizontal and vertical signages. A module is being created for the inventory system for each study element, utilizing a uniform data source of road videos/images and employing deep learning techniques to develop classification or object detection models for the purpose of inventorying the road elements. The focus of this thesis is to identify the type of pavement and the distresses on asphalt layers, with the aim of integrating this information into a network Pavement Management System (PMS). Additionally, a distinct researcher is currently involved in the development of signage modules.

As previously mentioned, the Automated Pavement Diagnostic System comprises four studies. The first study focuses on the development of a CNN-based model for constructing the Pavement Type Inventory (Figure 50), while the second study presents a model designed for identifying defects and creating the Asphalt Pavement Defects Inventory (Figure 51).

Figure 50 – System Module: Surface Pavement Type
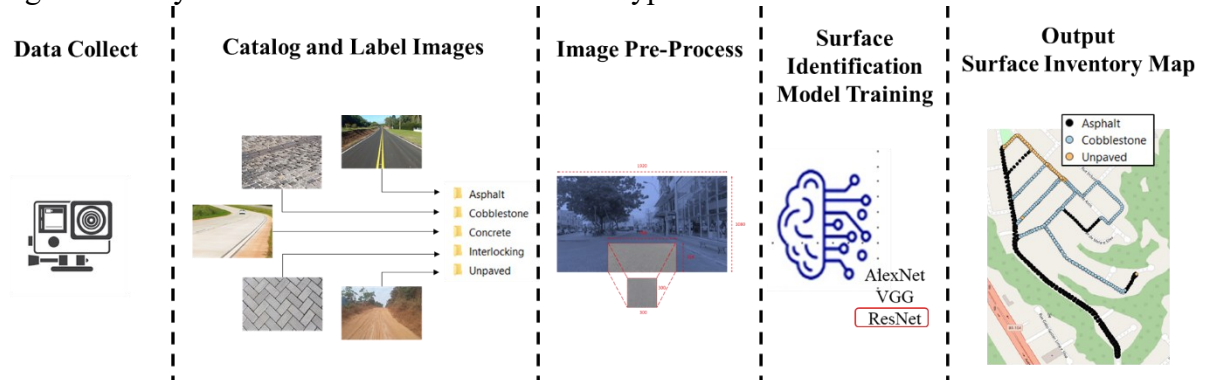
Source: elaborated by the author.

Figure 51 – System Module: Pavement Defects
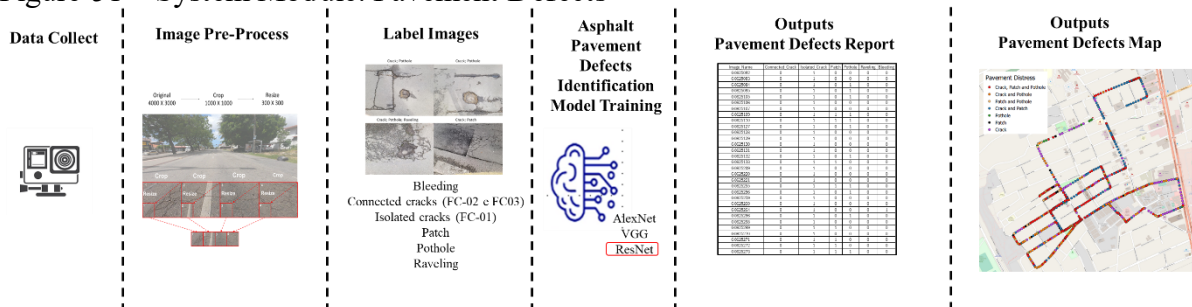
Source: elaborated by the author.

Figure 50 presents the development structure of the System Module: Surface Pavement Type, detailing the input data, model training, and initial module output in the form of a map. Images identified as asphalt pavement are further analyzed by the second system module, as illustrated in Figure 51. Figure 51 outlines the development structure of the System Module: Pavement Defects, which includes input data, labelling, CNN training, and outputs in the form of a report and/or map.

In the third study, a pavement quality index was proposed–the Simplified Pavement State Index (SPS) based on defect information extracted from 2D images. This index was then compared with the IGG DNIT 006/2003-PRO, the widely adopted pavement condition assessment index in Brazil. Figure 52 presents a summary of the elements that define pavement quality based on the IGG and SPS indexes, illustrating the key components used to assess pavement conditions.

Figure 52 – Summary of Elements Defining Pavement Quality Based on IGG and SPS Indexes
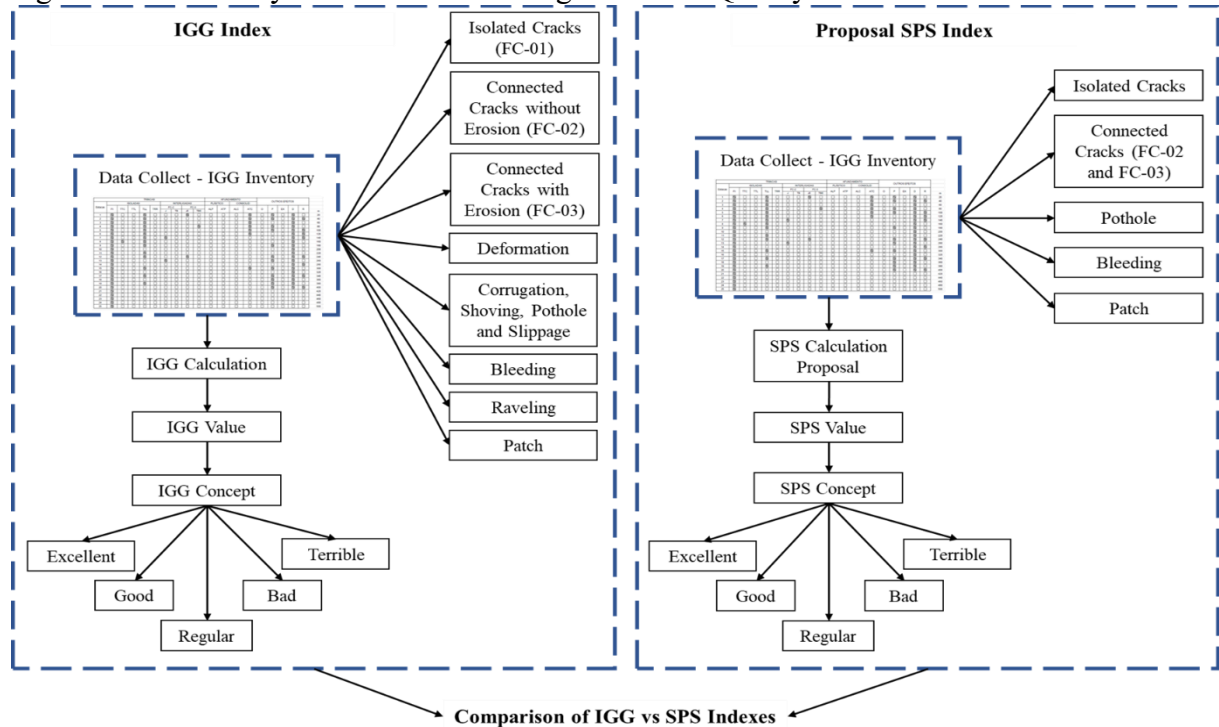
Source: elaborated by the author.

The final system module structure (Automated Pavement Diagnostic System) is presented in Figure 53. Initially, images are captured using GPS-equipped cameras, such as GoPro, Garmin, or mobile phones. Latitude and longitude data are then extracted to facilitate the creation of inventory maps. The automatic pavement surface type classification models are applied. Additionally, for asphalt pavement images, the multi-label distress classification module can be integrated with the inventory system, followed by the calculation of the Simplified Pavement State Index (SPS). The system outputs can be generated in the form of reports or maps.

Figure 53 – Automated Pavement Diagnostic System
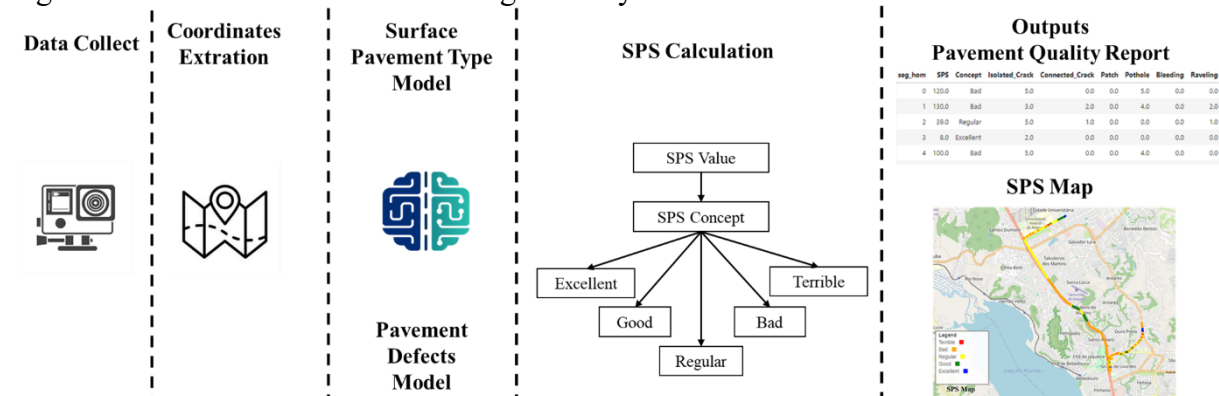
Source: elaborated by the author.

This method chapter details the stages and sequence of actions required for the development of the tree research papers, including data collection, datasets, image preprocessing, model training, evaluation metrics, and the IGG database, which contributed to the proposal of the Simplified Pavement State Index (SPS).

3.1 Datasets and Database

The first three papers utilized different input data sources for research development. The first study focused on developing an automatic pavement type classification model, and its dataset consisted of pavement surface images representing five distinct pavement types: asphalt, concrete, unpaved, interlocking, and cobblestone/stone. The second paper aimed to develop an automatic classification model for pavement defects that can be identified through 2D images. Consequently, the training and testing datasets included images containing bleeding, connected cracks, isolated cracks, patches, potholes, raveling, or no defects.

In contrast, the third paper no longer used images as input data but rather distress inventory information based on the IGG DNIT 006/2003-PRO standard methodology (IPR; DNIT, 2003b). The database was composed of records specifying the presence of pavement distresses rather than visual representations. The defects considered by the IGG methodology include isolated cracks (FC-01), connected cracks without erosion (FC-02), connected cracks with erosion (FC-03), deformation, corrugation, shoving, potholes and slippage, bleeding, raveling, and patches. Figure 54 summarizes the input data used in each paper.

Figure 54 – Types of data used for the development of the thesis papers
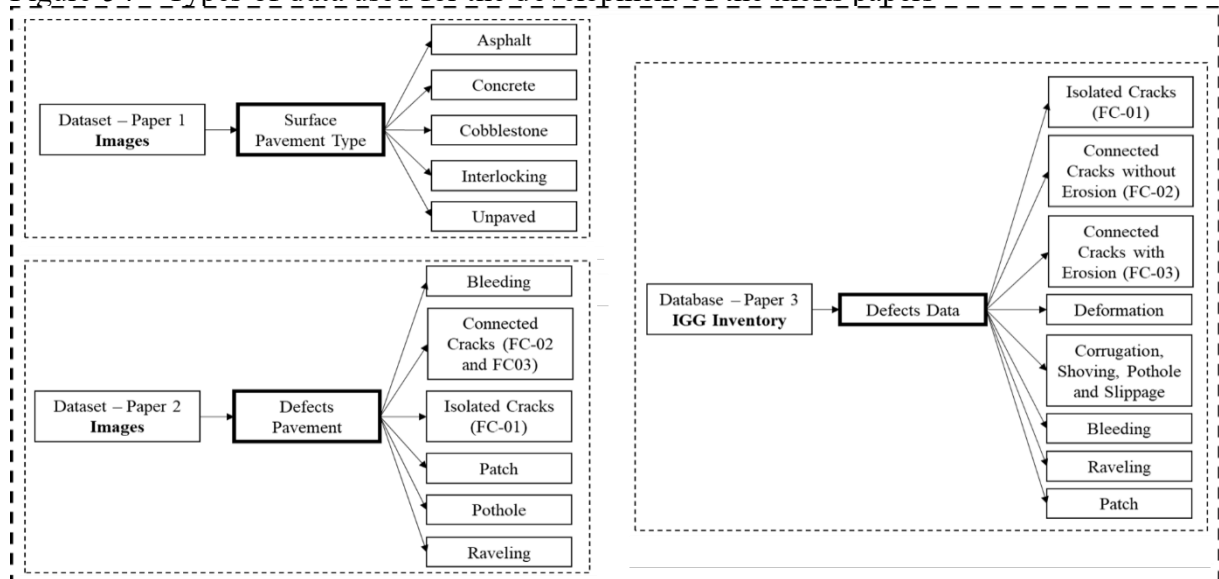
Source: elaborated by the author.

The studies by Espíndola *et al.* (2023a,b) highlight the limitations of training a classification model exclusively on images from a single dataset. Their findings indicate that such an approach often results in inaccurate predictions or, in some cases, fails to generate any meaningful labels. To address this, they emphasize the importance of incorporating diverse datasets, which enhances the generalizability of the classification algorithm, simplifies its implementation, and facilitates its integration into network-level pavement management systems (PMS). This approach improves the model's robustness and reliability in practical applications.

3.1.1 Images Data Collection

The dataset utilized for this research comprises data from 116 roads obtained from five different sources, including four private companies and one internally collected dataset. The sources are designated as Dataset A (own data), B and C (source 2), D (source 3), E (source 4) and F (source 5). These datasets include images captured from federal, state, and urban roads located in the Brazilian states of Alagoas, Ceará, Rio de Janeiro, São Paulo, and Sergipe, as well as roads in London, UK. In total, more than 280,000 images extracted from survey videos were available for analysis.

During the image analysis, variations were observed in the camera positioning within the vehicle (Figure 55), with differences in viewing angles (front-facing, downward-facing, or rear-facing) and image dimensions.

Figure 55 – Camera positioning in the vehicle, dataset A, B, C, D, E and F
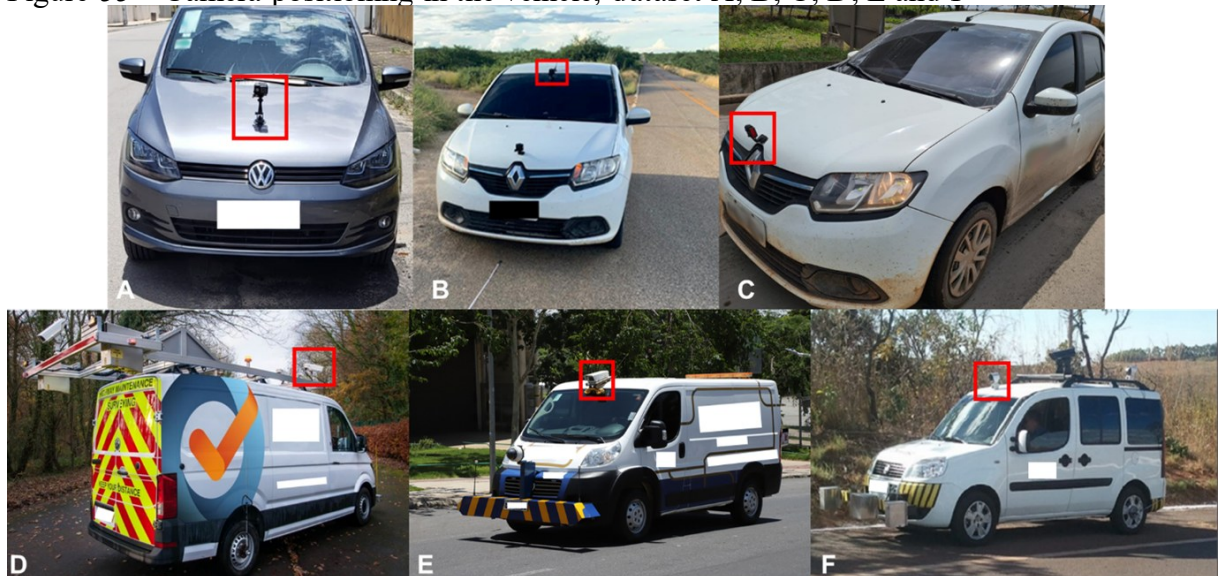
Source: elaborated by the author.

A total of six datasets were utilized, encompassing a diverse range of geographical regions and incorporating variations in camera types, vehicle positions, and viewing angles. Figure 55 provides an overview of the vehicles used for image acquisition in each dataset. The images in Dataset A were captured using a GoPro camera, while Datasets B and C were obtained with Garmin cameras. These devices, which are relatively affordable (costing less than $300), come equipped with essential accessories such as a protective case and a suction mount for secure installation. Notably, Dataset A was the only one collected using a GPS-enabled camera, enabling precise georeferencing of the captured images.

Three private companies provided images for this study and supplied information regarding the Functional Evaluation Vehicle used in their data collection processes. While the camera positions used for image acquisition were specified (Figure 55, datasets D, E and F), details regarding the specific camera models were not available. In addition to capturing images, the Functional Evaluation Vehicle records pavement quality indicators such as the International Roughness Index (IRI), rutting, and cracking area; however, this study did not have access to these additional data.

It sought to uphold the velocity of the vehicle within the range of 40 km/h to 70 km/h, constantly adhering to the prescribed speed limits on highways. The images were recorded during the day and in good weather conditions, as requested by DNIT (2003). Also encompass a range of scenarios with varying lighting conditions and shadows cast by items located at the periphery or on the sidewalks of the streets. This is demonstrated in the mosaic displayed in Figure 56, which exhibits an example of acquired images from each dataset.

Figure 56 – Examples of images collected, dataset A, B, C, D, E and F
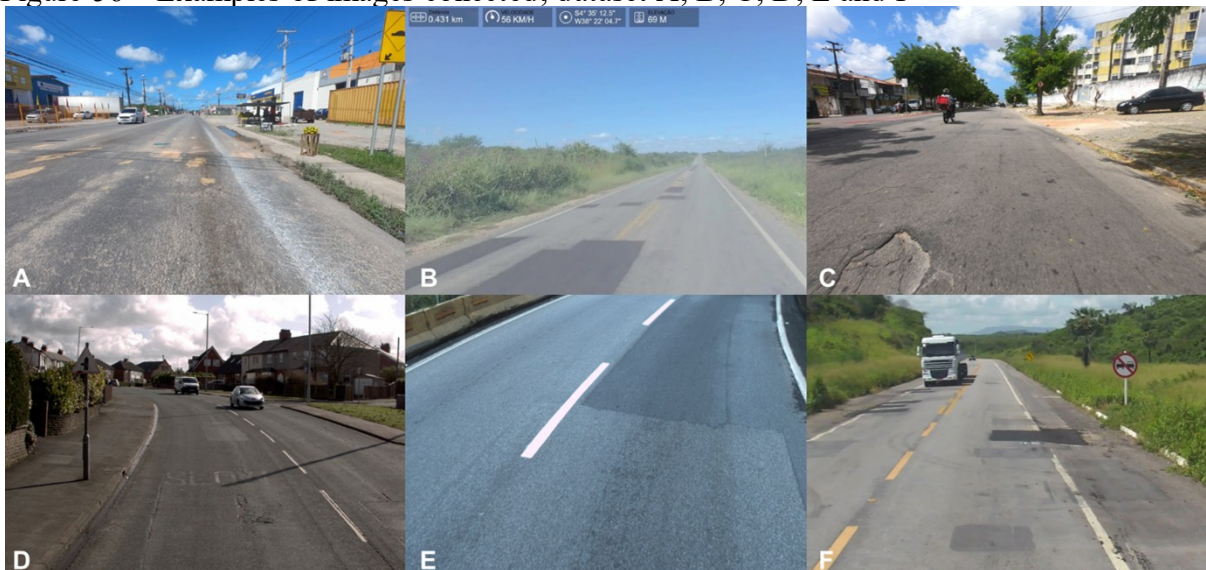
Source: elaborated by the author.

All datasets were collected in JPG format; however, only Dataset A includes geolocation information necessary for georeferencing. Consequently, images used for final system testing were exclusively selected from this dataset, as one of the system's objectives is to generate a map with the geolocated results of the model predictions. The data collection process was carried out specifically for this research, with the explicit goal of creating datasets which comprises data obtained from low-cost cameras for training the CNN and designated for the case study analysis. Table 28 presents the image dimensions for each dataset, highlighting the variability in resolution across different sources, which contributes to a more diverse training dataset.

Table 28 – Summary of images dimensions for each dataset

Dataset	Original Size of Image
A	1920x1080
B	3840x2160
C	1600x1200
D	1280x720
E	4000x3000
F	1600x1200

Source: elaborated by the author.

3.1.2 Surface Pavement Type Datasets

Different pavement surface types were considered in this study, including asphalt, concrete, unpaved, interlocking, and cobblestone/stone, as illustrated in Figure 57.

Figure 57 – Pavement Classes (A) Asphalt, (B) Concrete, (C) Unpaved, (D) Interlocking and (E) Cobblestone/Stone
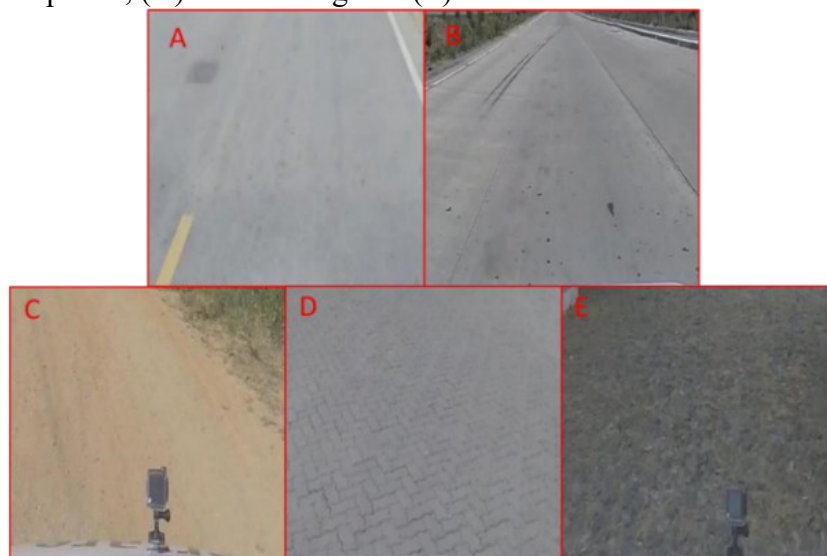
Source: elaborated by the author.

The majority of the images corresponded to asphalt roads. Part of the dataset was pre-labelled by the companies that provided the images, while the remaining portion was manually labelled by the researchers. Initially, a subset of 10,000 asphalt images and 5,000 images for each of the other pavement types was randomly selected for model training. Each image was individually examined to exclude misclassified samples, images that did not depict pavement surfaces, low-quality images, and highly similar images. The similarity check was performed using the Structural Similarity Index Measure (SSIM). After this filtering process, the number of images retained for training represented approximately 9.3% of the total dataset, and these images were subsequently processed for model development. Table 29 provides details on the composition of the datasets, the total number of images for each pavement type, and the number of images selected for model training.

Table 29 – Datasets composition

Label	Road Dataset Quantity	Image Quantity (Total)		Image Quantity (included in the model)	
		n	%	n	%
Asphalt	99	184,449	65.1%	9,576	3.4%
Concrete	11	11,372	4.0%	4,266	1.5%
Cobblestone	39	51,702	18.2%	4,092	1.4%
Interlocking	22	13,862	4.9%	4,214	1.5%
Unpaved	40	22,016	7.8%	4,237	1.5%
Total	**116**	**283,401**	**100.0%**	**26,385**	**9.3%**

Source: elaborated by the author.

There is significant variability in pavement surfaces, particularly along rural and urban roads. Among the available datasets, 56 contained more than one type of pavement, of which 24 datasets included two types of pavements, 28 datasets contained three types, and four datasets featured four distinct pavement types.

It was decided to work with ROI (Region of Interest) in the image to enable the analysis and block training. The input images represent 25% of the original size, in which the nine different variations of the images were grouped into two types. Type 1 is more focused on the pavement, and at the bottom of the image, the vehicle does not appear, and type 2 is a broader image, and the vehicle hood appears (Figure 58). This procedure is adopted because the large part of the image is not pavement elements and has a low sharpness of the road sections is far from the vehicle. After defining the ROI, the images are resized to 300x300 pixels.

Figure 58 – Pre-processing action to ROI
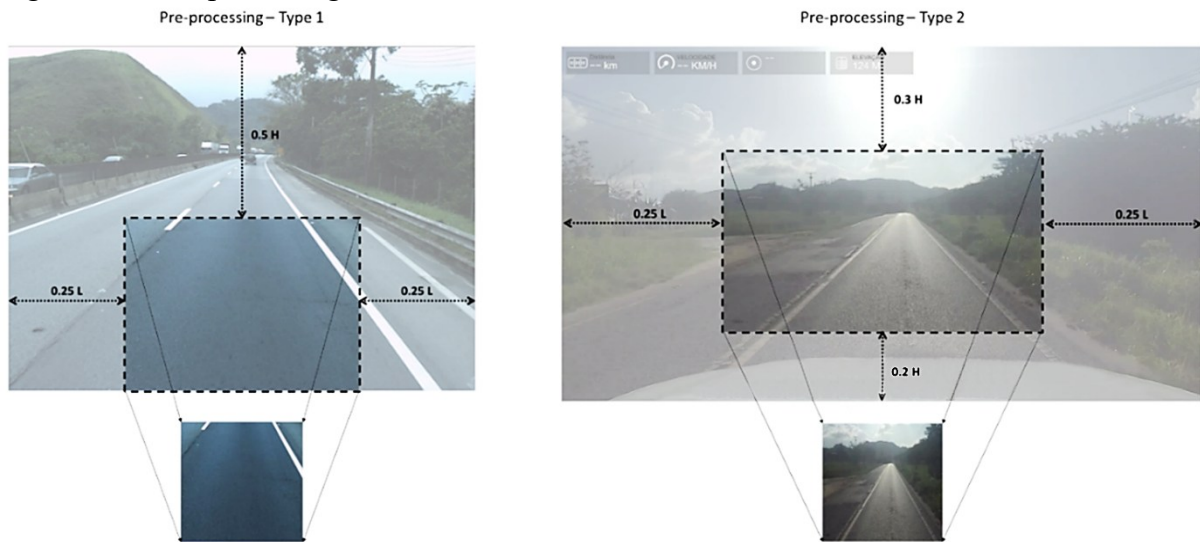
Source: elaborated by the author.

An important observation is that after pre-processing and defining the region of interest (ROI), the dataset was partitioned, with 80% of the images allocated to the training set and 20% to the validation set. The image allocation was performed randomly to ensure a diverse and representative distribution of samples across both sets.

3.1.3 Defects Pavement Datasets

The same sets of images were used for the development of the second study; however, the number of roads and images analyzed was significantly smaller. This reduction was due to the specific focus of the second study on identifying defects in asphalt pavements. As a result, images containing other types of pavement surfaces were excluded from the research, as well as roads that were in good condition.

Like the first study, image pre-processing was required. The train model methodology proposed in Espíndola *et al.* (2023b) focuses on identifying defects in image fragments rather than the entire image. Consequently, the dataset was retained using this methodology. Each dataset contains images of various dimensions, necessitating distinct pre-processing for each. The pre-processing stage of the study images involved following the sequence indicated in Figure 59 to perform image cropping and after resizing. Figure 59 illustrates the pre-processing of dataset E, where the four parts (1000x1000 pixels) at the bottom of the image are cropped. This allows the model to focus on and analyze areas predominantly composed of pavement. Finally, the dimensions of each cropped section were reduced to 300x300 pixels to standardize input data for the model.

Figure 59 – Pre-processing images of Dataset E: Crop and Resize
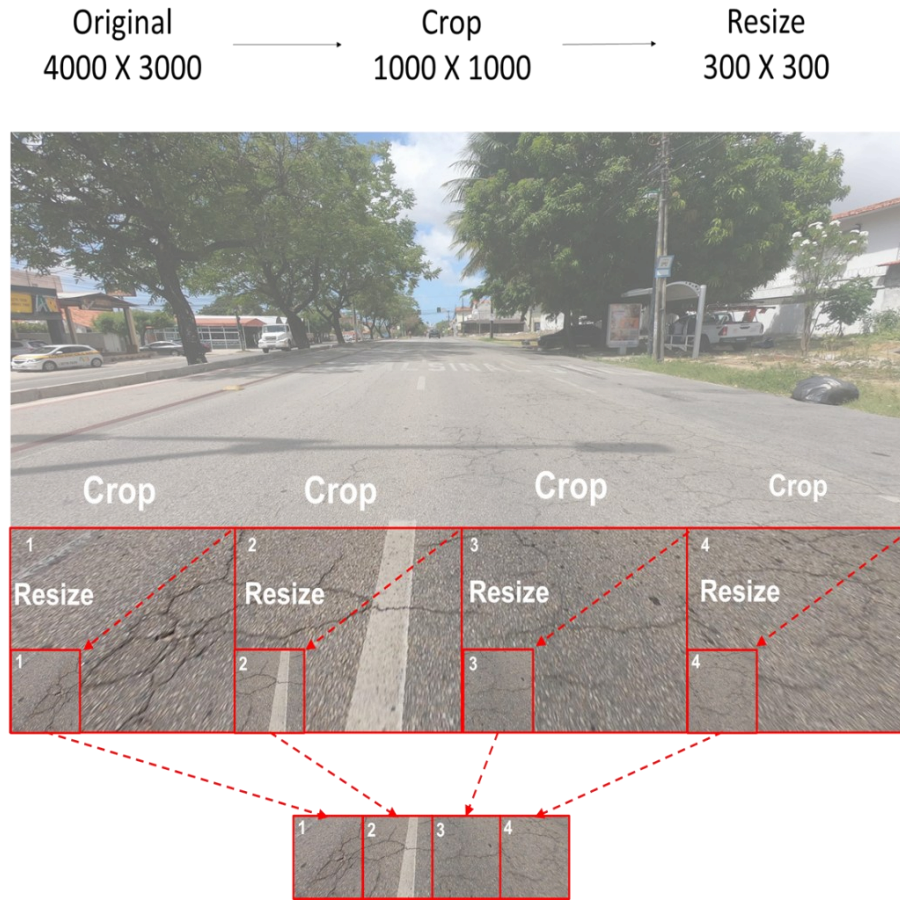
Source: elaborated by the author.

Table 30 presents the original image dimensions for each dataset, along with the sizes and quantities of the cropped sections. All cropped parts were then resized to a uniform size of 300x300 pixels, which corresponds to the model's input size. It is important to highlight that both the cropping and resizing processes were performed automatically through a scripted approach. The only task that required manual intervention was the annotation of pavement distress in the images.

Table 30 – Summary of dataset preprocessing

Dataset	Original Size of Image	Pre Processar	
		Crop	Resize
A	1920x1080	Crop in 3 parts of 640x640	300x300
B	3840x2160	Crop in 4 parts of 960x960	300x300
C	1600x1200	Crop in 3 parts of 533x533	300x300
D	1280x720	Crop in 3 parts of 426x426	300x300
E	4000x3000	Crop in 4 parts of 1000x1000	300x300
F	1600x1200	Crop in 3 parts of 533x533	300x300

Source: elaborated by the author.

The model includes six different classes of defects: Bleeding, Connected Crack, Isolated Crack, Patch, Pothole and Raveling. Like the model in the first study, the datasets used were divided between training and validation. The dataset was partitioned, with 80% of the images allocated to the training set (32,304 images) and 20% to the validation set (8,076 images), with the allocation performed randomly. The number of images and defects in each dataset can be observed in Table 31.

Table 31 – Datasets Summary

Dataset	Quantity							
	Images Total	Distress Total	Connected Crack	Isolated Crack	Patch	Pothole	Raveling	Bleeding
A	6,300	8,752	257	477	3,815	3,528	73	602
	20.10%	138.92%	2.94%	5.45%	43.59%	40.31%	0.83%	6.88%
B	2,500	3,099	412	1,292	723	188	476	8
	3.90%	123.96%	13.29%	41.69%	23.33%	6.07%	15.36%	0.26%
C	10,000	13,990	422	6,335	4,444	138	2,011	640
	24.76%	139.90%	3.02%	45.28%	31.77%	0.99%	14.37%	4.57%
D	5,100	4,172	404	676	544	348	464	1,736
	12.63%	81.80%	9.68%	16.20%	13.04%	8.34%	11.12%	41.61%
E	10,120	12,410	1,062	4,098	3,626	1,526	1.168	930
	27.10%	122.63%	8.56%	33.02%	29.22%	12.30%	9.41%	7.49%
F	6,360	7,424	296	2,852	1,080	16	3,144	36
	9.80%	116.73%	3.99%	38.42%	14.55%	0.22%	42.35%	0.48%
All Together	40,380	49,847	2,853	15,730	14,232	5,744	7,336	3,952
	100.00%	123.44%	5.72%	31.56%	28.55%	11.52%	14.72%	7.93%

Source: elaborated by the author.

The Table 31 presents the distribution of images and pavement distress types across different datasets. Each dataset contains a varying number of total images and defect types, with Isolated Crack and Patch being the most frequent, appearing in approximately 30% of the dataset images. In contrast, Pothole and Raveling defects were present in just over 12% of the images, while Bleeding appeared in only 8%. Connected Crack, despite representing only 5.7% of the dataset, demonstrated better performance than Bleeding, suggesting that increasing the number of samples for these defects could improve model accuracy.

3.1.4 Defects Inventory Database

The input data for the third study were no longer images but rather information regarding road quality and the presence or absence of specific defects in the road samples. This shift in data format allowed for a broader analysis of pavement conditions, focusing on structured data rather than visual assessments.

It is important to note that the IGG encompasses a comprehensive assessment of 20 distinct flaws, which are further categorized into 8 different types of defect classification. The defects considered by the IGG methodology include isolated cracks (FC-01); connected cracks without erosion (FC-02); connected cracks with erosion (FC-03); deformation; corrugation, shoving, potholes and slippage; bleeding; raveling; and patches. Figure 60 presents an example of a defect inventory conducted during the research on Rua Santos Ferraz (ID:27), demonstrating the practical application of the IGG methodology in pavement condition assessment.

Figure 60 – Original Defects Inventory Spreadsheet of Rua Santos Ferraz (ID:27)

Estacas	TRINCAS									AFUNDAMENTO				OUTROS EFEITOS					
	ISOLADAS					INTERLIGADAS				PLÁSTICO		CONSOLID							
						FC-2		FC-3											
	FI	TTC	TTL	TLL	TRR	J	TB	JE	TBE	ALP	ATP	ALC	ATC	O	P	EX	D	R	m
1	✓			✓				✓					✓		✓		✓		20
2	✓			✓									✓		✓		✓	✓	40
3	✓			✓									✓				✓		60
4	✓			✓					✓				✓		✓		✓		80
5	✓			✓									✓		✓		✓	✓	100
6	✓			✓									✓				✓	✓	120
7	✓			✓		✓											✓	✓	140
8	✓	✓		✓													✓		160
9	✓			✓											✓		✓		180
10	✓																✓		200
11	✓			✓													✓		220
12	✓			✓				✓							✓		✓	✓	240
13	✓					✓									✓		✓		260
14	✓																✓	✓	280
15	✓			✓									✓		✓		✓		300
16	✓																✓		320
17	✓			✓											✓		✓	✓	340
18	✓														✓		✓		360
19	✓			✓											✓		✓		380
20	✓														✓		✓	✓	400
21	✓																		420
22	✓																		440
23	✓																		460
24	✓																		480
25	✓																		500

Source: elaborated by the author.

A comprehensive analysis was conducted on a total of 30 streets, avenues, and highways, all of which were single-lane roads with asphalt surfaces and varying hierarchical classifications, as shown in Table 32 and Figure 61. Two highways (ID: 3 - AL-101 and ID: 22 - AL-110) fall outside the metropolitan territory of Maceió but serve as linking routes to the inland areas of the state of Alagoas.

It is important to note that some roads were assessed multiple times between 2015 and 2023, leading to a total of 40 inventoried segments following the DNIT 006/2003-PRO

norm. These segments collectively cover a total distance of 16.86 km, as detailed in Table 32. At certain intervals, the condition of the streets was reevaluated to track the progression of pavement deterioration and identify any interventions carried out by the authorities.

Table 32 – The compilation of streets and avenues inventory

Date	ID	Location	Extension (m)
May, 2016	3	AL-101	220
May, 2018	22	AL-110	520
October, 2016	4	Av. Almirante Álvaro Calheiros	300
October, 2016	9	Av. Cachoeira do Mirim	520
November, 2017	15	Av. Empresário Carlos da Silva Nogueira	520
May, 2018	23	Av. Empresário Nelson Oliveira Menezes	520
December, 2022	35	Av. Empresário Nelson Oliveira Menezes	200
February, 2022	28	Av. Eng. Paulo Brandão Nogueira	200
May, 2016	1	Av. Francisco Freire Ribeiro	200
October, 2016	8	Av. Governador Luis Cavalcante	520
October, 2016	2	Av. Governador Luis Cavalcante	220
May, 2017	11	Av. Governador Luís Cavalcante	520
November, 2017	19	Av. José Airton Gondim Lamenha	520
November, 2017	16	Av. José Haílton dos Santos	200
June, 2022	31	Av. Menino Marcelo	220
June, 2018	21	Av. Paulo Holanda	500
July, 2022	32	Av. Paulo Holanda	600
December, 2022	34	Av. Paulo Holanda	500
October, 2023	40	Av. Paulo Holanda	520
April, 2023	39	Av. Pierre Chalita	420
December, 2022	36	Av. Presidente Getúlio Vargas	500
May, 2017	14	Rua Adelaíde Melo Mota	520
February, 2022	29	Rua Desportista Cláudio da Rocha Lima	240
May, 2017	12	Rua Dona Alzira Aguiar	500
November, 2017	17	Rua Empresário Jorge Montenegro	200
February, 2022	25	Rua General Hermes	440
June, 2022	37	Rua General Hermes	520
October, 2016	5	Rua Hugo Corrêa Paes	520
October, 2016	7	Rua 01	520
December, 2022	33	Rua 01	500
April, 2023	38	Rua 01	200
May, 2017	13	Rua João Norberto de Lima	520
February, 2022	26	Rua José Henrique do Nascimento	200
July, 2022	30	Rua Paulina Maria de Mendonça	500
November, 2017	18	Rua Paulina Maria de Mendonça	500
October, 2016	6	Rua Professor Antônio Nemésio de Albuquerque	540
November, 2017	20	Rua Professor Manuel Coelho Neto	500
February, 2022	27	Rua Santos Ferraz	500
May, 2017	10	Rua Senador Firmino de Vasconcelos	500
May, 2018	24	Rua Senador Rui Palmeira	500
Total			**16,860.0**

Source: elaborated by the author.
* In 2016, two assessments were conducted on Av. Governador Luis Cavalcante, focusing on separate segments.

Figure 61 – Segments assessed within the Maceió-AL metropolitan area
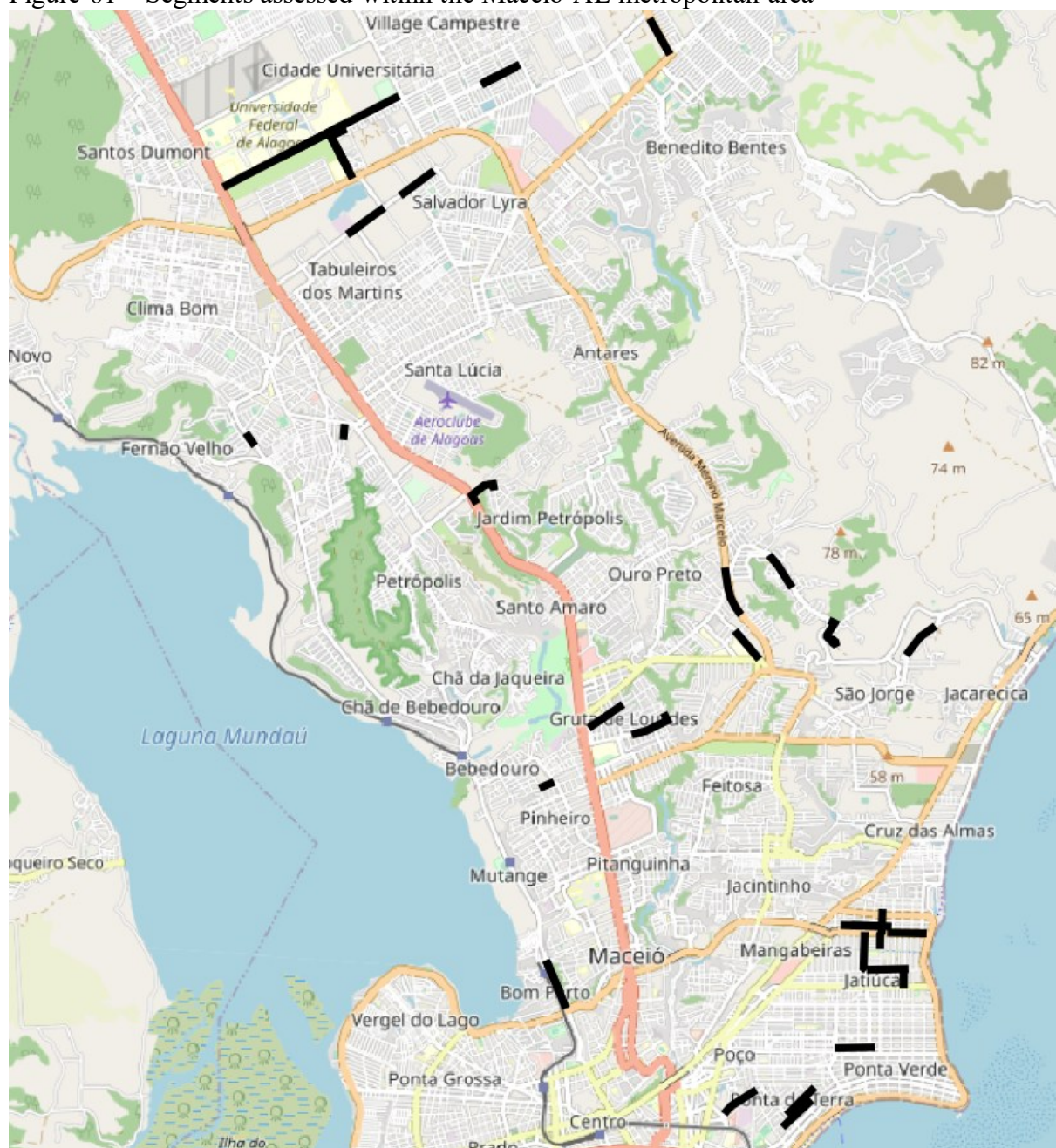
Source: elaborated by the author.

The data that was gathered pertained to the inventory of defects found in the 40 segments indicated in Table 32. The guidelines outlined in the DNIT 006/2003-PRO norm were adhered to, which permits the computation of a pavement quality metric by utilizing an assessment of pavement distresses. The IGG is determined on a sampling basis for samples with a predetermined area, spacing between them of 20 m, and when switching between tracks rather than for the total road area. The evaluation is specified in each sample by a 6-meter-long region, 3 meters before and 3 meters after the road stakes, as illustrated in Figure 62 (IPR; DNIT, 2003b).

Figure 62 – Demarcation of areas to register defects
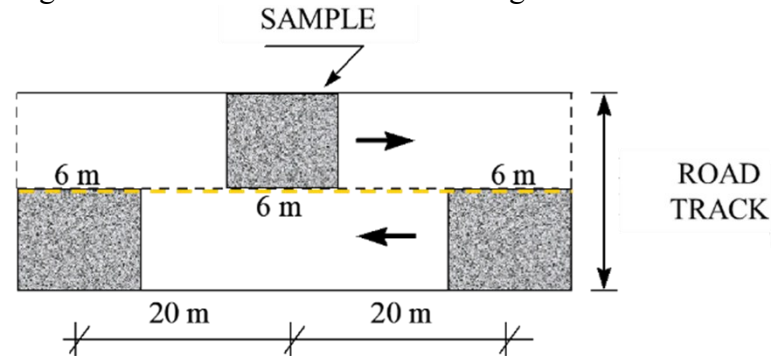
During the process of inventorying the road samples, a comprehensive photographic report was generated, documenting the many problems observed. For illustrative purposes, Figure 63 displays a compilation of recorded flaws observed along ID: 35 - Av. Empresário Nelson Oliveira Menezes in December 2022.

Figure 63 – Images from the Photographic Report of the ID assessment: 35 - Av. Empresário Nelson Oliveira Menezes. (A) Sample 5 – Connected crack (alligator crack) and patch; (B) Sample 4 – Connected crack (alligator crack) and patch; (C) Sample 7 – Raveling and patch; (D) Sample 11 – Isolated crack (longitudinal).
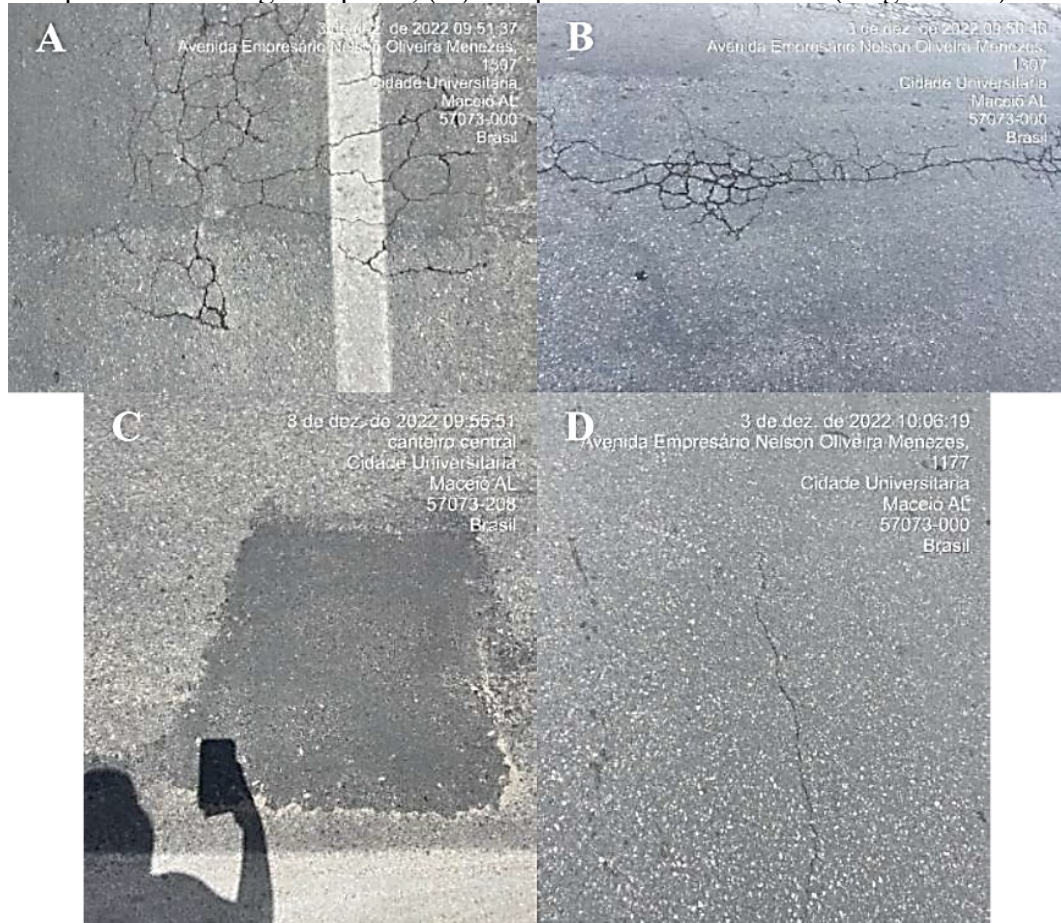
Source: elaborated by the author.

3.2 Model training

During the training process, various input parameters and optimization techniques of the Convolutional Neural Network (CNN) must be tested and analyzed to evaluate the model's performance metrics. The parameters assessed included image size, batch size, number of epochs, CNN architectures, and learning rate (LR).

Both models employed the following optimization techniques: AdamW, weight decay, and L2 regularization. AdamW enhances optimization efficiency by decoupling weight decay from gradient updates, leading to improved generalization. Meanwhile, weight decay and L2 regularization serve as regularization techniques that mitigate overfitting by penalizing large weight values, thereby enhancing the model's robustness and stability.

3.2.1 Training the network – Surface Pavement Type

During CNN training, mathematical optimization is performed using a loss function to automatically adjust the network's weights and biases. This process involves presenting a labeled training dataset to the optimization procedure to minimize the loss function globally. In this study, multiple CNN models were trained based on AlexNet, VGG16, ResNet34, and ResNet50 architectures.

Given the high computational demand of CNN training, GPU acceleration was essential for parallel processing. Google Colab's computing resources were utilized, with GPU allocation varying per model. The Tesla P100-PCIE-16GB GPU was assigned for the final training of AlexNet and ResNet50, while the Tesla T4 GPU was used for ResNet34 and VGG16. The training was conducted using PyTorch (version 1.5.0+cu101), integrated with the fastai v2 library.

The model training process utilized AlexNet, VGG16, ResNet34, and ResNet50 as pre-trained models, with input images resized to 224×224 pixels and a batch size of 64. Training was conducted in two stages, each consisting of 15 epochs, using cross-entropy (flattened loss) as the loss function. The dataset was randomly split, with 80% of the images used for training and 20% for validation.

The effectiveness of transfer learning depends on the similarity between the new dataset and the data the CNN was originally trained on. Since different layers capture different levels of abstraction, adjustments to the pre-trained weights vary. The final layers often require

significant modifications to adapt to specific classification tasks, while earlier layers, which detect fundamental features such as edges and textures, typically need only minor adjustments.

Since transfer learning is applied in the network architecture, the CNN is fine-tuned in two stages. In the first stage, the initial layer group is frozen, preventing its parameters from being updated, while the remaining layers are trained for 15 epochs. In the second stage, the initial layer group is unfrozen, and training resumes from the previous state for an additional 15 epochs. This approach ensures that the initial layers, which capture fundamental features, are trained only after the later layers have been optimized. This method maximizes the benefits of transfer learning by focusing on adjusting the final layers of the pre-trained models, which include AlexNet, VGG16, ResNet34, and ResNet50.

To determine the learning rate, a cyclical learning rates (CLR) method, also known as "one-cycle" training, is used. This method plots the validation loss against the learning rate, allowing for the selection of an optimal learning rate just before the validation loss increases sharply (Figure 64). The learning rate can be fixed or set within a specific range. For the first stage of training, the optimal range was between 10^{-4} and 6×10^{-2}, while for the second stage, it was between 10^{-7} and 2×10^{-6}. The cyclical learning rate method was applied across all pre-trained models, and since the resulting graphs showed minimal variation, the learning rate values were kept consistent across all cases to ensure fair metric comparisons.

Figure 64 – The graph of validation loss vs. LR of ResNet50
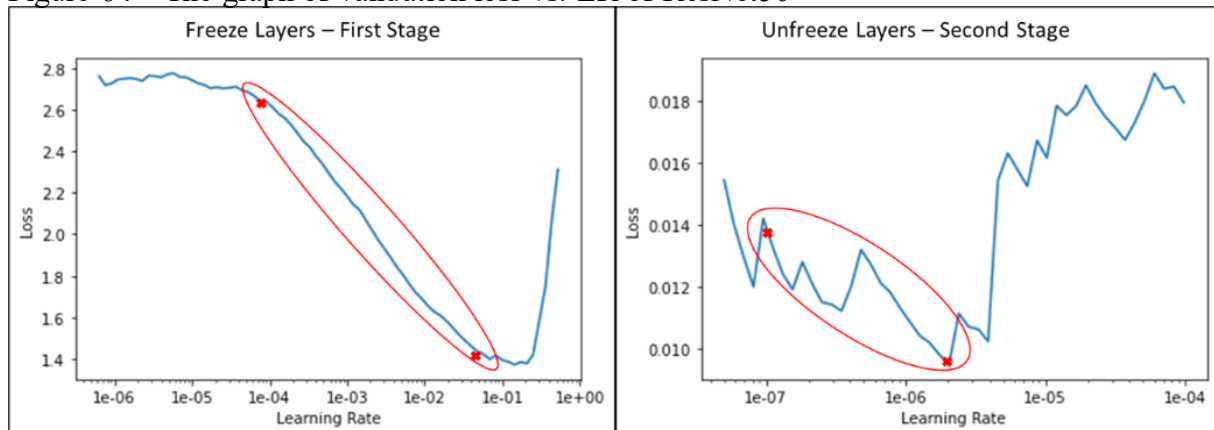
Source: elaborated by the author.

After training each model, the loss graphs were analyzed against the number of processed batches to monitor and prevent overfitting. This step was essential in ensuring that the models generalized well without excessive memorization of the training data.

Figure 65 presents the parameters used during the optimization of the neural network and the final construction of the Surface Pavement Type Model classification. The

parameters highlighted in bold in Figure 65 represent the configurations selected for the final version of the model, indicating those that demonstrated the best performance during the training.

Figure 65 – Training parameters tested – Surface Pavement Type Model

Image Size: **224x224**
Batch Size: **64;** 32; 16
Epochs: 10; **15**; 20
Loss Function: **cross-entropy**
CNN Architectures: AlexNet; VGG16; ResNet34; **ResNet50**
Learning Rate (LR) *: **Stage 1: LR = 10^{-4} to 6.10^{-2}**
                      Stage 2: LR = 10^{-7} to 2.10^{-6}
*Based on results of Cyclical Learning Rates (CLR)

Source: elaborated by the author.

3.2.2 Training the Network – Defects Pavement

The training methodology for the automatic pavement defect classification model followed the same sequence of steps as the previous model, with minor parameter adjustments.

ResNet50 was selected as the pre-trained architecture for developing the Convolutional Neural Network (CNN) due to its proven effectiveness in identifying pavement distress, as demonstrated in studies by Ranjbar, Nejad and Zakeri (2020), Espíndola *et al.* (2023b), and Sirhan, Bekhor and Sidess (2024). The model classifies pavement defects into six categories: bleeding, connected crack, isolated crack, patch, pothole, and raveling.

The datasets used for training and evaluation are detailed in Table 4. The dataset partitioning followed the same approach as the previous model, with 80% of the images allocated for training and 20% for validation. The image allocation was performed randomly to ensure unbiased distribution of data across the training and validation sets.

The batch sizes tested included 64, 32, and 16. Reducing the batch size from 64 to 32 improved the model's performance metrics. However, further reduction to 16 had a minimal impact on the results while significantly increasing the training time. As a result, a batch size of 32 was selected for the final training.

The training process was initially conducted with 10, 20, 25, and 30 epochs, with loss values monitored at the end of each set. The results indicated stable performance after 25 epochs, beyond which overfitting began to occur. Consequently, 25 epochs were chosen as the optimal training duration.

During the initial training phase, images of varying dimensions–224×224, 300×300, and 512×512 pixels–were evaluated. Training with 512×512 images considerably increased computational time, while using 224×224 images led to a decline in model performance. In contrast, 300×300-pixel images provided a balance between accuracy and efficiency. This resolution enhanced model predictions without significantly extending the training duration, making it the preferred choice.

Transfer learning was implemented in the network design, leading to the fine-tuning of CNNs in two stages. In the first stage, the initial layer group froze, preventing any updates to its parameters, while the final layer groups were trained for 25 epochs. In the second stage, the initial layer group was unfrozen, and training continued for an additional 25 epochs, building upon the previously trained state.

The cyclical learning rates (CLR) method was also applied to determine the optimal learning rate (LR). The LR values could either be fixed or within a specified range. For the first stage, the optimal fixed value was approximately 10^{-3}, while for the second stage, the LR range was set between 10^{-6} and 10^{-3}.

After completing the training process, the relationship between loss and the number of processed batches was analyzed, as shown in Figure 66, to assess model performance and mitigate overfitting.

Figure 66 – Learning Curve – Loss vs. batches processed
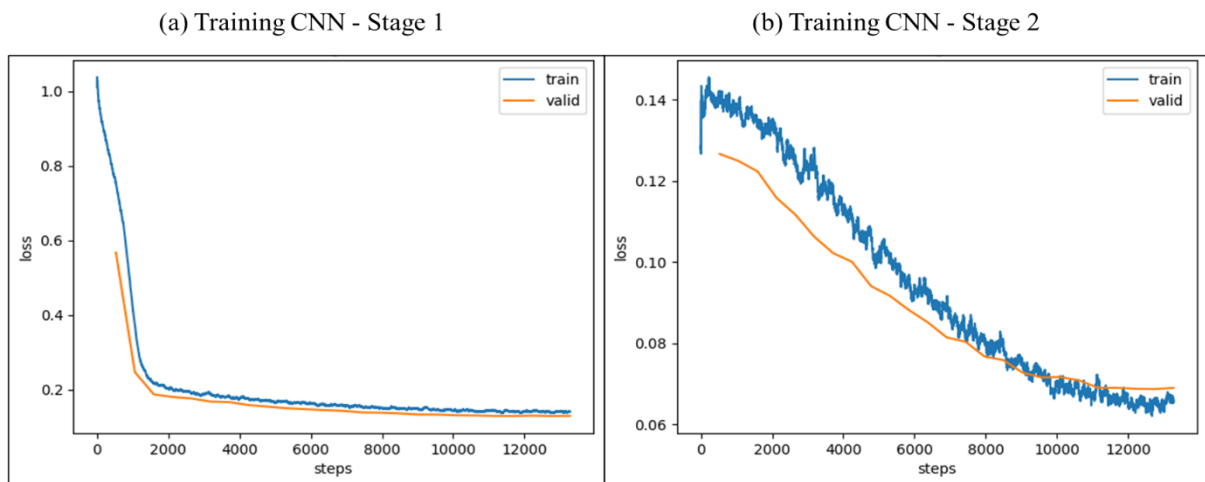
Source: elaborated by the author.

In Figure 67, the input parameters tested during the training of the multi-label classification defects pavement model are presented. The parameters highlighted were those used in the final version of the model. The ResNet50-based models adopted a batch size of 32, 25 training epochs, cross-entropy as the loss function, and an image size of 300×300 pixels.

Figure 67 – Training parameters tested – Defects Pavement Model

Image Size: 224x224; **300x300;** 512x512
Batch Size: 64; **32**; 16
Epochs: 10; 20; **25;** 30
Loss Function: **cross-entropy**
CNN Architectures: **ResNet50**
Learning Rate (LR) *: **Stage 1: LR = 10^{-3}**
                          Stage 2: LR = 10^{-6} to 10^{-3}
*Based on results of Cyclical Learning Rates (CLR)

Source: elaborated by the author.

3.2.3 Metric for Evaluating Classification Models

Accuracy, precision, recall, and F1-score were utilized to evaluate the models' performances, with the confusion matrix employed to identify which classes the classifier is misclassifying and the extent of this confusion. Accuracy is appropriate when class distribution is similar, offering a straightforward measure of correct predictions. However, in scenarios with imbalanced classes, the F1-score provides a more reliable metric as it balances precision and recall, ensuring both true positives and false negatives are adequately considered. True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are fundamental components in these evaluations, providing detailed insights into the classifier's performance across different metrics.

3.2.3.1 Accuracy

Accuracy is a fundamental metric for evaluating the performance of classification models. It is defined as the proportion of correctly predicted examples to the total number of instances, calculated using Equation 16.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (16)$$

Accuracy provides a straightforward assessment of a model's overall performance by indicating the ratio of correct predictions, making it particularly useful when all classes in the dataset are of equal importance as it gives a general overview of how well the model performs across all classes. However, its utility diminishes in situations where there is class imbalance. In such cases, the dataset is dominated by one or more classes, leading to potentially

misleading accuracy scores. A model might achieve high accuracy by predominantly predicting the majority class, thereby failing to capture the underlying patterns of the minority classes. This limitation underscores the need for additional metrics to evaluate model performance comprehensively in imbalanced datasets, ensuring that the model accurately represents all classes within the dataset (Sokolova; Lapalme, 2009).

3.2.3.2 Precision

Precision, also known as positive predictive value, measures the proportion of true positive predictions among all positive predictions made by the model. It is calculated using Equation 17:

$$Precision = \frac{TP}{TP+FP} \tag{17}$$

Precision is particularly significant in scenarios where the cost of false positives is high, such as in medical diagnostics, where a high precision indicates that the model has a low rate of false positives, ensuring that most of the predicted positive cases are indeed positive. This metric aids in understanding the exactness of the model's predictions by focusing on the accuracy of the positive predictions made. Precision is crucial in fields where minimizing the occurrence of false positives is essential, thereby enhancing the reliability and trustworthiness of the model's outputs. By providing a clear measure of the proportion of relevant instances among the retrieved instances, precision assists in assessing the model's performance in discriminating between classes, ensuring that the model accurately identifies positive cases without overestimating their prevalence (Powers, 2011).

3.2.3.3 Recall

Recall, also known as sensitivity or the true positive rate, measures the proportion of true positive predictions among all actual positive instances. It is calculated using Equation 18:

$$Recall = \frac{TP}{TP+FN} \tag{18}$$

Recall is crucial in situations where it is important to capture as many positive instances as possible, such as in disease screening, where a high recall ensures that most actual positive cases are identified by the model, thereby minimizing the number of false negatives. This metric helps in understanding the completeness of the model's predictions by focusing on the ability to correctly identify positive instances. In contexts where missing positive cases could have severe consequences, such as in medical diagnostics or fraud detection, a high recall is essential. By providing a measure of the proportion of actual positives correctly identified, recall allows researchers to assess the sensitivity of the model to the positive class, ensuring that the model does not overlook critical instances (Sokolova; Lapalme, 2009).

3.2.3.4 F1-Score

The F1-Score is the harmonic means of precision and recall, providing a single metric that balances both aspects. It is calculated using Equation 19.

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{19}$$

The F1-Score effectively summarizes the predictive performance of a model by combining precision and recall, metrics that often compete. This balance is particularly useful in scenarios where evaluating the trade-off between precision and recall is necessary, such as in cases of class imbalance. In such contexts, accuracy alone might be misleading, as it does not account for the distribution of classes. A high F1-Score indicates that the model achieves a good balance between precision and recall, offering a more comprehensive measure of performance (Chinchor, 1992).

The relevance of the F1-Score is particularly underscored in domains where both false positives and false negatives carry significant consequences, such as medical diagnostics, fraud detection, and information retrieval. By considering both precision and recall, the F1-Score ensures that the model captures as many positive instances as possible while maintaining a low rate of false positives. This dual consideration makes the F1-Score a critical metric for assessing the effectiveness of classification models, especially in imbalanced datasets (Sitarz, 2023).

3.2.3.5 Confusion Matrix

The confusion matrix is a critical tool for evaluating the performance of machine learning classification models, particularly in scenarios where the output can belong to two or more classes. This matrix is a tabular representation that compares the predicted classifications made by the model to the actual classifications in the dataset. Each element of the matrix provides a count of instances corresponding to the combination of predicted and actual class labels, thus allowing for a detailed breakdown of the model's performance into True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), Figure 68. For multi-class classification problems, the confusion matrix extends to a more complex structure where each row represents the instances of an actual class, and each column represents the instances of a predicted class. This expansion allows for the assessment of the model's performance across multiple classes, highlighting any classes where the model might be performing poorly (Sokolova; Lapalme, 2009).

Figure 68 – Confusion Matrix

	Actual positive TP+FN	Actual negative FP+TN
Classified positive TP + FP	TP	FP
Classified negative FN + TN	FN	TN

Source: Sitarz (2023).

The confusion matrix is particularly valuable in contexts where class imbalance exists. In such scenarios, simple accuracy metrics can be misleading, as a model might appear to perform well by predominantly predicting the majority class. The confusion matrix reveals the true performance by showing how well the model differentiates between all classes, regardless of their distribution. This detailed performance analysis is essential for developing robust models that perform well across all classes. By analyzing the confusion matrix, researchers can identify patterns of misclassification, which can inform model adjustments and improvements, leading to better overall model performance (Sitarz, 2023).

4 MANUSCRIPT PAPER 01: AUTOMATIC PAVEMENT SURFACE SYSTEM FOR ROAD INVENTORY

Abstract

There is a clear need to implement new technologies in the pavement management system; computer vision and machine learning are techniques used in developing an automatic pavement evaluation system. The models that are being developed assume that all segments evaluated are asphalt paved roads. However, developing countries have various pavement types, mainly urban roads. Therefore, the present work developed a model based on a deep convolution neural network to identify the type of road covering capable of classifying pavements in concrete, asphalt, interlocking, cobblestone/stone, and unpaved roads. The model developed obtained an accuracy of almost 100% for both the database and street images used for validation. Allowing the construction of an automatic inventory system for the type of pavement will facilitate the choice of system and parameters for evaluating the quality of the road's surface directed to the specific type of pavement.

4.1 Introduction

Transportation infrastructures are by far the most critical asset of any country. Therefore, mature, and well-maintained highway networks are crucial to developing a nation's economy, mainly for countries with a transport matrix based on the highway modal.

In general, developing countries have an unbalanced transport matrix and a high concentration of road transport. Therefore, the logistic issue is aggravated when the roads system does not have the structure to absorb the demand. Such as Brazil, where road transport concentrates approximately 61% of the movement of cargo and 95% of passengers. Besides, only 12.4% of the roads are paved, and of those, 67.5% were considered inadequate on at least one of the criteria (geometry, signage, or pavement), with 56.7% of the roads assessed showing some pavement defect (CNT; SEST SENAT, 2023).

Brazil has 213,500.0 km of paved roads, 30.8% of which are under federal administration and 69.2% under the care of state and municipal (CNT; SEST SENAT, 2023). For road network management, it is necessary to have a minimum of information, at least as length, type of pavement, traffic, weather, and some quality index along the road. The agency at the federal level (National Department of Transport Infrastructure [DNIT]/Brazil) has a compatible human and financial resource's structure to manage the road network under its administration. However, municipal or state agencies are interested in having a pavement management system (PMS) but do not have the structure to implement and maintain an effective PMS, mainly due to the difficulty of obtaining the minimum necessary data (Causim, 2001).

The absence of primary data on the network makes the implementation of infrastructure management unfeasible, preventing the definition of the maintenance and restoration plan, the investment plan for implementing new highways, and guidelines for urban networks and low-volume roads.

There are two approaches to road data collection, manual and automatic. The manual approach fell out of favor due to its subjectivity, high financial costs, time, and human resources, turning impractical. With technological advances, there is an increasing tendency to use automated surveys to assess the pavement with equipment ranging from simple cameras to high technology, such as sensors and lasers. Analysis techniques have also progressed steadily to deal with the vast amount of data (Gopalakrishnan *et al.*, 2017).

Researchers have used the neural network to classify pavement deterioration and quantify severity for the past twenty years. In the previous eight years, a convolutional neural network (CNN), a particular type of neural network widely known as deep learning (DL), has gained notable popularity in many sectors, especially in computer vision (Gopalakrishnan, 2018). However, DL techniques' application is constantly evolving in pavement engineering, with advances in hardware and software technologies. After a literature review about the application of deep learning in the pavement, the main focuses are on road defects and reveals that attention has been predominantly on the classification, detection, and segmentation of cracks (Gopalakrishnan *et al.*, 2017).

The first step in implementing the PMS is to inventory the road network and obtain the primary data. The management of the road infrastructure is directly dependent on the pavement, the survey methodology/data collection, equipment used, inventoried defects, maintenance solutions, and costs vary for each type of pavement. In addition, there is a significant variation in the pavement type in some urban networks, which can make it challenging to collect data for network management.

In this context, the work aims to present a method that incorporates a deep convolutional neural network-based model in automating the registration of the pavement type to auxiliary pavement management system and identify unpaved road zones for urban planning to cities.

4.2 Literature review

Deep learning methods focus on learning hierarchies of features with features from higher levels of the hierarchy formed by a combination of lower-level features. Learning

features at multiple levels of abstraction allow a system to learn complex functions mapping the input to the output directly from data, without depending entirely on human-crafted features (Ikram, 2019). A Convolutional Neural Network (CNN) is a network architecture for deep learning algorithms and is specifically used for image recognition and tasks that involve the processing of pixel data, or rather computer vision tasks like image classification, object detection, and segmentation, Figure 1.

Figure 1 – Example of cracks and potholes identifications by the methods (A) Multi-label Classification, (B) Detection, and (C) Segmentation
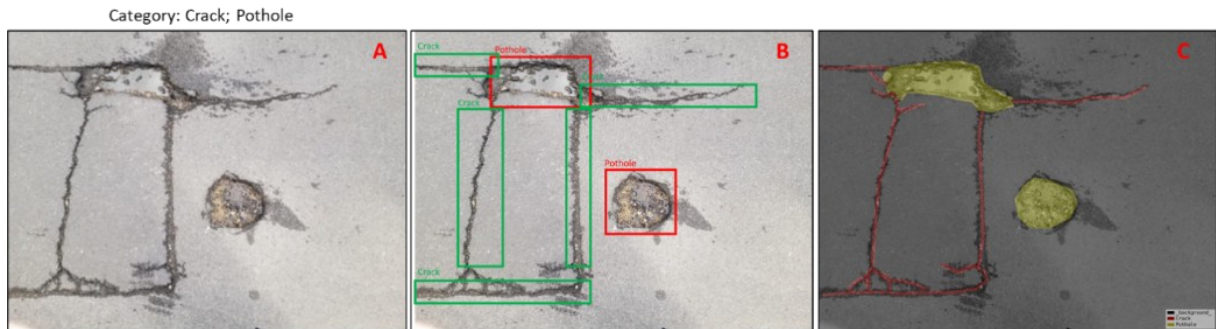
Source: Espíndola, Freitas and Nobre Júnior (2021) and Espíndola, Nobre Júnior and Silva Júnior (2021).

DL typically has billions of floating-point operations per second (FLOPS); however, when analyzing FLOPs in structures and architectures for the different computer vision tasks, classification structures require significantly less computational power. For example, architectures for classifications such as VGG16 have 15.5 Giga Floating Point Operations per Second (GFLOPS) and ResNet50 only 4 GFLOPS (He *et al.*, 2016), for another side, the architecture more used to object detection using YOLOv7 requires 103 GFLOPS (Ma *et al.*, 2024), and U-Net architecture to segmentation 221 GFLOPS (Ding *et al.*, 2020). Therefore, in the development of this research, the multi-label classification was used, as it presents significantly smaller FLOPS and, as will be shown below, satisfactory performance.

4.2.1 Transfer learning

A CNN typically requires many labelled images to achieve high prediction accuracy. However, it is impossible to collect hundreds of thousands of images and to label them manually. Transfer learning method allows building accurate models with fewer input resources. Rather of starting the learning process from scratch (or full training), it chooses a model already trained in a similar and a larger dataset to solve a similar problem, such as ImageNet (Aparna *et al.*, 2019). Unlike existing learning methods that randomly initialize the

weights of CNN, transfer learning uses the weights of pre-trained CNN as initial weights before the training process. Transfer learning is useful in solving image classification problems of various domains, and it is not different into pavement field (Tajbakhsh *et al.*, 2016).

4.2.2 Systematic learning rate finder

The learning rate (LR) is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum loss function (Murphy, 2012). Smith proposed (Smith, 2017) a method to find the optimum LR, based on the cyclical learning rates (CLR) method, called a "one-cycle" training schedule. This study shows that increasing the LR might have a short-term negative effect and still reach a longer-term beneficial effect. Furthermore, it reinforces the idea that the LR could vary within a range of values instead of adopting a stepwise fixed or exponentially decreasing value. One sets the minimum and maximum boundaries, and the LR cyclically varies between these bounds. The study proposes one training run for a few epochs to estimate the minimum and maximum boundary values. It is an "LR range test", run your model for a few epochs while letting the LR increase linearly between low and high LR values, Figure 2. The Smith (2017) method allowed for higher LR and was still able to achieve great accuracy results on CIFAR-10 and CIFAR-100 in fewer epochs.

Figure 2 – Example Loss x Learning Rate graph in both training stages
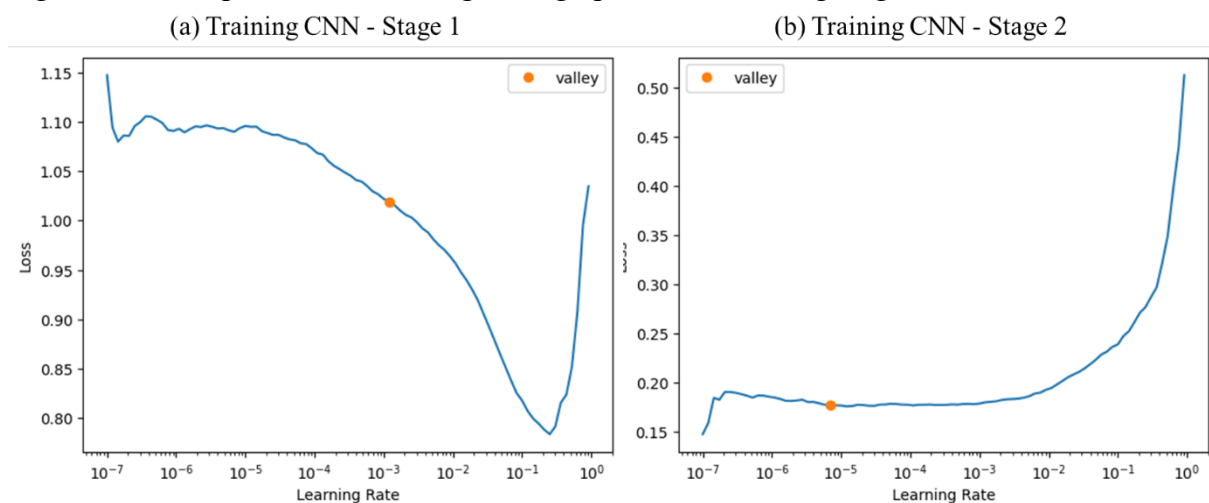
Source: elaborated by the author.

4.2.3 Related work

It is interesting to determine the type of pavement surface to avoid rework or manual work when determining the pavement condition. Because the pavement types of influence defects type, it is interesting to have a distress automatic detection model for each type of pavement. For example, there is no rutting on concrete pavements, only on asphalt. So, the first step in a fully automatic pavement management system is to segregate the samples by pavement type and evaluate each with the most appropriate methodology for the specific pavement type of the sample. This way, Table 1 shows other researchers developed to identify the road pavements, primarily focused on autonomous vehicles, but can also do the surface inventory road.

Table 1 – Road pavement research

Research	Data	Labels	Description
Varona, Monteserin and Teyseyre (2020)	Accelerometer information	Concrete panels, cobblestones, asphalt, and dirt road	Testing CNN, long short-term memory network (LSTM), and Reservoir Computing (RC) models, the best model was based-CNN, accuracy, precision, recall, and F-measure of 85%.
Menegazzo and Wangenheim (2020)	Accelerometer information	Dirt roads, cobblestone, and asphalt	The proposed models were based on LSTM, GRU, and CNN. CNN had best performance and obtained an average training accuracy of 93.04%, classifying asphalt segments with an F1-score of 98.60%, cobblestone with 86.09%, and dirt with 90.78%.
Pereira *et al.* (2018)	Image from Smartphone	Paved and unpaved road	Random forest and support vector machine models were compared with the proposed CNN. The model achieved the best performance to test the dataset, with precision 98.0%, recall 98.4%, and F1-Score 98.2%.
Tumen, Yildirim and Ergen (2018)	Image from Google Street View	Divide roads, express roads, paved street roads, unpaved street roads, parquet streets, and distorted roads	The proposed approach shows that the road types were determined with an accuracy of 91.41%. The model had the best performance to identify express roads with 96.0% and the worst accuracy to paved street roads with 79.2%.
Rateke, Justen and Wangenheim (2019)	Image from a camera into survey vehicle	Asphalt, paved and unpaved road	A classifier-based deep CNN model was proposed, using three datasets during the training KITTI (Geiger *et al.*, 2013), CaRINA (Shinzato *et al.*, 2016), and RTK (Rateke; Justen; Wangenheim, 2019). The lowest accuracy result was 86.05% for RTK, 76.23% to KITTI and 99.9% to CaRINA.
Espíndola, Freitas and Nobre Júnior (2021)	Image from a camera into survey vehicle	Asphalt, cobblestone, concrete, interlocking, and unpaved	Used one dataset with almost 50 thousand images to develop a classifier based ResNet50 to obtain accuracy, precision, and recall values of almost 100%.

Source: elaborated by the author.

4.3 Methods

The research aims to develop a low-cost automated method to invent the road elements (pavement type, vertical signage, and pavement defects, among others). For each study element, a module is being developed for the inventory system, all with the same data source (road videos) and application of Machine learning in developing the classification or object detection models to invent the road elements.

Figure 3 presents the system's working flow; initially, the images are captured by cameras that have GPS, for example, GoPro, Garmin, or even cell phones; then extracts the latitude and longitude information (enable the generation of inventory maps); and proceeds the application the automatic detection models, which can be signaling and pavement type; and for asphalt pavement images, the pavement evaluation module can be applied with the inventory of existing defects.

Figure 3 – Inventory system working flow
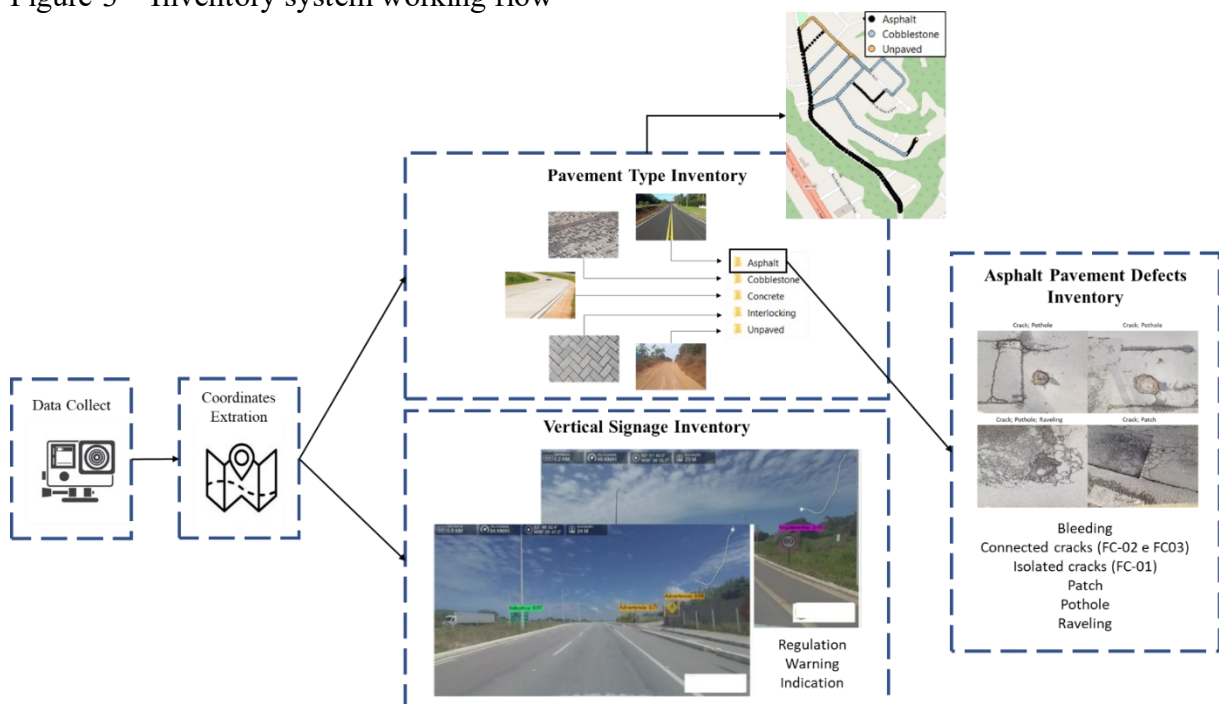
Source: elaborated by the author.

The paper presents the results of the pavement type inventory module, the data processing methodology, neural network training, the results of performances during validation and testing, and the outputs of the test segment inventory module. In the research, five different pavement surfaces were considered: asphalt, concrete, unpaved, interlocking, and cobblestone/stone, as shown in Figure 4.

Figure 4 – Pavement Classes (A) Asphalt, (B) Concrete, (C) Unpaved, (D) Interlocking and (E) Cobblestone/Stone
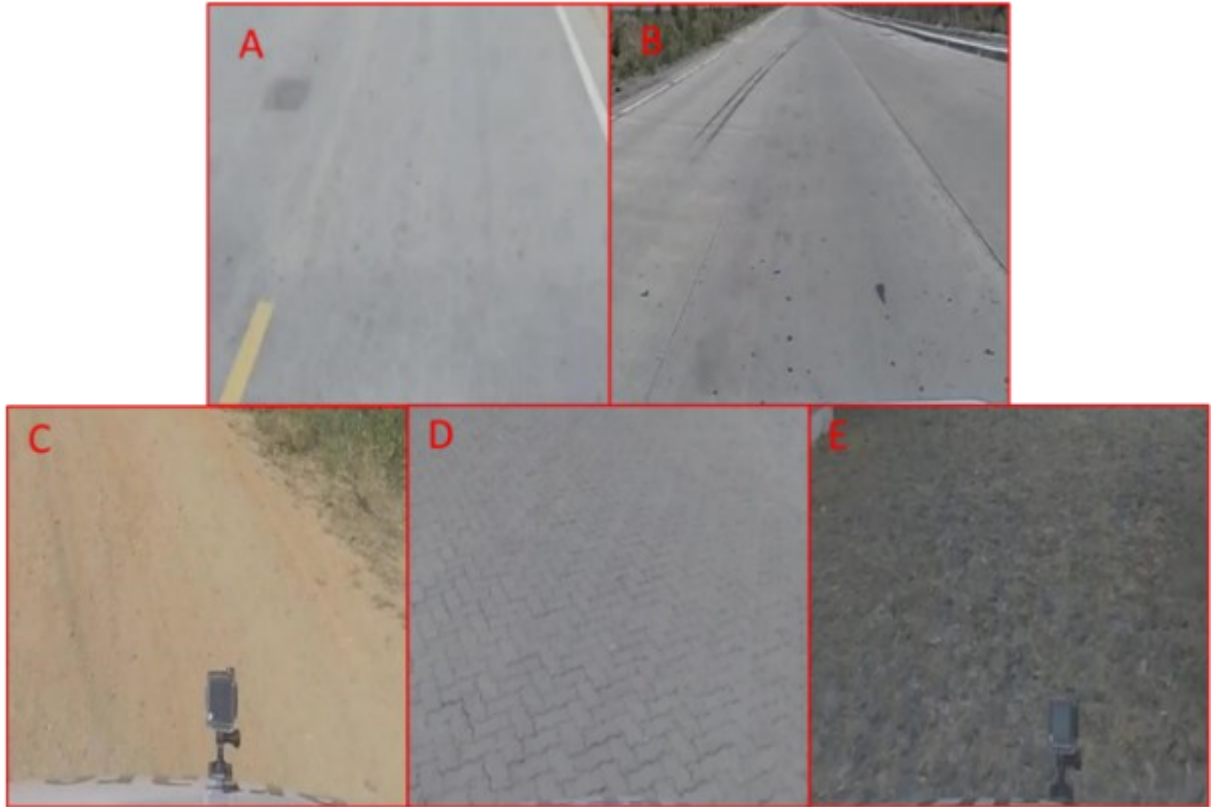
Source: elaborated by the author.

The development of the pavement type inventory module followed the flow shown in Figure 5; data collection (videos and images of roads from cameras attached to vehicles), extraction of latitude and longitude coordinates from images that had GPS data, cataloguing and treatments of available datasets, labelling of images; pre-processing of randomly selected images for neural network training; training the neural network with several different architectures; selection of the best performance model during validation, application of the model in the test segment and generation of the pavement type inventory map.

Figure 5 – The flow of development of the pavement type inventory module
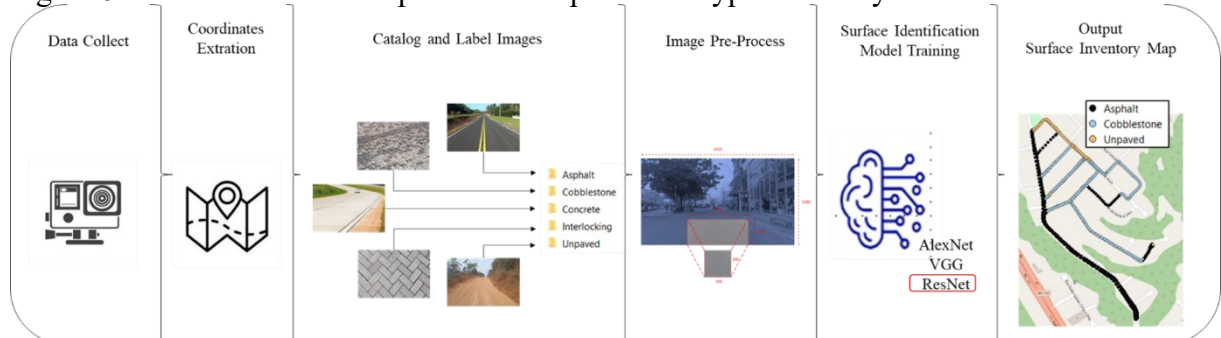
Source: elaborated by the author.

4.3.1 Dataset

In total, 116 different datasets were used in this research. They are composed of images from federal, state, and urban roads in Alagoas, Ceará, Rio de Janeiro, São Paulo, and Sergipe states from Brazil and London from the UK. More than 280 thousand images from survey videos were available for research; The images were predominantly of asphalted roads. Part of the images were labelled by the companies that made the images available; the researchers labelled others. Initially, it was determined to use only 10 thousand images of asphalt and 5 thousand of the other labels to enable training, randomly selected.

All images were analyzed individually to exclude images that might have been mislabeled, did not show the pavement, were of inferior quality, and were significantly similar. The similarity was verified by applying the structural similarity index measure (SSIM). The number of images of each label was reduced to about 9.3% of the total images, and those images passed to the pre-processing stage. Table 2 shows the number of datasets and images for each type of pavement and the number of images used for model development.

Table 2 – Datasets composition

Label	Road Dataset Quantity	Image Quantity (Total)		Image Quantity (included in the model)	
		n	%	n	%
Asphalt	99	184,449	65.1%	9,576	3.4%
Concrete	11	11,372	4.0%	4,266	1.5%
Cobblestone	39	51,702	18.2%	4,092	1.4%
Interlocking	22	13,862	4.9%	4,214	1.5%
Unpaved	40	22,016	7.8%	4,237	1.5%
Total	**116**	**283,401**	**100.0%**	**26,385**	**9.3%**

Source: elaborated by the author.

There is considerable variability in the pavement surface along rural and mainly urban roads. Of the dataset available for the research, 56 datasets had more than one type of pavement, of these 24 datasets had two types of pavements, 28 datasets had three types of pavements, and four datasets had four different types of pavements. It reinforces the importance of registering the pavement along the roads to speed up the analysis of the PMS and make the correct application of assessment and M&R methodologies for the different types of pavements.

4.3.2 Image pre-process

As data from 116 roads were used from 5 different sources (from 4 companies and one own), when analysing the images, nine different types of images of highways were identified, with variations in the positioning of the camera in the vehicle (Figure 6), angle (front, down face or rear) or image dimensions.

Figure 6 – Camera positioning in the vehicle

Source: elaborated by the author.

It was decided to work with ROI (Region of Interest) in the image to enable the analysis and block training. The input images represent 25% of the original size, in which the nine different variations of the images were grouped into two types. Type 1 is more focused on the pavement, and at the bottom of the image, the vehicle does not appear, and type 2 is a broader image, and the vehicle hood appears (Figure 7). This procedure is adopted because the large part of the image is not pavement elements and has a low sharpness of the road sections is far from the vehicle. After defining the ROI, the images are resized to 300x300 pixels.

Figure 7 – Pre-processing action to ROI
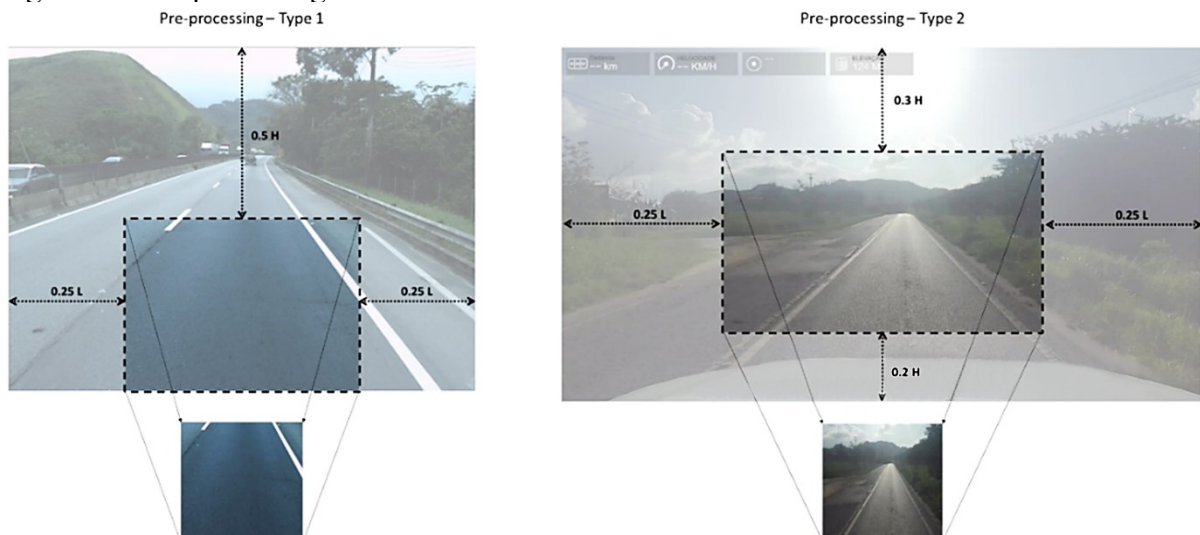
Source: elaborated by the author.

4.3.3 Training the network

During the training of a CNN mathematical optimization happens that uses a loss function, so the neural network's weights and biases can be automatically tuned. In this process, the training dataset with expected labels is presented to the optimization procedure to find a global minimum loss function. During the research, several CNN models were trained, based on the following architectures: AlexNet, VGG16, ResNet34 and ResNet50.

These architectures were selected for this research because they are architectures with excellent performance in the ImageNet challenge and have average or low FLOP values. AlexNet won the 2012 ImageNet challenge with a top-5 error rate of 16.4%. VGG was the winner along with GoogleNet for the 2014 ImageNet challenge with a top-5 error rate of 7.3%. While ResNet won the 2015 challenge with a top-5 error rate of 5.71%.

CNN training has massive computational time consumption. Therefore, it is necessary to use GPUs that can process data parallelly. For this purpose, the computing resources of Google Colab were used to conduct CNNs training, in which Colab randomly allocated GPU to each model performed. Tesla P100-PCIE-16GB GPU was allocated for final training of AlexNet, and ResNet50 architecture-based models, and Tesla T4 GPU to ResNet34 and VGG16. The DL framework used is PyTorch version1.5.0+cu101 with the fastaiv2 library.

The input parameters used for model training are AlexNet, VGG16, ResNet34, and ResNet50, as pre-train models; image size of 224x224 pixels; the batch size of 64; two-stage of training with 15 epochs each and cross-entropy (flattened loss) the loss function. And the dataset was divided into 80% of the images used for the training set and 20% for the validation set, each chosen randomly.

Accuracy, precision, recall, and F1-score were used to evaluate the models' performances. In addition, the confusion matrix will be used to identify which classes are being mistaken by the classifier and the intensity of this confusion. Accuracy can be used when the class distribution is similar, while the F1-score is a better metric when there are imbalanced classes.

Accuracy is a metric that generally describes how the model performs across all classes. It is useful when all classes are of equal importance. It is calculated as the ratio between the number of correct predictions to the total number of predictions, as shown in Equation 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision can define precision as the ratio of true positive (TP) (true predictions) and the total number of predicted positives (total predictions). Equation 2 is given as such:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

The recall is the ratio of true positive (true predictions) and the total of ground truth positives (total number). Equation 3 is given as such:

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The F1 score sums up the predictive performance of a model by combining two otherwise competing metrics – precision and recall, as the harmonic meaning between them, as shown in equation 4.

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

The level of training of models based on transfer learning depends on the resemblance of the data that the CNN will deal with in its lifecycle, the weights produced by training need to be altered to a greater or lesser extent. Thus, the last layers of the network often need enormous changes in parameters, while deeper levels that are already well trained in detecting basic features (such as edges and outlines) need minor modifications.

Since transfer learning is used in network architecture, CNN is fine-tuned in two steps. The first step freezes the first layer group so that all the parameters in it are not updated– the last layer-groups train for 15 epochs. In the second step, the first layer group is unfrozen. Then, the training is continued from the previous training state for an additional 15 epochs. In other words, the first layer group is trained after the other layer groups are well optimized. This procedure takes maximum advantage of transfer learning by training the last layers of the pre-trained models (AlexNet, VGG16, ResNet34, and ResNet50).

A cyclical learning rates (CLR) method or "one-cycle" training is used to define Learning Rate (LR). CLR method plots the validation loss graph against the LR (Figure 8); it is used to choose the values LR at the point before the validation loss increases sharply. LR can be fixed values or ranges, with the best results for Stage 1 being a range between 10^{-4} to 6.10^{-2}, and to stage 2 there was an LR range between 10^{-7} and 2.10^{-6}. The CLR was performed for all

cases and in all pre-trained models. The graphs only varied marginally, so keeping the LR values equal for all analyzed cases was decided. It will allow comparing the metrics results with equal hyperparameters by changing only the pre-trained model. After training each model, the graphs of Loss vs. Batches processed were analyzed to check and avoid overfitting.

Figure 8 – The graph of validation loss vs. LR of ResNet50
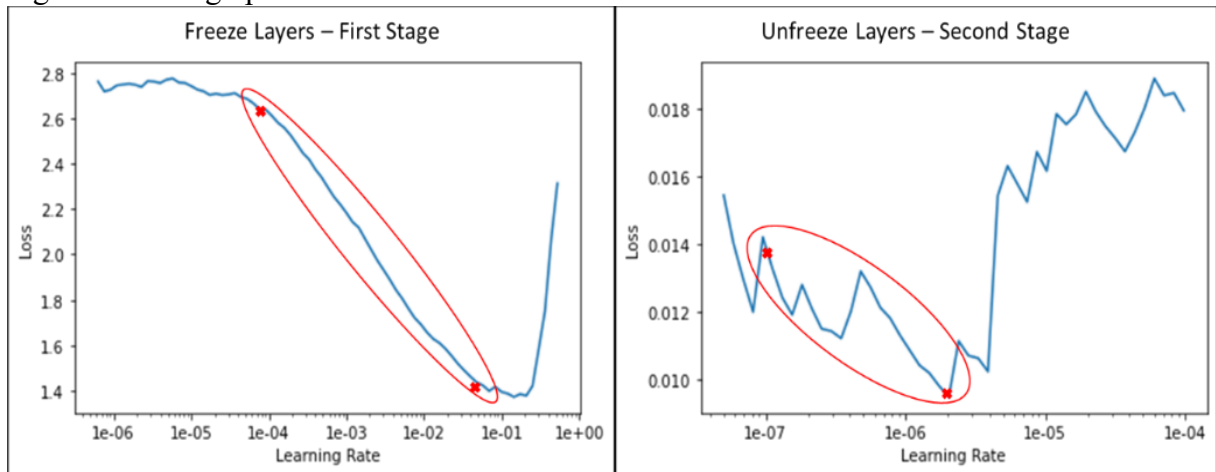
Source: elaborated by the author.

4.4 Results and discussion

Regardless of the architecture or type of pavement, the models present excellent results in the four metrics, all above 95%. They show that the models can incorporate it into the road inventory system. The CNN-based model's general results are presented in Table 3.

Table 3 – Performance of Pavement Surface Type Model per Architecture

(continue)

Architecture / Class	Accuracy	Precision	Recall	F1-score
AlexNet				
Asphalt	98.5%	97.4%	98.7%	98.0%
Cobblestone	99.1%	97.8%	97.3%	97.5%
Concrete	99.9%	100.0%	99.4%	99.7%
Interlocking	99.0%	97.3%	96.3%	96.8%
Unpaved	98.9%	96.3%	95.0%	95.7%
Avegare	99.1%	97.8%	97.3%	97.5%
VGG16				
Asphalt	99.6%	99.5%	99.5%	99.5%
Cobblestone	99.6%	99.3%	98.5%	98.9%
Concrete	99.9%	99.9%	99.4%	99.6%
Interlocking	99.9%	99.5%	99.9%	99.7%
Unpaved	99.5%	97.3%	98.5%	97.9%
Avegare	99.7%	99.1%	99.2%	99.1%

Table 3 – Performance of Pavement Surface Type Model per Architecture

(conclusion)

Architecture / Class	Accuracy	Precision	Recall	F1-score
ResNet34				
Asphalt	99.6%	99.6%	99.4%	99.5%
Cobblestone	99.7%	99.5%	98.6%	99.1%
Concrete	100.0%	100.0%	99.9%	99.9%
Interlocking	99.8%	99.0%	99.4%	99.2%
Unpaved	99.6%	97.6%	99.1%	98.4%
Avegare	99.7%	99.1%	99.3%	99.2%
ResNet50				
Asphalt	99.7%	99.8%	99.5%	99.6%
Cobblestone	99.9%	99.9%	99.5%	99.7%
Concrete	100.0%	100.0%	99.9%	99.9%
Interlocking	99.9%	99.6%	99.9%	99.8%
Unpaved	99.8%	98.5%	99.7%	99.1%
Avegare	99.9%	99.6%	99.7%	99.6%

Source: elaborated by the author.

Some highlights based on Table 3 results are the metrics in the best and lowest performance situations. For example, the lowest accuracy was 98.5% for asphalt in AlexNet, and the best was 100% for concrete in architectures ResNet 34 and 50. As for precision metrics, the lowest was 96.3% for unpaved on AlexNet, and the highest was 100% for concrete on ResNet 34 and 50 architectures. Regarding recall, the lowest recall is 95% unpaved on AlexNet, and the best recall is 99.9% for concrete on ResNet 34 and 50 architectures. The F1-score metric that compiles the recall and precision obtained a result of 95% for the unpaved in AlexNet and the best of 99.9% for the concrete in the architectures ResNet 34, and 50.

When analyzing the performance of the models considering all types of pavement surfaces, the architecture that presented the best performance was obtained by ResNet50. Another advantage of the ResNet50 architecture is that it has a low computational cost, makes the inference analysis of new images faster, and does not require high-performance computers.

When it analyzed the confusion matrix of each architecture (Figure 9), it verified that, regardless of the architecture, the class with the highest number of errors was the unpaved one, reaching 33 errors for AlexNet and only two errors for ResNet50. The second class with the highest errors was Asphalt, with 27 errors in AlexNet, and the smallest in ResNet50, with ten errors. VGG16 and ResNet34 had intermediate and close results. As the model with the best performance and the lower mistake was ResNet50, it is suggested that the management system incorporate this model into its structure.

Figure 9 – Confusion Matrix
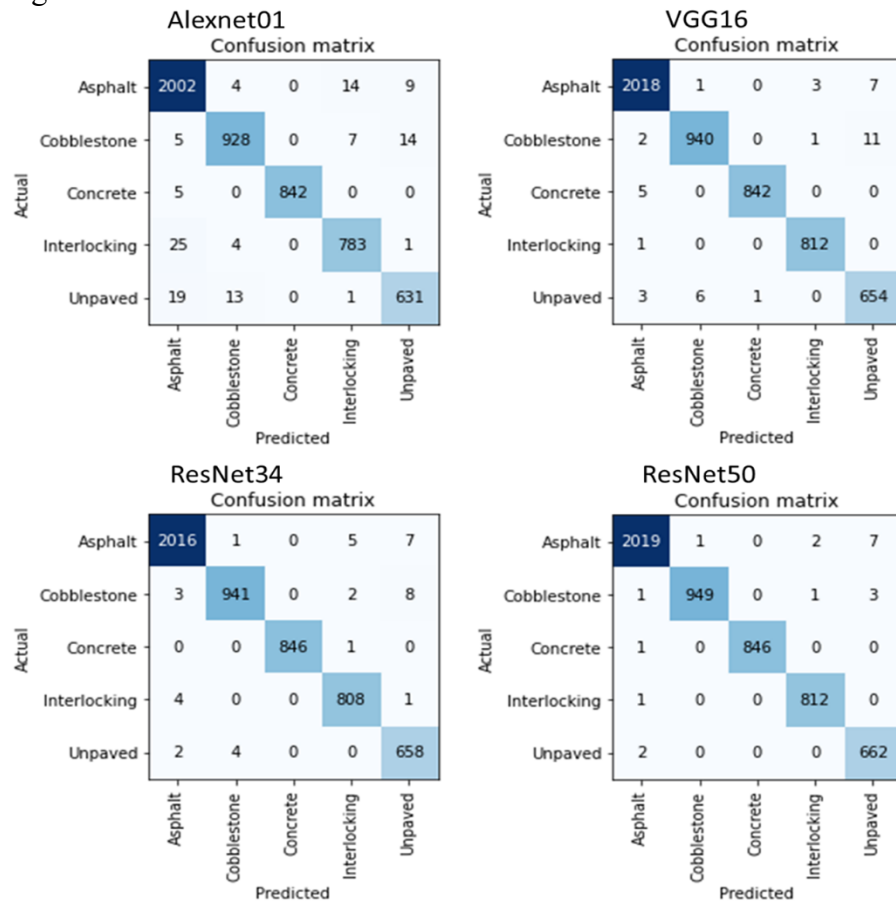
Source: elaborated by the author.

Figure 10 presents an example in which the neural network mistakes the classes. The original label was the cobblestone/stone road, and the model predicts unpaved road. However, it is an understandable error since the surface has deteriorated, covered with dirt, and rough stones are not visible. And Figure 11 presents several images that have been correctly classified by the model based on ResNet50.

Figure 10 – An example image that the model has misclassified

Source: elaborated by the author.

Figure 11 – Examples of images and the results of the classification of the pavement surface of the system
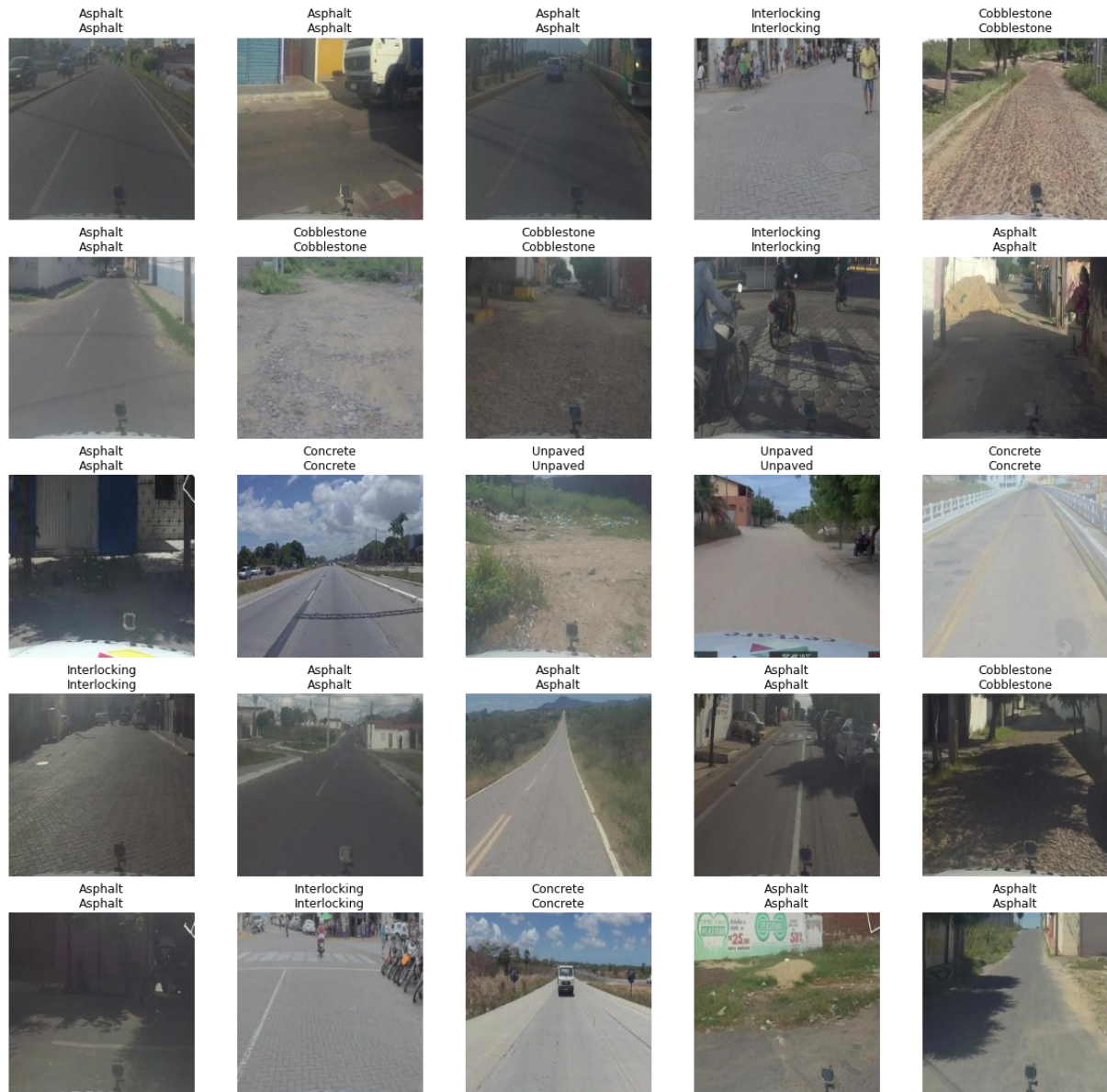
Source: elaborated by the author.

To validate the model, it tested the pavement surface module in an urban area with three of the five types of pavements available, asphalt, unpaved, and cobblestone. As a result, a total of 1592 images of the analysed streets were obtained, and the model had an accuracy of 99.69%, with only four errors in total, as shown in Table 4 and Figure 12.

Table 4 – Classification errors during validation

Image ID	Original Label	Prediction Label
G0489525 (a)	Unpaved	Asphalt
G0489526 (b)	Unpaved	Asphalt
G0489698 (c)	Cobblestone	Unpaved
G0489958 (d)	Cobblestone	Asphalt

Source: elaborated by the author.

As can be seen from the images and the types of errors that the model made are understandable errors. In Figure 12 (a) and (b) the vehicle was returning from the street, and the images refer to the parking lot of a church that is composed of a dark-colored soil mixed with old asphalt rap. And in Figure 12 (c) and (d) are at the interface of changing pavement types.

Figure 12 – Images with the classification prediction errors during validation: (a) G0489525, (b) G0489526, (c) G0489698 and (d) G0489958

Source: elaborated by the author.

Therefore, when applying the pavement surface module of the system, it is possible to catalog the pavement surface of the streets, and, as the images are georeferenced, it allows the production of maps with this information. Figure 13 shows the two stages of map generation. In the first map, each point represents an image and the surface classification; in the second, the information from the points was transformed into segments.

Figure 13 – Pavement surface inventory maps

Source: elaborated by the author.

4.5 Conclusions

The following conclusions are drawn from the research:

– The robust image bank and the use of Transfer learning and "systematic learning rate finder" techniques facilitate outstanding performance achievement.

– Among the deep learning architectures focused on task classification, ResNet50 presented the best results. However, the other architectures also achieved similar performances to humans, with accuracies above 95%, and can be used in the module of the pavement inventory system.

– In the model validation stage, the classification results were in line with expectations, with an almost perfect performance, with only four classification errors occurring and corroborating that the model can be used in the system, generating the necessary information for the inventory either in a database or in a map.

– An approach being investigated to reduce errors is to enhance the model by including the ability to identify many types of pavements within an image. This would be particularly useful when the image or sample is located precisely at the boundary where the pavement type changes.

– Another potential field for research is the categorization of different kinds of hot asphalt mixes, such as Stone Matrix Asphalt (SMA), Dense-Graded Mixes (DGA), or Open Graded Asphalt (OGA), which allows for the observation of variations in the texture of the asphalt mix.

4.6 References

APARNA *et al.* Convolutional neural networks based potholes detection using thermal imaging. **Journal of King Saud University – Computer and Information Sciences**, [*s. l.*], 2019. Available from: https://doi.org/10.1016/j.jksuci.2019.02.004. Access in: 15 June 2024.

CAUSIM, Patrícia Bolsonaro. **Estudo de um sistema de gerência de pavimentos para cidades de pequeno e medio porte**. 2001. 128 f. Universidade Estadual de Campinas, Campinas, 2001.

CNT; SEST SENAT. **Pesquisa CNT de rodovias 2023**. Brasília: [*s. n.*], 2023.

DING, Lei *et al.* MP-ResNet: Multi-path residual network for the semantic segmentation of high-resolution PolSAR images. **arXiv**, [*s. l.*], p. 1-5, 2020.

ESPÍNDOLA, Aline Calheiros; FREITAS, Gabriel Tavares de Melo; NOBRE JÚNIOR, Ernesto Ferreira. Pothole and patch detection on asphalt pavement using deep convolutional neural network. *In*: IBERO-LATIN-AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 42.; PAN-AMERICAN CONGRESS ON COMPUTATIONAL MECHANICS, 3., 2021, Rio de Janeiro. **Proceedings** [...]. Rio de Janeiro: ABMEC-IACM, 2021. Available from: https://repositorio.ufc.br/handle/riufc/63680. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; NOBRE JÚNIOR, Ernesto Ferreira; SILVA JÚNIOR, Elias Teodoro da. Pavement surface type classification based on deep learning to the automatic pavement evaluation system. *In*: IBERO-LATIN-AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 42.; PAN-AMERICAN CONGRESS ON COMPUTATIONAL MECHANICS, 3., 2021, Rio de Janeiro. **Proceedings** [...]. Rio de Janeiro: ABMEC-IACM, 2021. Available from: https://repositorio.ufc.br/handle/riufc/63676. Access in: 15 June 2024.

GEIGER, A. *et al.* Vision meets robotics: the KITTI dataset. **The International Journal of Robotics Research**, [*s. l.*], n. October, p. 1-6, 2013.

GOPALAKRISHNAN, Kasthurirangan *et al.* Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. **Construction and Building Materials**, [*s. l.*], v. 157, p. 322-330, 2017. Available from: https://doi.org/10.1016/j.conbuildmat.2017.09.110. Access in: 15 June 2024.

GOPALAKRISHNAN, Kasthurirangan. Deep learning in data-driven pavement image analysis and automated distress detection: a review. **Data**, [*s. l.*], v. 3, n. 3, 2018.

HE, Kaiming *et al.* Deep residual learning for image recognition. *In*: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [*S. l.: s. n.*], 2016. p. 770-778. Available from: http://image-net.org/challenges/LSVRC/2015/. Access in: 15 June 2024.

IKRAM, Chaudhry Rehan. **A benchmark for evaluating deep learning based image analytics**. Oslo: University of Oslo, 2019.

MA, Na *et al.* Wheat seed detection and counting method based on improved YOLOv8 Model. **Sensors**, [*s. l.*], v. 24, n. 5, 2024.

MENEGAZZO, Jeferson; WANGENHEIM, Aldo Von. Multi-contextual and multi-aspect analysis for road surface type classification through inertial sensors and deep learning. **IEEE X Brazilian Symposium on Computing Systems Engineering (SBESC)**, [*s. l.*], 2020.

MURPHY, Kevin P. **Machine learning**: a probabilistic perspective. [*S. l.: s. n.*], 2012.

PEREIRA, Vosco *et al.* Classification of paved and unpaved road image using convolutional neural network for road condition inspection system. **ICAICTA 2018 – 5th International Conference on Advanced Informatics: Concepts Theory and Applications**, [*s. l.*], p. 165-169, 2018.

RATEKE, Thiago; JUSTEN, Karla Aparecida; WANGENHEIM, Aldo von. Road surface classification with images captured from low-cost camera-road traversing knowledge (RTK) dataset. **Revista de Informatica Teorica e Aplicada**, [*s. l.*], v. 26, n. 3, p. 50-64, 2019.

SHINZATO, Patrick Y. *et al.* CaRINA dataset: An emerging-country urban scenario benchmark for road detection systems. **IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC**, [*s. l.*], p. 41-46, 2016.

SMITH, Leslie N. Cyclical learning rates for training neural networks. **Proceedings – 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017**, [*s. l.*], n. April, p. 464-472, 2017.

TAJBAKHSH, Nima *et al.* Convolutional neural networks for medical image analysis: full training or fine tuning? **IEEE Transactions on Medical Imaging**, [New York], v. 35, n. 5, p. 1299-1312, May 2016. DOI: https://doi.org/10.1109/TMI.2016.2535302. Available from: https://ieeexplore.ieee.org/document/7426826. Access in: 15 June 2024.

TUMEN, Vedat; YILDIRIM, Ozal; ERGEN, Burhan. Recognition of road type and quality for advanced driver assistance systems with deep learning. **Elektronika ir Elektrotechnika**, [*s. l.*], v. 24, n. 6, p. 67-74, 2018.

VARONA, Braian; MONTESERIN, Ariel; TEYSEYRE, Alfredo. A deep learning approach to automatic road surface monitoring and pothole detection. **Personal and Ubiquitous Computing**, [*s. l.*], v. 24, n. 4, p. 519-534, 2020.

5 MANUSCRIPT PAPER 02: A DEEP LEARNING-BASED METHOD FOR IDENTIFYING MULTIPLE DISTRESSES ON ASPHALT PAVEMENT

Abstract

Pavement deterioration measurements are crucial for pavement management systems (PMS). Image-based distress diagnosis is increasingly being employed in the analysis of road traffic speed and vehicle pavement conditions. This allows for a full assessment of road conditions throughout the whole network. Thus, a precise location of concern within the lane is unnecessary. Convolutional Neural Networks (CNN)-based multi-label classification (MLC) can be used to identify distress in a network-level Right-of-Way (ROW) video survey. MLC has the benefit of reduced computational requirements due to the utilization of lightweight classification networks. The ResNet-50 architecture of CNN was employed in this study to detect bleeding, connected crack, isolated crack, patch, pothole, and raveling utilizing MLC models. The optimal model achieved a 96% accuracy rate and an F1-score of 91% in identifying distress, even in the presence of variations in imaging hardware. This allows for the possibility of applying the classification method in a broader sense, making it possible to integrate it into other applications and pavement management systems on a larger scale. This method presents the possibility of quickly and accurately identifying distress from a video survey, without using expensive sensors like laser scanners.

Keywords: Deep learning, multi-label classification, distress diagnosis, pavement management

5.1 Introduction

The primary purpose of the transport service is to facilitate connectivity by enabling the movement of individuals, resources, and commodities, thereby enhancing market accessibility. Transportation is crucial for the operation of any country's economy. Nevertheless, the effectiveness of the service is constrained by the caliber and amount of infrastructure accessible for its operation. Efficient and properly maintained transportation networks are crucial for economic development. From this standpoint, the transportation infrastructure of a nation is its most valuable resource. Still, the task of upholding infrastructure to a superior level has been progressively difficult for authorities worldwide, primarily due to financial limitations in monitoring extensive networks. According to the National Transport Confederation (CNT) in 2023, 56.8% of Brazilian roadway pavements were in a state of deterioration (CNT; SEST SENAT, 2023). This issue is not limited to developing nations; as per the American Society of Civil Engineers, the road networks in the United States will be in a deteriorating state, graded as 'D', by 2021 (ASCE, 2021). The primary factors contributing to the absence of timely identification of decaying pavement are maintenance delays and

inadequate investment. Therefore, monitoring and studying pavements on a frequent basis are essential for developing optimal maintenance and rehabilitation (M&R).

Investigations of pavement are often undertaken on two levels. The major objective of the network-level survey is to determine the network's general condition for budgeting purposes and to identify locations of defective pavement for prioritizing specific investigation plans. The second step is to conduct a project-level study of the selected sites identified in the network-level survey, generally to determine the intensity of distress for planning maintenance and rehabilitation alternatives and the associated costs.

Attached to the survey vehicle is frequently high-tech, pricey equipment for evaluating pavement. The functional evaluation vehicle is outfitted with a satellite tracking system, digital barometer, GPS navigation, a precise odometer, laser sensors (profilometers), HD (high definition) videography, and regular image gathering. Alternately, the Intelligent Pavement Assessment Vehicle (iPAVe), which can conduct functional and structural assessments with the same equipment as before and the Traffic Speed Deflectometer, can be utilized. Occasionally, the cost and difficulty of evaluating the pavement using surveys truck prevents the assessment and monitoring of the quality of roadways, hence raising the costs associated with highway recovery when intervention occurs in a belated manner. Nevertheless, if it is possible to gather information for network-level management at a cheap cost using commercial cameras and automatic distresses recognition models, it will allow investments in iPAVe for project-level management to be deployed solely in prioritized sites, minimizing assessment costs, and permitting more investment in M&R operations.

Accurately identifying multiple types of distress from a single image enhances the potential to extract comprehensive information using a single sensor. However, the primary challenge lies in accurately labeling these images and developing robust analysis techniques to identify the various types of distress. Recently, convolutional neural networks (CNNs) in deep learning (DL) architectures have emerged as a promising solution to this challenge (Gopalakrishnan, 2018). In the field of pavement engineering, the application of DL techniques is a relatively recent development, showcasing significant potential for research and innovation. Literature reveals a growing focus on applying DL tools to the classification, detection, and segmentation of road surface cracks and potholes (Gopalakrishnan *et al.*, 2017). Researchers have been working on developing DL-based classifiers capable of identifying different defects by training on images containing a single defect (Aslan *et al.*, 2019; Eisenbach *et al.*, 2017; Nie; Wang, 2018; Ranjbar; Nejad; Zakeri, 2020). However, real-world pavement images often

contain multiple defects, making it more practical and realistic to develop models capable of handling multiple defects within a single image.

Multiple defects per image can be identified using object detection and segmentation techniques. Studies conducted by Ikram (2019), Cao, Liu and He (2020) and Dadashova, Dobrovolny and Tabesh (2021) have employed these methods. Regrettably, the computational cost of those approaches might be high due to the huge amount of data collected at the network level. This is because it necessitates laborious manual work to label images for training the model. To address this issue, it is necessary to have a clear understanding of the typical defects that are commonly observed in a Right of Way (ROW) image. By accurately identifying these defects, one can then select certain methods that are suitable for quantifying the severity of these defects. Quantifying defects at the network level is not crucial. Instead, the focus should be on identifying images with defects, which would reduce the need for significant human and computing resources in the classification process. By recognizing various sorts of defects present in an image, it is possible to assess the overall condition of a certain section of a road based on a series of images taken along that segment.

The objective of the research is to create a precise and computationally efficient deep learning-based multi-label classifier (MLC) model. This model will be used to identify multiple pavement defects from traffic speed network-level surveys. This paper aims to present the first stage of the research, focusing on a comparative analysis of various deep learning architectures to evaluate their efficiency in identifying multiple types of distress on asphalt surfaces. This study provides a concise review of literature on the measurement of pavement distress using deep learning techniques. It also includes a detailed explanation of the development of the model and the validation results obtained from a test section. The analysis is based on around 40,000 images from different roads surveyed by three commercially available systems. The six most widely identifiable distresses, such as potholes, isolates cracks, connected cracks, patches, raveling and bleeding, are included in this research.

5.2 Literature review

The literature extensively documents the use of traditional computer vision (CV) approaches for identifying pavement distress. Techniques such as image processing, which includes intensity-thresholding, edge and corner detection, and seed-based methods, as well as machine learning approaches including unsupervised learning, supervised learning, and deep learning, have been extensively employed for the identification and measurement of distress

(Du *et al.*, 2021; Ghosh; Smadi, 2021; Gong; Wang, 2021; Li *et al.*, 2021; Li *et al.*, 2011; Paulya *et al.*, 2017; Song; Wang, 2021; Zhang *et al.*, 2016; Zhou *et al.*, 2012).

The popularity of Deep Learning (DL) is increasing due to the widespread availability of datasets, Graphics Processing Units (GPUs), and DL algorithms (Kulkarni; Parikh; Abbott, 2017). Deep learning enables computer vision engineers to attain superior precision in tasks such as image classification, segmentation, and object detection as compared to classic computer vision techniques. Deep learning (DL) provides more versatility than computer vision (CV) methods, as Convolutional Neural Network (CNN) models and frameworks can be trained again using a customized dataset for each distinct application. In contrast, CV algorithms are generally more restricted to specific fields (O'Mahony *et al.*, 2019). Deep learning approaches, CNN-based, are designed to acquire knowledge about the hierarchical structure of features. In this approach, features at higher levels of the hierarchy are produced by combining lower-level characteristics. Acquiring features at multiple levels of abstraction enables a system to acquire intricate functionalities that directly transfer input to output simply from data, without depending on features manually designed by humans (Ikram, 2019). DL methods have shown success in diverse computer vision tasks, including image classification, object detection, and imagine segmentation.

Artificial neural networks (ANNs) have the capability to construct classifiers for binary, multi-class, and multi-label tasks. Binary classification algorithms determine whether images include defects or not, as shown by the output (Cao; Liu; He, 2020). In multi-class learning, each object is only associated with a single label, while in multi-task learning, different tasks may involve different domains and different datasets. Traditional two-class and multi-class problems can be transformed into multi-label problems by constraining each instance to have only one label (Zhou *et al.*, 2012).

Classification methods establish the label of the image, Figure 1 (A). If the image contains a certain number of defective pixels, the image is labeled as defective. The objective of object detection is to accurately classify and precisely locate items inside an image by enclosing them with bounding boxes, this is illustrated in Figure 1 (B). Image segmentation is the procedure of categorizing every individual pixel in the image according to a certain label, which is denoted as Figure 1 (C).

Figure 1 – Example of cracks and potholes identifications by the methods (A) Multi-label Classification, (B) Detection, and (C) Segmentation
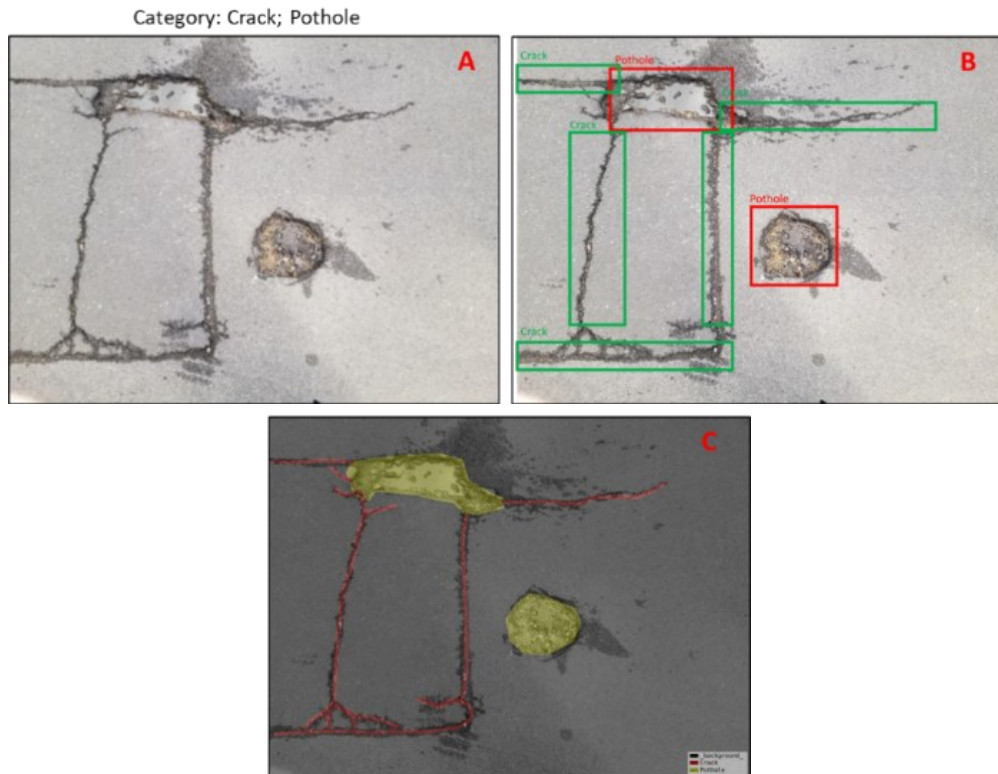
Source: Espíndola, Freitas and Nobre Júnior (2021) and Espíndola, Nobre Júnior and Silva Júnior (2021).

The computational demand of a DL network can be measured by its running speed, which is quantified by the number of floating-point operations it needs to execute per unit of time. This metric is commonly referred to as floating-point operations per second (FLOPS). Classification networks typically have lower computational requirements compared to other deep learning-based computer vision architectures. Classifier networks such as ResNet-34 and ResNet-50 have Giga Floating Point Operations per Second (GFLOPS) ranging from 4 (He *et al.*, 2016). Object detection using YOLOv7 requires 103 GFLOPS (Ma *et al.*, 2024), whereas segmentation with U-Net requires 221 GFLOPS (Ding *et al.*, 2020). It is important to note that once the model has been trained and is prepared, the process of doing inference or classification requires far fewer computer resources. The main constraint is the training model itself.

The pavement assessment has mostly concentrated on doing research in the domains of classification, object identification, and segmentation. Nevertheless, the focus of this work is to identify the several categories of defects seen in the pavement, rather than precisely determining their extent or location within the image. Hence, the computer vision methodology utilized in this study is multi-label classification. The study offers a detailed survey of the critical literature on deep learning-based models used for diagnosing these distresses, Table 1.

Table 1 – Research on pavement assessment using DL

Computer Vision Task	Classes	Relevant References
Classification	Crack	Aslan *et al.*, 2019; Chen; Jahanshahi, 2018; Eisenbach *et al.*, 2017; Espíndola *et al.*, 2023a; Gopalakrishnan *et al.*, 2017; Hoang; Nguyen; Tran, 2018; Li *et al.*, 2018; Nhat-Duc; Nguyen; Tran, 2018; Nie; Wang, 2018; Paulya *et al.*, 2017; Ranjbar; Nejad; Zakeri, 2020; Sirhan; Bekhor; Sidess, 2024; Wang; Hu, 2017; Xia, 2018; Yusof *et al.*, 2018; Zhang *et al.*, 2016; Zhou; Song, 2020
	Pothole	An *et al.*, 2018; Aparna *et al.*, 2019; Aslan *et al.*, 2019; Eisenbach *et al.*, 2017; Espíndola *et al.*, 2023a; Gopalakrishnan *et al.*, 2017; Pereira *et al.*, 2018; Sirhan; Bekhor; Sidess, 2024; Srinidhi; Devi, 2020; Xia, 2018; Ye *et al.*, 2019
	Others Distress	Eisenbach *et al.*, 2017; Espíndola *et al.*, 2023a; Nie; Wang, 2018; Sirhan; Bekhor; Sidess, 2024

Source: elaborated by the author.

Prior studies have predominantly employed binary and multi-class classifiers for classification purposes. Both binary and multi-class classifiers operate on the assumption that a pavement image contains just one instance of distress. Nevertheless, to fulfill this assumption, a substantial amount of manual labor is required to precisely isolate and remove all other elements from the photos, leaving only one defect visible. This action necessitates human intervention, as it cannot be fully automated. The presence of defects in an image affects the cropping process, making it impossible to achieve uniform image sizes. In addition, achieving flawless cropping may not always be feasible, for instance, when a crack and pothole are situated near each other in an entangled manner, as depicted in Figure 2 (a) and (b). Pavement images typically exhibit multiple defects within a single image, as depicted in Figure 2 (a)-(d). Therefore, it is crucial to detect any defects in such images to automate the process of defect identification.

Figure 2 – Multiple defects in pavement images
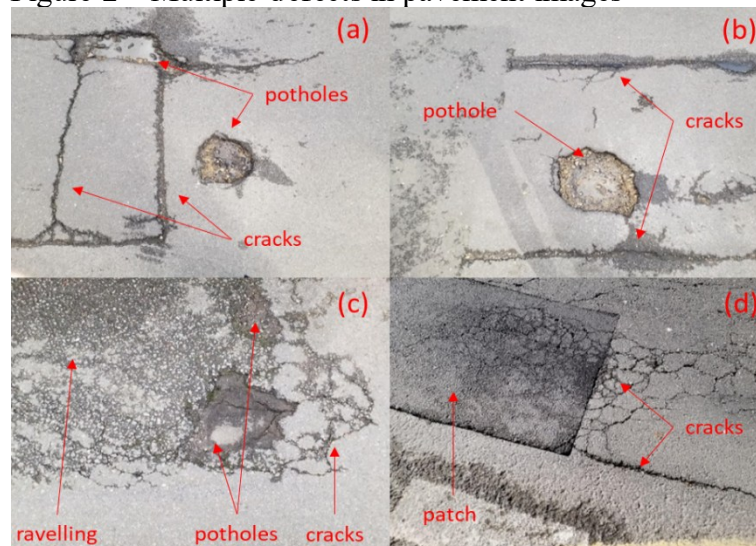
Source: Espíndola *et al.* (2023a).

The method of using multi-label classifiers to identify numerous defects concurrently on pavement was introduced by research such as Espíndola *et al.* (2023a) and Sirhan, Bekhor and Sidess (2024). This strategy aims to reduce or eliminate the need for manual image pre-processing to identify the distinct defects present in the image. This characteristic enables a more efficient and rapid examination of the pavement, potentially leading to a decrease in the financial expenses, workforce, and time required for the development of network-level PMS reports.

Creating a convolutional neural network (CNN) from scratch requires a large amount of labeled images, such as the 460,000 images provided by the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) for training. However, gathering and labeling such a vast number of images is often impractical. Transfer learning addresses this by using pre-trained models on similar large datasets, significantly reducing the need for extensive data and computational resources (Aparna *et al.*, 2019). This method enhances efficiency in training CNNs by leveraging pre-existing knowledge.

Figure 3 illustrates the sequential procedures required to train a Convolutional Neural Network (CNN) using Transfer learning. The primary pre-training models of CNN have evolved within the context of the ILSVRC.

Figure 3 – Transfer learning for ship classification and tracking in five steps
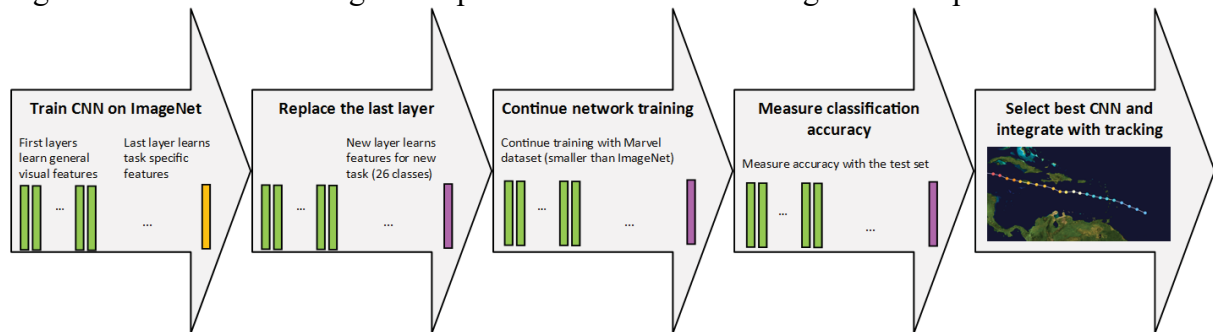
Source: Leclerc *et al.* (2018).

In 2016, He *et al.* (2016) introduced a deep Residual Learning Network (ResNet) and examined its performance by adding more layers. The study involved using CNNs with 18, 34, 50, 101, and 152 layers layered on top of each other (Figure 4). The network design achieved victory in multiple categories at the 2015 ImageNet and DAWNBench competitions (DAWNBench being a benchmark suite for complete deep learning training and inference). ResNet achieved a test accuracy of 94% on the CIFAR-10 dataset and top-5 accuracy scores of 93% on ImageNet (Coleman *et al.*, 2017).

Figure 4 – ResNet layers

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3{\times}3,\,64 \\ 3{\times}3,\,64 \end{bmatrix}{\times}2$	$\begin{bmatrix} 3{\times}3,\,64 \\ 3{\times}3,\,64 \end{bmatrix}{\times}3$	$\begin{bmatrix} 1{\times}1,\,64 \\ 3{\times}3,\,64 \\ 1{\times}1,\,256 \end{bmatrix}{\times}3$	$\begin{bmatrix} 1{\times}1,\,64 \\ 3{\times}3,\,64 \\ 1{\times}1,\,256 \end{bmatrix}{\times}3$	$\begin{bmatrix} 1{\times}1,\,64 \\ 3{\times}3,\,64 \\ 1{\times}1,\,256 \end{bmatrix}{\times}3$
conv3_x	28×28	$\begin{bmatrix} 3{\times}3,\,128 \\ 3{\times}3,\,128 \end{bmatrix}{\times}2$	$\begin{bmatrix} 3{\times}3,\,128 \\ 3{\times}3,\,128 \end{bmatrix}{\times}4$	$\begin{bmatrix} 1{\times}1,\,128 \\ 3{\times}3,\,128 \\ 1{\times}1,\,512 \end{bmatrix}{\times}4$	$\begin{bmatrix} 1{\times}1,\,128 \\ 3{\times}3,\,128 \\ 1{\times}1,\,512 \end{bmatrix}{\times}4$	$\begin{bmatrix} 1{\times}1,\,128 \\ 3{\times}3,\,128 \\ 1{\times}1,\,512 \end{bmatrix}{\times}8$
conv4_x	14×14	$\begin{bmatrix} 3{\times}3,\,256 \\ 3{\times}3,\,256 \end{bmatrix}{\times}2$	$\begin{bmatrix} 3{\times}3,\,256 \\ 3{\times}3,\,256 \end{bmatrix}{\times}6$	$\begin{bmatrix} 1{\times}1,\,256 \\ 3{\times}3,\,256 \\ 1{\times}1,\,1024 \end{bmatrix}{\times}6$	$\begin{bmatrix} 1{\times}1,\,256 \\ 3{\times}3,\,256 \\ 1{\times}1,\,1024 \end{bmatrix}{\times}23$	$\begin{bmatrix} 1{\times}1,\,256 \\ 3{\times}3,\,256 \\ 1{\times}1,\,1024 \end{bmatrix}{\times}36$
conv5_x	7×7	$\begin{bmatrix} 3{\times}3,\,512 \\ 3{\times}3,\,512 \end{bmatrix}{\times}2$	$\begin{bmatrix} 3{\times}3,\,512 \\ 3{\times}3,\,512 \end{bmatrix}{\times}3$	$\begin{bmatrix} 1{\times}1,\,512 \\ 3{\times}3,\,512 \\ 1{\times}1,\,2048 \end{bmatrix}{\times}3$	$\begin{bmatrix} 1{\times}1,\,512 \\ 3{\times}3,\,512 \\ 1{\times}1,\,2048 \end{bmatrix}{\times}3$	$\begin{bmatrix} 1{\times}1,\,512 \\ 3{\times}3,\,512 \\ 1{\times}1,\,2048 \end{bmatrix}{\times}3$
	1×1	average pool, 1000-d fc, softmax				

Source: He *et al.* (2016).

The study conducted by Espíndola *et al.* (2023a) examined various classification architectures and determined that ResNet50, despite its low floating-point operations per second, exhibited superior performance for identifying multiple pavement distress. Nevertheless, the study highlighted the necessity of augmenting the quantity of images and broadening the range of datasets used to train the classifier. This will ensure that the classifier accurately reflects real-world road conditions and enhances the effectiveness of the pavement management system on a network-wide scale.

5.3 Methods

The project aims to provide an economical automated method for cataloguing different road components, such as pavement type, vertical signage, and pavement defects. Each study component is provided with a module for the inventory system. This module utilizes the same data source, which is road videos or images, and employs a convolutional neural network (CNN) to develop classification or object detection models. These models are then used to invent the road components. Figure 5 displays the workflow of the system. To commence, images are collected using GPS-enabled cameras, such as GoPro, Garmin, or even mobile devices. Subsequently, the latitude and longitude data are retrieved to facilitate the production of inventory maps. Lastly, there exist automated identification models. The pavement evaluation module depends on an inventory of pavement defects generated by automated classification.

Figure 5 – Inventory system working flow
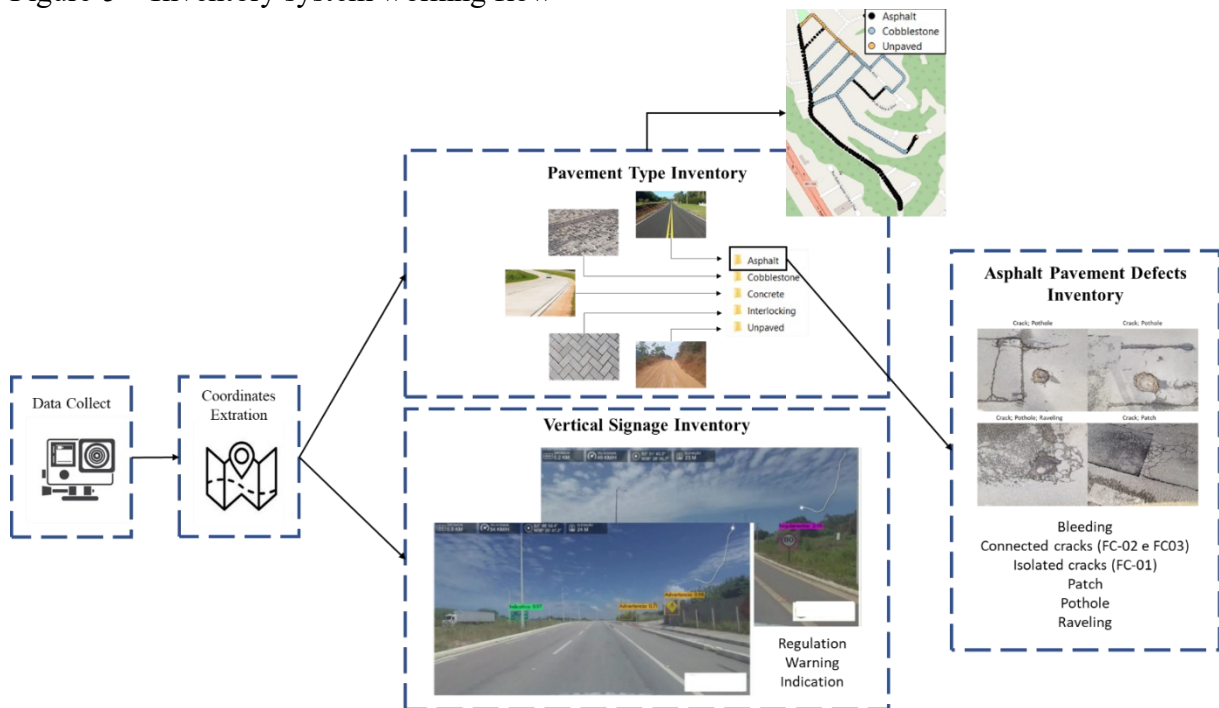
Source: elaborated by the author.

The distress inventory module was developed by adhering to a specific sequence of processes: Collection of data using cameras attached to vehicles, collecting videos and images of highways and streets. Video data is processed by extracting frames and converting them into images. Additionally, latitude and longitude information are extracted from images using GPS data, if it is provided. Processing and organizing dataset for treatments and grouping, labelling distresses on images. Preparing a collection of randomly selected images for training a neural network; Training a convolutional neural network using transfer learning with a pre-trained architecture, specifically ResNet50, is performed. The most efficient model is selected during the validation process. The model is then applied in the test section. Finally, a pavement distress inventory map is generated. Figure 6 displays the training flowchart of the Convolutional Neural Network (CNN) model for multi-label classification of pavement distress.

Figure 6 – CNN Training Flowchart
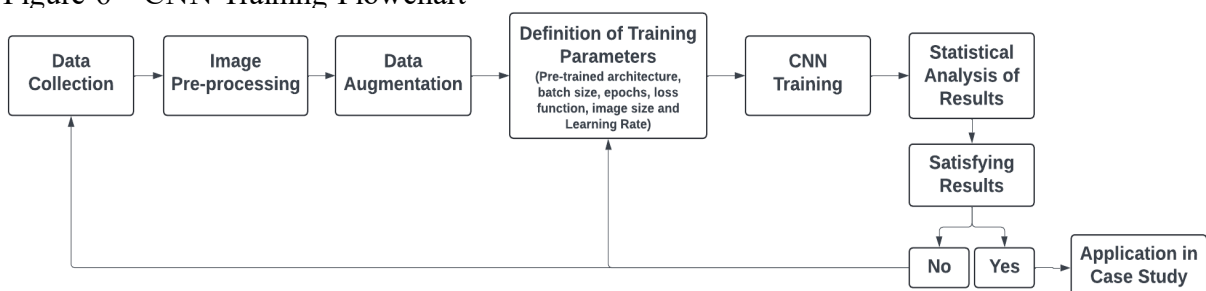
Source: elaborated by the author.

5.3.1 Data collection

The initial findings reported in Espíndola *et al.* (2023a,b) demonstrate that employing a model trained exclusively on images from a single dataset is ill-advised. These studies have demonstrated that extrapolated applications may yield inaccurate predictions or fail to produce any label as an outcome. Furthermore, the incorporation of diverse datasets enables the classification algorithm to be generalized, simplifies its implementation, and assists in its integration into network-level PMS.

The dataset for model development has been expanded based on the acquired results. A total of six datasets were utilized, sourced from various geographical regions, camera types, vehicle positions, and viewing angles. Figure 7 displays the vehicles employed for image collecting in each dataset. The images in dataset A, B, and C were acquired using a GoPro and Garmin camera, which cost less than $300 and included the camera itself, a protective case, and a suction holder.

Figure 7 – Camera positioning in the vehicle, dataset A, B, C, D, E and F
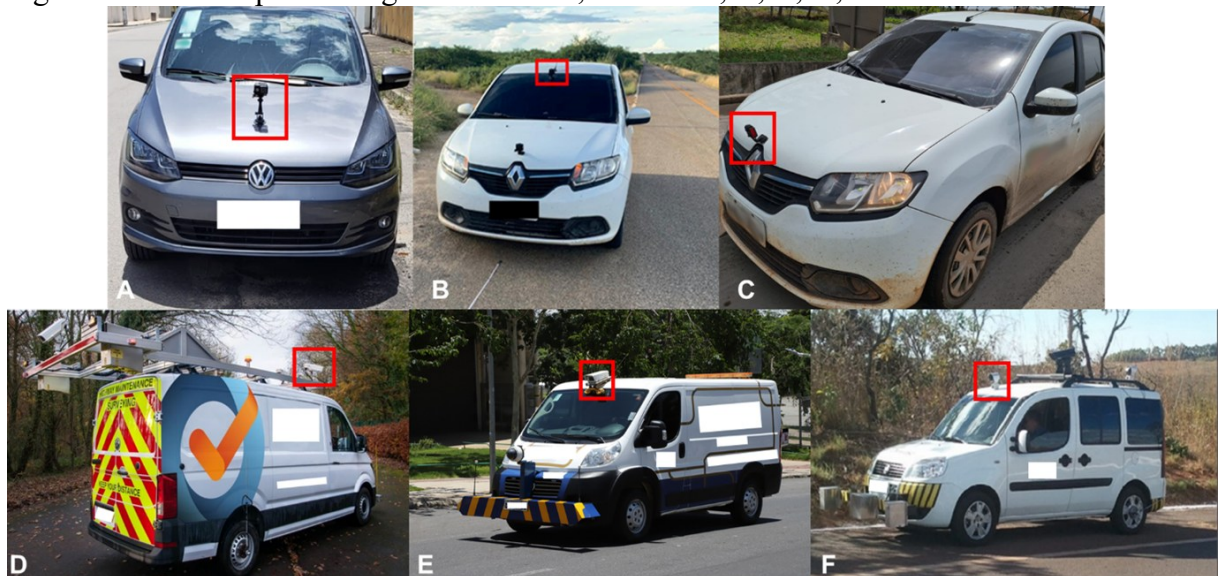
Source: elaborated by the author.

Three companies supplied images for the study and furnished examples of the Functional Evaluation Vehicle employed. The positions of the cameras that utilized images were also determined (Figure 7 D, E and F). However, the exact cameras employed were not specified. The Functional Evaluation Vehicle not only records images but also collects data on quality indicators such as the international roughness index (IRI), rutting, and cracking area. Nevertheless, those data were likewise not obtainable.

It sought to uphold the velocity of the vehicle within the range of 40 km/h to 70 km/h, constantly adhering to the prescribed speed limits on highways. The images were recorded during the day and in good weather conditions, as requested by DNIT (2003). Also encompass a range of scenarios with varying lighting conditions and shadows cast by items located at the periphery or on the sidewalks of the streets. This is demonstrated in the mosaic displayed in Figure 8, which exhibits an example of acquired images from each dataset.

Figure 8 – Examples of images collected, dataset A, B, C, D, E and F
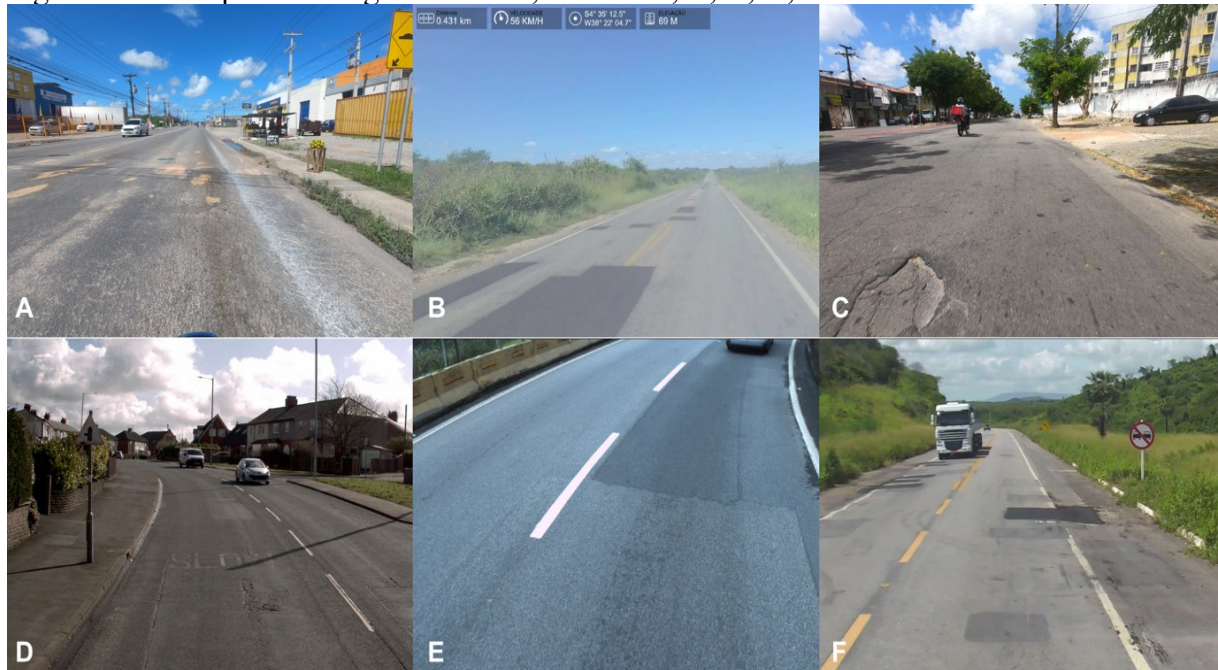
Source: elaborated by the author.

All datasets were collected using the JPG image format. However, only dataset A contains the geolocalization information necessary for georeferencing pavement defects. The data collection process was conducted during the research study with the explicit purpose of creating two datasets: Dataset A, which consisted of data obtained from low-cost cameras to include in the dataset for training CNN, and the dataset for the study case.

5.3.2 Image pre-processing

Data augmentation is a technique used to enhance the number of samples (images) in datasets, which aids in the learning process of Convolutional Neural Networks (CNNs). The train model methodology proposed in Espíndola *et al.* (2023b) focuses on identifying defects in image fragments rather than the entire image. Consequently, the dataset was retained using

this methodology. Each dataset contains images of various dimensions, necessitating distinct pre-processing for each.

The pre-processing stage of the study images involved following the sequence indicated in Figure 9 to perform image cropping and after resizing. Figure 9 illustrates the pre-processing of dataset E, where the four parts (1000x1000 pixels) at the bottom of the image are cropped. This allows the model to focus on and analyze areas predominantly composed of pavement. The dimensions of each image's parts were reduced to 300x300 pixels.

Figure 9 – Pre-processing images of Dataset E: Crop and Resize
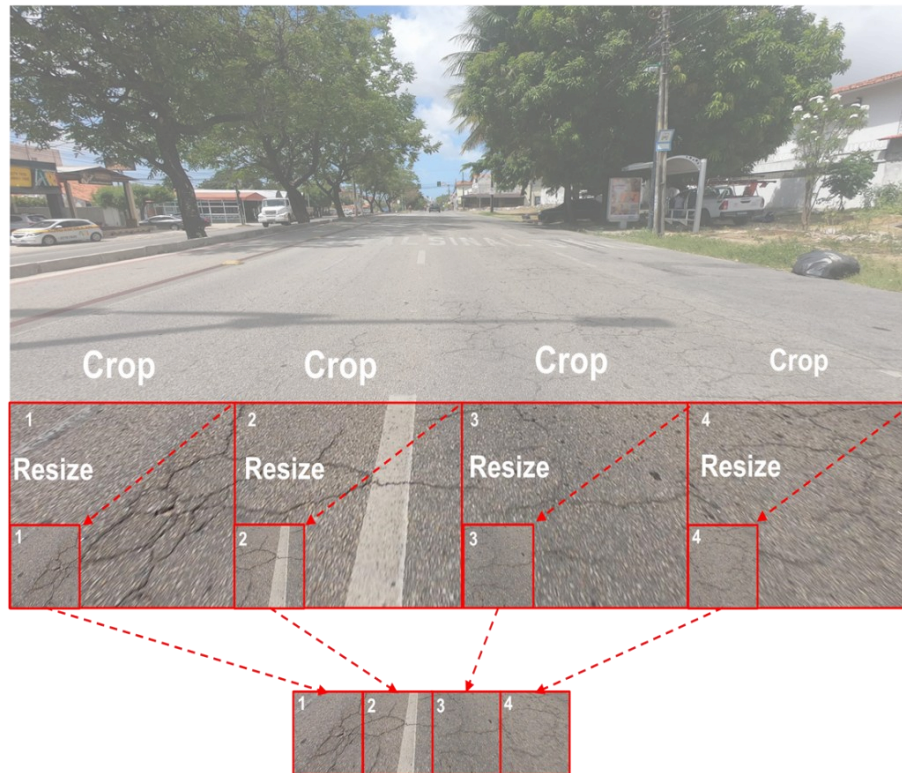
Source: elaborated by the author.

Table 2 displays the dimensions of the original images for each dataset, as well as the sizes and quantities of the cropped parts. All parts were subsequently resized to the same size (300x300), known as the model input size. It is important to emphasize that the cropping and resizing were executed automatically via code. The only activity that necessitated manual labor was to annotate the pavement distress in the images.

Table 2 – Summary of dataset preprocessing

Dataset	Original Size of Image	Pre Processar	
		Crop	Resize
A	1920x1080	Crop in 3 parts of 640x640	300x300
B	3840x2160	Crop in 4 parts of 960x960	300x300
C	1600x1200	Crop in 3 parts of 533x533	300x300
D	1280x720	Crop in 3 parts of 426x426	300x300
E	4000x3000	Crop in 4 parts of 1000x1000	300x300
F	1600x1200	Crop in 3 parts of 533x533	300x300

Source: elaborated by the author.

5.3.3 Training and testing

Training involves the utilization of mathematical optimization, employing a loss function to automatically adjust the weights and biases of the neural network. The optimization technique utilizes the training data, which includes predicted labels, to identify the global minimum of the loss function. The pre-trained architecture ResNet50 is utilized for developing a Convolutional Neural Network (CNN) due to its previous success in identifying pavement distress, as demonstrated by Gopalakrishnan *et al.* (2017), Ranjbar, Nejad and Zakeri (2020), Espíndola *et al.* (2023b), and Sirhan, Bekhor and Sidess (2024). The model includes 6 different classes of defects, namely: Bleeding, Connected Crack, Isolated Crack, Patch, Pothole and Raveling. The datasets utilized for the training and evaluation of the models are provided in Table 3. The dataset was partitioned, with 80% of the images allocated to the training set (32,304 images) and 20% to the validation set (8,076 images). The image allocation was done randomly.

Table 3 – Datasets Summary

Dataset	Quantity							
	Images Total	Distress Total	Connected Crack	Isolated Crack	Patch	Pothole	Raveling	Bleeding
A	6,300	8,752	257	477	3,815	3,528	73	602
	20.10%	138.92%	2.94%	5.45%	43.59%	40.31%	0.83%	6.88%
B	2,500	3,099	412	1,292	723	188	476	8
	3.90%	123.96%	13.29%	41.69%	23.33%	6.07%	15.36%	0.26%
C	10,000	13,990	422	6,335	4,444	138	2,011	640
	24.76%	139.90%	3.02%	45.28%	31.77%	0.99%	14.37%	4.57%
D	5,100	4,172	404	676	544	348	464	1,736
	12.63%	81.80%	9.68%	16.20%	13.04%	8.34%	11.12%	41.61%
E	10,120	12,410	1,062	4,098	3,626	1,526	1.168	930
	27.10%	122.63%	8.56%	33.02%	29.22%	12.30%	9.41%	7.49%
F	6,360	7,424	296	2,852	1,080	16	3,144	36
	9.80%	116.73%	3.99%	38.42%	14.55%	0.22%	42.35%	0.48%
All Together	**40,380**	**49,847**	**2,853**	**15,730**	**14,232**	**5,744**	**7,336**	**3,952**
	100.00%	**123.44%**	**5.72%**	**31.56%**	**28.55%**	**11.52%**	**14.72%**	**7.93%**

Source: elaborated by the author.

These are the defects that have the highest quantity of samples (images). Approximately 30% of the images exhibited Isolated Crack and Patch, while just over 12% of the images had pothole and raveling. The lowest performance was observed when bleeding was present in just approximately 8% of the images. Despite having only, a 5.7% presence in the dataset images, connected crack yielded somewhat superior results compared to bleeding. Suggesting that enhancing the database of these defects could potentially offer opportunities for improvement.

5.3.4 Parameter initialization and performance metrics

The input parameters employed were a batch size of 32, 25 epochs, cross-entropy as the loss function, and an image size of 300x300 pixels for all ResNet50-based models. The batch sizes that were evaluated comprised 64, 32, and 16. Decreasing the batch size from 64 to 32 led to enhanced metrics. However, decreasing it further to 16 had a negligible effect on the measures and resulted in an escalation in training duration. Consequently, a batch size of 32 was chosen. The training process commenced with 10, 20, 25 and 30 epochs, and the loss values were observed at the end of each set of epochs. The results demonstrated consistent performance after 25 epochs, after this number of epochs, overfitting begins to occur, leading to the decision to choose this value for the final training.

Saponara and Elhanashi (2022) investigates the influence of image resizing on two key aspects: the length of model training and its overall performance. Small images accelerate the training process of deep learning models, leading to faster conclusions. When the size of the input image is increased, the neural network must handle a larger number of pixels, resulting in extended training durations for the design. The images evaluated during the initial training stage had dimensions of 224x224, 300x300, and 512x512. Utilizing models trained with images of dimensions 512x512 led to a substantial increase in the time of the training process. The utilization of a 224x224 pixel dimension resulted in a decline in the model's quality. Conversely, the utilization of 300x300 pixel images resulted in moderate outcomes. Although there was no substantial increase in the training duration, there was an enhancement in the accuracy of model predictions.

Accuracy is the quantification of correctly detected instances. It is primarily employed when all the classes have equal significance. The F1-score is a statistical measure that combines precision and recall using the harmonic mean. It provides a more accurate assessment of misclassified situations compared to the accuracy metric. Precision quantifies the

proportion of accurately detected positive cases out of all the projected positive cases, making it particularly valuable in situations when the consequences of false positives are significant. Conversely, recall quantifies the measures the correct identification of positive cases out of all the actual positive cases. It is particularly important when the consequences of false negatives cases are significant. Recall is commonly evaluated using metrics such as % accuracy and F1-score in a macro-level analysis. The computation involves averaging the F1-scores for each label, considering that there are several labels per pattern and a contingency table for each label. The macro technique calculates a single performance indicator for each label and then averages the numbers across all categories.

5.3.5 Parameters optimization

Depending on CNN's data, the weights require modification. The last layers of the model frequently require significant adjustments to the parameters. Conversely, more profound layers that have already been well trained in identifying fundamental characteristics (such as edges and contours) require more subtle adjustments. Transfer learning was employed in network design, resulting in the fine-tuning of CNNs in two stages. In the initial stage, the first layer group is freezing, preventing any updates to its parameters. Subsequently, the last layer groups undergo training for a total of 25 epochs. During the second stage, the initial layer group is thawed, and the training is extended for a further 25 epochs, starting from the prior training state. Basically, the first layer group is not trained until the subsequent layer groups have been optimized. By fine-tuning the final layers of the pre-trained models, this technique fully optimizes the benefits of transfer learning.

The cyclical learning rates (CLR) method, often known as "one-cycle" training, is employed to determine the optimal Learning Rate (LR). The CLR approach is utilized to generate a graph that displays the validation loss in relation to the LR. This graph is employed to identify the LR values just before a significant increase in validation loss occurs. The LR values can either be fixed or within a range. For Stage 1, the optimal fixed values are around 10^{-3}, For Stage 2, the LR range should be between 10^{-6} to 10^{-3}. A CLR was conducted for all cases and across all pre-trained models. Due to the minimal variation in the graphs, it was determined to maintain similar LR values for all cases investigated. Following the completion of model training, the graphs depicting the relationship between Loss and Batches processed (Figure 10) were examined to assess and prevent overfitting.

Figure 10 – Learning Curve – Loss vs. batches processed

(a) Training CNN - Stage 1 (b) Training CNN - Stage 2
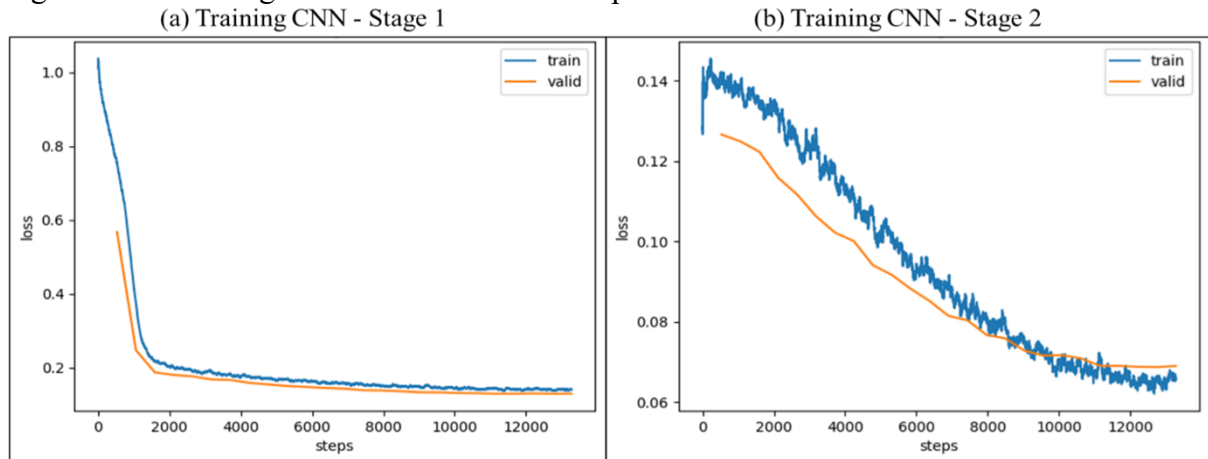
Source: elaborated by the author.

5.4 Results and discussion

This topic contains results and analysis of the model that shown the higher performance in the metrics, while contrasting predictions of defects recognized in the image by the CNN to the defects ground truth.

It is important to note that the model was trained using just 80% of the images, while the remaining 20% were utilized for quality analysis of the model generated. The specifics of the validation set, including the default numbers of images and how it is organized by dataset, are displayed in Table 4. The neural network did not encounter these specific images during its training process. However, these images share similar features with the ones used for training, such as the type of pavement (old or new), lighting conditions, image angle, interference, and objects in the right-of-way.

Table 4 – Validation Dataset

Valid Set	ValidSet – Quantity							
	Images Total	Distress Total	Connected Crack	Isolated Crack	Patch	Pothole	Raveling	Bleeding
A	1,236	1,730	57	101	754	691	10	117
	15.30%	139.97%	3.29%	5.84%	43.58%	39.94%	0.58%	6.76%
B	501	614	85	252	147	34	94	2
	6.20%	122.55%	13.84%	41.04%	23.94%	5.54%	15.31%	0.33%
C	1,979	2,735	90	1,231	872	29	380	133
	24.50%	138.20%	3.29%	45.01%	31.88%	1.06%	13.89%	4.86%
D	1,044	866	91	133	118	78	92	354
	12.93%	82.95%	10.51%	15.36%	13.63%	9.01%	10.62%	40.88%
E	2,056	2,523	234	826	738	314	226	185
	25.46%	122.71%	9.27%	32.74%	29.25%	12.45%	8.96%	7.33%
F	1,260	1,458	52	571	210	3	618	4
	15.60%	115.71%	3.57%	39.16%	14.40%	0.21%	42.39%	0.27%
All Together	**8,076**	**9,926**	**609**	**3,114**	**2,839**	**1,149**	**1,420**	**795**
	100.00%	**122.91%**	**6.14%**	**31.37%**	**28.60%**	**11.58%**	**14.31%**	**8.01%**

Source: elaborated by the author.

The presence of imbalanced classes of defects, specifically bleeding and connected crack, can cause an impact on the results due to the limited number of images available for these defects. This imbalance is leading to substantial percentage variances, even with a modest number of errors.

The results of the comparison between ground truth and predicted, based on the metrics, are presented in Table 5. The human error rate was estimated to be 5.1% (Russakovsky *et al.*, 2015), so the goal was to attain a 95% into four metrics for all defects and datasets. Nevertheless, it was unattainable to accomplish this level of performance in every defect and in every measurement.

Table 5 – Metrics results of validation set for each dataset and pavement defect

#	Metric/Distress	Bleeding	Connected Crack	Isolated Crack	Patch	Pothole	Raveling
All Together	Accuracy	99.1%	99.7%	96.3%	97.9%	99.4%	98.9%
	Precision	99.7%	98.0%	96.1%	98.2%	97.3%	97.9%
	Recall	91.3%	97.4%	94.4%	95.9%	98.2%	95.7%
	F1 score	95.3%	97.7%	95.2%	97.0%	97.7%	96.8%
A	Accuracy	99.1%	99.7%	97.8%	97.6%	98.4%	99.8%
	Precision	100.0%	98.2%	92.0%	98.7%	97.6%	100.0%
	Recall	90.6%	94.7%	80.2%	97.3%	99.6%	80.0%
	F1 score	95.1%	96.4%	85.7%	98.0%	98.6%	88.9%
B	Accuracy	99.8%	100.0%	97.8%	99.2%	100.0%	99.6%
	Precision	100.0%	100.0%	98.0%	100.0%	100.0%	98.9%
	Recall	50.0%	100.0%	97.6%	97.3%	100.0%	98.9%
	F1 score	66.7%	100.0%	97.8%	98.6%	100.0%	98.9%
C	Accuracy	98.8%	99.6%	92.0%	97.4%	99.9%	97.2%
	Precision	99.1%	97.7%	94.1%	97.8%	100.0%	95.0%
	Recall	82.7%	94.4%	92.9%	96.3%	93.1%	90.3%
	F1 score	90.2%	96.0%	93.5%	97.1%	96.4%	92.6%
D	Accuracy	99.6%	99.9%	99.5%	99.6%	100.0%	99.6%
	Precision	99.7%	100.0%	99.2%	100.0%	100.0%	98.9%
	Recall	99.2%	98.9%	97.0%	96.6%	100.0%	96.7%
	F1 score	99.4%	99.4%	98.1%	98.3%	100.0%	97.8%
E	Accuracy	98.5%	99.2%	96.2%	96.4%	98.5%	98.9%
	Precision	100.0%	96.2%	96.5%	97.0%	95.5%	96.4%
	Recall	83.2%	97.0%	93.9%	93.0%	94.9%	93.8%
	F1 score	90.9%	96.6%	95.2%	94.9%	95.2%	95.1%
F	Accuracy	100.0%	100.0%	98.8%	99.6%	100.0%	99.6%
	Precision	100.0%	100.0%	98.8%	99.5%	100.0%	99.8%
	Recall	100.0%	100.0%	98.6%	98.1%	100.0%	99.4%
	F1 score	100.0%	100.0%	98.7%	98.8%	100.0%	99.6%

Source: elaborated by the author.

Upon analysing the metrics of all the images in the validation set, as displayed in Table 5, it is evident that the results are promising. The accuracy and precision metrics achieved results above 96% for all defects, which is comparable to the performance of an experienced professional. Nevertheless, the measures that were affected by false positives, particularly the

F1-score and Recall, consistently provided results above 91%. Demonstrating the model's potential for enhancement.

The model exhibited satisfactory performance in the following sequence of defects: connected crack, bleeding, pothole, raveling, patch, and isolated crack. By consolidating all the images into a single dataset, the results demonstrate significant promise as errors are evenly distributed throughout the entire collection. Nevertheless, it is obvious that datasets and defects with a smaller number of images still require adjustments.

In many instances, the Recall and F1-score measure yielded values below 90%, with the bleeding defect exhibiting the worse performance in datasets "B", "C", and "E". This indicates the occurrence of false positive errors. Due to the limited and unbalanced amount of images utilized for training and validation, it was anticipated that the performance would be poorer. The dataset "B" included a total of 8 images (0.26%) that exhibited bleeding. Out of these, two images were utilized for validation purposes. The dataset "C" comprised 4.57% of the total images, while the dataset "E" contained 7.49%. Due to the tiny size of both datasets, there was a higher number of errors in establishing the labels. This can be a challenge for CNN to effectively learn the distinctive features in such scenarios.

Upon studying the process of adapting the model to classify errors by comparing variations in predicted features within each dataset, the outcome deviated from the observed pattern in certain specific datasets. As anticipated, classifying minor defects proved to be more challenging, particularly in the images from dataset "A". Upon studying the pictures in this dataset, it is evident that they belong to the category of images with the greatest magnitude and poorest resolution. These characteristics may have contributed to the inaccurate detection of isolated cracks and raveling.

The occurrence of false positives consistently followed the expected pattern in all datasets. The presence of limited image quantities in datasets led to decreased recall and F1-score values. However, dataset F stands out due to its limited size, containing only 13 images of potholes in the training dataset and a meager 3 images in the validation set. In addition, the training dataset contained 32 images depicting bleeding, but the validation collection only had 4 such images. The model demonstrated impeccable performance in all metrics, precisely detecting the exact defects without any instances of false positives or false negatives.

The model exhibited satisfactory results, comparable to those achieved by people, and has the potential to be integrated into network-level pavement management models. Nevertheless, while achieving an accuracy rate of over 95%, no model is flawless. The model may exhibit errors in its labeling process, such including failure to identify defects, inaccurate

identification of defects, and identifying defects in instances where there are none, as depicted in Figure 11.

Figure 11 – Examples of images and the results of the defects pavement classification
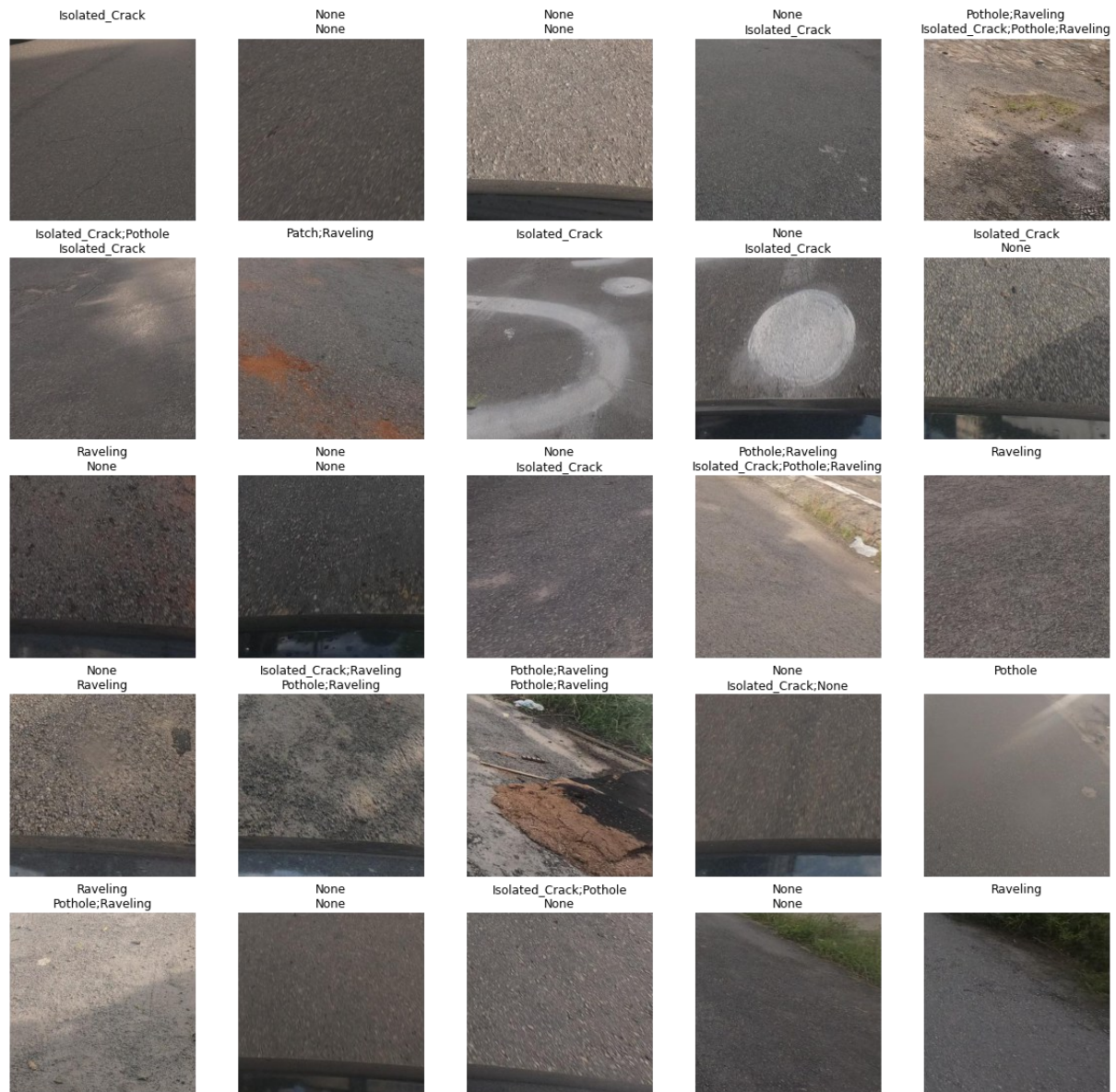
Source: elaborated by the author.

Of the total of 8,086 images in the validation set, 6,914 images were found to have at least a single type of image distress. No sample exhibited the presence of the 6 simultaneous defects that are considered by the classification model. Approximately 14.4% of the images displayed none of the 6 defects that were assessed in the study. The majority of samples (images) often contain a single defect, accounting for 54.1% of the total. Nevertheless, the presence of multiple types of defects in the samples is indicative, amounting to a combined total of 31.6%, the data is visible in the summary presented in Table 6.

Table 6 – Number of images based on the presence of multiple defects simultaneously

Quantity of simultaneous defects	6	5	4	3	2	1	0	Total
Images Quantity	0	6	27	392	2,123	4,366	1,162	8,076
%	0.0%	0.1%	0.3%	4.9%	26.3%	54.1%	14.4%	100.0%

Source: elaborated by the author.

The purpose of the proposal is to assess whether the image accurately identified which of six defects was visible in the image. Or rather, the model accurately verifies the presence or absence of each defect. The differential of the model is to verify the presence of more than one defect simultaneously into the sample.

Of the images that contained defects, an examination of the performance of simultaneous identification (Table 7) revealed that the model perfectly predicted the presence or absence of defects in 92.2% of the samples. However, 7.0% of predictions contained one labeling error, 0.7% contained two labeling errors.

Table 7 – Proportion of errors for each combination of simultaneous defects

Quantity of simultaneous defects	5	4	3	2	1	0	Total
No mistakes	6	24	320	1,850	4,166	1,082	7,448
	100.0%	88.9%	81.6%	87.1%	95.4%	93.1%	92.2%
1 mistake	0	3	66	251	172	76	568
	0.0%	11.1%	16.8%	11.8%	3.9%	6.5%	7.0%
2 mistakes	0	0	6	23	27	4	60
	0.0%	0.0%	1.5%	1.1%	0.6%	0.3%	0.7%

Source: elaborated by the author.

The current model exhibits ample potential for enhancement, including improved generalization and the ability to detect a wider range of defects in 2D images. The study aimed to detect defects in "controlled" images, which consisted of fragments of images with or without defects, to aid in the training and classification of different types of defects. Moreover, the excessive extrapolations may not yield the same accurate outcome as the presented model.

In the preliminary study conducted by Espíndola *et al.* (2023a), a total of three datasets and just over 10,000 images were analyzed, resulting in positive conclusions. However, in this following round of improving the model, three new datasets and 30,000 images were incorporated, encompassing a broader variety of image types. By integrating supplementary images, it improved the accuracy and F1-score results.

The concept is to incorporate portions of the images into subsequent training iterations when surveys are conducted, aiming to develop a highly comprehensive model that encompasses a wide range of defects, pavements, and roadways. The ongoing refinement of the

model would be intriguing, the model should be regularly updated with new highway images from various areas. These are essential for determining the minimum number of images required to update the model while keeping a high level of accuracy and F1-score.

5.5 Conclusions

The following conclusions are drawn from the research:

– It is feasible to develop a strong and computationally efficient convolutional neural network (CNN) based on multi-label classification (MLC) for the purpose of identifying and classifying various types of distress. In addition, transfer learning and cyclic learning rate (CLR) contribute to the effective performance of the models.

– The findings indicate that it is possible to accurately identify multiple defects from a single image with a reliability rate of up to 96%. Therefore, it offers substantial advantages for any automated defect identification system. This also obviates the necessity of manually cropping images to display a categorization that can only accommodate a solitary type of defect in each image. In addition, the top-performing model maintained satisfactory performance for all labels, even if there was an uneven distribution of one dis tress label in the training set.

– Using a model constructed using images from a few datasets as a foundation for diverse images is not recommended. This study has demonstrated that this expanded application may yield inaccurate forecasts or fail to produce any label as an output. Moreover, the incorporation of diverse datasets enables the classification method to be generalized, facilitates its application in various contexts, and aids in its integration into network-level PMSs.

– The utilization of MLC enables the enhancement of methods for automating pavement evaluations to produce reports on estimated pavement condition and identify the specific sections of the roadway that require pavement assessments to be conducted at the project level. Furthermore, doing on-site assessment or implementing segmentation or object identification models is only necessary for specific segments, resulting in decreased financial and time expenses for network-level PMS.

5.6 References

AN, Kwang Eun *et al.* Detecting a pothole using deep convolutional neural network models for an adaptive shock observing in a vehicle driving. **2018 IEEE International Conference on Consumer Electronics, ICCE 2018**, [*s. l.*], 2018.

APARNA *et al.* Convolutional neural networks based potholes detection using thermal imaging. **Journal of King Saud University – Computer and Information Sciences**, [*s. l.*], 2019. Available from: https://doi.org/10.1016/j.jksuci.2019.02.004. Access in: 15 June 2024.

ASCE. **2021 Report Card for America's Infrastructure**. [*s. l.*], 2021. Available from: https://infrastructurereportcard.org/. Access in: 15 June 2024.

ASLAN, O. D. *et al.* Using artifical intelligence for automating pavement condition assessment. **International Conference on Smart Infrastructure and Construction 2019, ICSIC 2019: Driving Data-Informed Decision-Making**, [*s. l.*], v. 2019, n. 2018, p. 337-341, 2019.

CAO, Wenming; LIU, Qifan; HE, Zhiquan. Review of pavement defect detection methods. **IEEE Access**, [*s. l.*], v. 8, p. 14531-14544, 2020.

CHEN, Fu-Chen; JAHANSHAHI, Mohammad R. NB-CNN: deep learning-based crack detection using convolutional neural network and naive bayes data fusion. **IEEE Transactions on Industrial Electronics**, [*s. l.*], v. 65, n. 5, p. 4392-4400, 2018.

CNT; SEST SENAT. **Pesquisa CNT de rodovias 2023**. Brasília: [*s. n.*], 2023.

COLEMAN, Cody *et al.* DAWNBench: an end-to-end deep learning benchmark and competition. *In*: Long Beach, CA, USA. **31st Conference on Neural Information Processing Systems**. Long Beach, CA, USA: [*s. n.*], 2017.

DADASHOVA, Bahar; DOBROVOLNY, Chiara Silvestri; TABESH, Mahmood. **Detecting pavement distresses using crowdsourced dashcam Camera ImagesSafe-D National UTC**. [*S. l.: s. n.*], 2021.

DING, Lei *et al.* MP-ResNet: Multi-path residual network for the semantic segmentation of high-resolution PolSAR images. **arXiv**, [*s. l.*], p. 1-5, 2020.

DU, Yuchuan *et al.* Pavement distress detection and classification based on YOLO network. **International Journal of Pavement Engineering**, [*s. l.*], v. 22, n. 13, p. 1659-1672, 2021. Available from: https://doi.org/10.1080/10298436.2020.1714047.

EISENBACH, Markus *et al.* How to get pavement distress detection ready for deep learning? A systematic approach. **Proceedings of the International Joint Conference on Neural Networks**, [*s. l.*], v. 2017-May, p. 2039-2047, 2017.

ESPÍNDOLA, Aline Calheiros; FREITAS, Gabriel Tavares de Melo; NOBRE JÚNIOR, Ernesto Ferreira. Pothole and patch detection on asphalt pavement using deep convolutional neural network. *In*: IBERO-LATIN-AMERICAN CONGRESS ON COMPUTATIONAL

METHODS IN ENGINEERING, 42.; PAN-AMERICAN CONGRESS ON COMPUTATIONAL MECHANICS, 3., 2021, Rio de Janeiro. **Proceedings** [...]. Rio de Janeiro: ABMEC-IACM, 2021. Available from: https://repositorio.ufc.br/handle/riufc/63680. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; NOBRE JÚNIOR, Ernesto Ferreira; SILVA JÚNIOR, Elias Teodoro da. Pavement surface type classification based on deep learning to the automatic pavement evaluation system. *In*: IBERO-LATIN-AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 42.; PAN-AMERICAN CONGRESS ON COMPUTATIONAL MECHANICS, 3., 2021, Rio de Janeiro. **Proceedings** [...]. Rio de Janeiro: ABMEC-IACM, 2021. Available from: https://repositorio.ufc.br/handle/riufc/63676. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; RAHMAN, Mujib; MATHAVAN, Senthan; NOBRE JÚNIOR, Ernesto Ferreira. Comparing different deep learning architectures as vision-based multi-label classifiers for identification of multiple distresses on asphalt pavement. **Transportation Research Record**: Journal of the Transportation Research Board, [New York], v. 2677, n. 5, p. 24-39, May 2023a. DOI: https://doi.org/10.1177/036119812227273. Available from: https://journals.sagepub.com/doi/abs/10.1177/03611981221127273. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; RAHMAN, Mujib; MATHAVAN, Senthan; NOBRE JÚNIOR, Ernesto Ferreira. Low-cost automated pavement diagnostic system to network pavement management systems. *In*: ANNUAL MEETING OF THE TRANSPORTATION REVIEW BOARD, 102., 2023, Washington, DC. **Proceedings** [...]. Washington, DC: Transportation Research, 2023b. Available from: https://annualmeeting.mytrb.org/Online ProgramArchive/Details/19122. Access in: 15 June 2024.

GHOSH, Rohit; SMADI, Omar. Automated detection and classification of pavement distresses using 3D pavement surface images and deep learning. **Transportation Research Record**, [*s. l.*], v. 2675, n. 9, p. 1359-1374, 2021.

GONG, Haitao; WANG, Feng. Pavement image data set for deep learning: a synthetic approach. **International Airfield and Highway Pavements Conference 2021**, [*s. l.*], p. 253-263, 2021.

GOPALAKRISHNAN, Kasthurirangan *et al.* Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. **Construction and Building Materials**, [*s. l.*], v. 157, p. 322-330, 2017. Available from: https://doi.org/10.1016/j.conbuildmat.2017.09.110. Access in: 15 June 2024.

GOPALAKRISHNAN, Kasthurirangan. Deep learning in data-driven pavement image analysis and automated distress detection: a review. **Data**, [*s. l.*], v. 3, n. 3, 2018.

HE, Kaiming *et al.* Deep residual learning for image recognition. *In*: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [*S. l.: s. n.*], 2016. p. 770-778. Available from: http://image-net.org/challenges/LSVRC/2015/. Access in: 15 June 2024.

HOANG, Nhat-duc; NGUYEN, Quoc-lam; TRAN, Van-duc. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. **Automation in Construction**, [*s. l.*], v. 94, n. July, p. 203-213, 2018.

IKRAM, Chaudhry Rehan. **A benchmark for evaluating deep learning based image analytics**. Oslo: University of Oslo, 2019.

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **NORMA DNIT 006/2003 - PRO**: avaliação objetiva da superfície de pavimentos flexíveis e semi-rígidos: procedimento. Rio de Janeiro: DNIT, 2003b.

KULKARNI, Amruta; PARIKH, Devi; ABBOTT, A Lynn. **Classification of faults in railway ties using computer vision and machine learning**. [*S. l.*], 2017. Available from: https://vtechworks.lib.vt.edu/handle/10919/86522. Access in: 15 June 2024.

LECLERC, Maxime *et al.* Ship classification using deep learning techniques for maritime target tracking. **2018 21st International Conference on Information Fusion, FUSION 2018**, [*s. l.*], p. 737-744, 2018.

LI, Baoxian *et al.* Automatic classification of pavement crack using deep convolutional neural network. **International Journal of Pavement Engineering**, [*s. l.*], v. 0, n. 0, p. 1-7, 2018. Available from: https://doi.org/10.1080/10298436.2018.1485917. Access in: 15 June 2024.

LI, Qingquan *et al.* FoSA: F * Seed-growing Approach for crack-line detection from pavement images. **Image and Vision Computing journal**, [*s. l.*], v. 29, n. 12, p. 861-872, 2011. Available from: https://doi.org/10.1016/j.imavis.2011.10.003. Access in: 15 June 2024.

LI, Yishun *et al.* RoadID: a dedicated deep convolutional neural network for multipavement distress detection. **Journal of Transportation Engineering, Part B**: Pavements, [*s. l.*], v. 147, n. 4, p. 1-12, 2021.

MA, Na *et al.* Wheat seed detection and counting method based on improved YOLOv8 Model. **Sensors**, [*s. l.*], v. 24, n. 5, 2024.

NHAT-DUC, Hoang; NGUYEN, Quoc Lam; TRAN, Van Duc. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. **Automation in Construction**, [*s. l.*], v. 94, n. July, p. 203-213, 2018.

NIE, Mingxin; WANG, Kun. Pavement distress detection based on transfer learning. **2018 5th International Conference on Systems and Informatics, ICSAI 2018**, [*s. l.*], n. Icsai, p. 435-439, 2018.

O'MAHONY, Niall *et al.* Deep Learning vs. Traditional Computer Vision. **Advances in Intelligent Systems and Computing**. [*s. l.*], v. 943, n. Cv, p. 128-144, 2019.

PAULYA, Leo *et al.* Deeper networks for pavement crack detection. *In*: Taipei, Taiwan. **International Symposium on Automation and Robotics in Construction**. Taipei, Taiwan: [*s. n.*], 2017. p. 479-485.

PEREIRA, Vosco *et al.* Classification of paved and unpaved road image using convolutional neural network for road condition inspection system. **ICAICTA 2018 – 5th International Conference on Advanced Informatics: Concepts Theory and Applications**, [*s. l.*], p. 165-169, 2018.

RANJBAR, Sajad; NEJAD, Fereidoon Moghadas; ZAKERI, H. An image-based system for pavement crack evaluation using transfer learning and wavelet transform. **International Journal of Pavement Research and Technology**, [*s. l.*], p. 1-13, 2020.

RUSSAKOVSKY, Olga *et al.* ImageNet large scale visual recognition challenge. **International Journal of Computer Vision**, [*s. l.*], p. 211-252, 2015.

SAPONARA, Sergio; ELHANASHI, Abdussalam. Impact of image resizing on deep learning detectors for training time and model performance. *In*: SAPONARA, S.; DE GLORIA, A. (ed.). **Applications in electronics pervading industry, environment and society**. ApplePies 2021. Lecture Notes in Electrical Engineering, [*s. l.*], v. 866, 2022.

SIRHAN, Mai; BEKHOR, Shlomo; SIDESS, Arieh. Multilabel CNN Model for Asphalt Distress Classification. **Journal of Computing in Civil Engineering**, [*s. l.*], v. 38, n. 1, p. 1-7, 2024.

SONG, Liang; WANG, Xuancang. Faster region convolutional neural network for automated pavement distress detection. **Road Materials and Pavement Design**, [*s. l.*], v. 22, n. 1, p. 23-41, 2021. Available from: https://doi.org/14680629.2019.1614969. Access in: 15 June 2024.

SRINIDHI, G.; DEVI, Renuka. **Pothole Detection using CNN and AlexNet**. [*S. l.*], 2020.

WANG, Xianglong; HU, Zhaozheng. Grid-based pavement crack analysis using deep learning. **2017 4th International Conference on Transportation Information and Safety, ICTIS 2017 – Proceedings**, [*s. l.*], p. 917-924, 2017.

XIA, Wei. An approach for extracting road pavement disease from HD camera videos by deep convolutional networks. **ICALIP 2018 – 6th International Conference on Audio, Language and Image Processing**, [*s. l.*], p. 418-422, 2018.

YE, Wanli *et al.* Convolutional neural network for pothole detection in asphalt pavement. **Road Materials and Pavement Design**, [*s. l.*], v. 0, n. 0, p. 1-17, 2019. Available from: https://doi.org/14680629.2019.1615533. Access in: 15 June 2024.

YUSOF, N. A. M. *et al.* Crack detection and classification in asphalt pavement images using deep convolution neural network. **Proceedings – 8th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2018**, [*s. l.*], p. 227-232, 2018.

ZHANG, Lei *et al.* Road crack detection using deep convolutional neural network. **Proceedings – International Conference on Image Processing, ICIP**, [*s. l.*], v. 2016-Augus, p. 3708-3712, 2016.

ZHOU, Shanglian; SONG, Wei. Deep learning–based roadway crack classification with heterogeneous image data fusion. **Structural Health Monitoring**, [*s. l.*], 2020.

ZHOU, Zhi Hua *et al.* Multi-instance multi-label learning. **Artificial Intelligence**, [*s. l.*], v. 176, n. 1, p. 2291-2320, 2012. Available from: https://doi.org/10.1016/j.artint.2011.10.002. Access in: 15 June 2024.

6 MANUSCRIPT PAPER 03: PROPOSAL OF SIMPLIFIED PAVEMENT STATE INDEX (SPS) TO PAVEMENT QUALITY LOW-COST DIAGNOSTICS

Abstract

The effectiveness of Pavement Management Systems (PMS) and the maintenance of the pavement quality standard depend significantly on the continuous and extensive examination of the pavement surface of a road network. By assessing the degree of deterioration in the pavement, it is feasible to determine the most urgent areas for intervention and conduct out repairs in a cost-effective manner, thereby prolonging the useful lifetime of the roadways while minimizing restoration expenses. Nevertheless, the costs associated with the evaluation may make it impractical to implement the PMS. This is due to the need for a significant number of qualified workers, expensive equipment, and a lengthy execution process. Another option is to lean to automated pavement assessments, which involve using road images and implementing a Convolutional Neural Network to categorize issues in the image. However, affordable technology, such as 2D images, is unable to capture the complete set of data regarding the current pavement distresses, such as rutting. Only specialized equipment, such as a laser profilometer, LiDAR, or a high-tech camera, allows for evaluations to be conducted quickly and efficiently. However, at the network level, the primary focus is to obtain a comprehensive overview of the condition of the road network, highlighting the segments or roads that are deteriorating the most, allocating priority to necessary actions, and determining the required budget. The proposal in this study is to create a simplified index called the Simplified Pavement State Index (SPS) for assessing the quality of pavement. This index is going to employ 2D images taken with low-cost commercial cameras to identify pavement defects. The index will be based on the Global Severity Index (acronym in Portuguese is IGG – *Índice Gravidade Global*) standard set by the National Department of Transport Infrastructure (DNIT/Brazil). When utilizing the SPS and IGG, the prioritization of segments for intervention showed a high degree of similarity, demonstrating that the proposed index effectively fulfils a Performance Measurement System (PMS).

Keywords: Defect Identification, PMS, Simplified Pavement State Index.

6.1 Introduction

The economic development of a country is directly related to Transport infrastructure because it is responsible for the integration between regions inside the country and internationally, allowing the movement of people and cargo. Therefore, the management of this critical asset must be adequate to allow progress, mainly in countries with a transport matrix focused on highways.

Nevertheless, due to limited resources available for maintenance versus the tremendous demand for investments, infrastructure suffers from continuous degradation of roads. For example, according to data from the Management Report of the National Transport Confederation (CNT) 2023, 67.5% of Brazilian highways managed by public management have

unsatisfactory conditions, showing deficiencies in geometry, signage, or pavements (CNT; SEST SENAT, 2023).

In situations of limited resources for maintenance, an alternative to reverse and improve the performance of pavements is the implementation of Pavement Management Systems (PMS). The PMS has been used for years to support managers in making decisions to maintain and extend the useful life of pavements and to plan interventions and necessary resources in the short, medium, and long term. Pavement management aims to efficiently apply financial resources, providing users with comfortable, safe, and economical pavements (Fernandes Junior, 2001). In other words, it allows for a better allocation of resources, following a prioritization system through the pavement evaluation and minimizing maintenance and rehabilitation costs. The basis for an effective PMS is to have reliable information about the actual condition of the pavement, to diagnose the pavement's state, and to identify critical segments and defects in the road network.

The pavement assessment can be through three stages, manual, semi-automatic, and automatic. However, manual inspection is becoming impractical due to the required scale and frequency, its subjectivity, high financial costs, time, and human resources (Gopalakrishnan, 2018). Semi-automatic assessment realized acquisition of pavement images and location information synchronous way, over time breakthroughs were made as techniques based on high-speed line scanning digital camera (Zalama; Jaime; Medina, 2014). However, these techniques need manual distress identification through indoor interpretation devices. Also, automatic inspection systems were promoted that brought convenience to realize high-precision and quick distress inspection; however, those used expensive sensors like laser scanners (Coenen; Golroo, 2017).

Those techniques can obtain accurate data, but it is challenging to use in high-frequency data acquisition due to complex operational and financial costs. Mainly in small and medium-sized cities or even developing countries because they have restricted human and financial resources for road assessment activities, which makes implementing the PMS unfeasible (Causim, 2001).

These algorithms can complete the work of feature extraction and object identification through the automatic learning of many samples.

In recent decades, technological progress has improved automatic systems based on image analysis to extract information and features from the images through the automatic learning of a large number of samples (Cha; Choi; Büyüköztürk, 2017). Several studies of different methods of automatic identification of damage in pavements based on computer vision

can be found in the literature. Methods such as image processing (intensity-thresholding, edge detection, and seed-based) and machine learning (unsupervised learning, supervised learning, and deep learning) have been widely used for defect identification and quantification (Li *et al.*, 2011).

The multi-defect classification model of pavement developed by Espíndola *et al.* (2023) identified four different types of defects: bleeding, crack, Patch and pothole. Furthermore, they analysed some of the Convolutional Neural Network (CNN) architectures and found that the model based on ResNet50 obtained the best performance results to predict pavement defects, presenting its optimization parameters and the necessary pre-processing in the images. In this present paper, the applicability of the model developed by Espíndola *et al.* (2023) will be tested and analysed an uncontrolled environment, in urban streets with characteristics different from the original model, using low-cost equipment to obtain the data and georeferencing the model results to facilitate the visualization of the sites that need rehabilitation and which are the most critical segments.

6.2 Literature review

The DNIT 005/2003 (IPR; DNIT, 2003a) standardizes the language and coding used to identification pavement defect, as demonstrated in Tables 1 and 2.

Table 1 – Summary table of cracks defects – coding and classification

Cracks				Code	Crack Class		
Fissure				**FI**	**-**	**-**	**-**
Cracks in the asphalt layer caused by excessive permanent deformation and/or due to fatigue	Isolated Cracks	Transversal	Short	TTC	FC-1	FC-2	FC-3
			Long	TTL	FC-1	FC-2	FC-3
		Longitudinal	Short	TLC	FC-1	FC-2	FC-3
			Long	TLL	FC-1	FC-2	FC-3
	Connected Cracks	Alligator	No erosion on crack edges	J	-	FC-2	-
			With erosion on crack edges	JE	-	-	FC-3
Cracks in the asphalt layer not caused by fatigue	Isolated Cracks	Due to thermal shrinkage or dissection of the base (soil-cement) or asphalt layer		TRR	FC-1	FC-2	FC-3
	Connected Cracks	Block crack	No erosion on crack edges	TB	-	FC-2	-
			With erosion on crack edges	TBE	-	-	FC-3

Source: DNIT 005/2003 (IPR; DNIT, 2003a).

Table 2 – Summary table of other defects – coding and classification

Other defects				Code
Deformation	Plastic	Local	Due to the plastic creep of one or more layers of the pavement or subgrade	ALP
		Wheel Rut		ATP
	Consolidation	Local	Due to differential consolidation occurring in pavement or subgrade layers	ALC
		Wheel Rut		ATC
Shoving/Corrugation – Transversal shoving caused by instability of the bituminous mixture constituting the asphalt layer or the base				O
Slippage (asphalt layer)				E
Bleeding of the bituminous binder in the asphalt layer				EX
Severe raveling on the surface of the asphalt layer				D
Potholes arising from breakdown of asphalt layer and sometimes from lower layers				P
Patch		Superficial Patch		RS
		Deep Patch		RP

Source: DNIT 005/2003 (IPR; DNIT, 2003a).

The defects exhibited in Tables 1 and 2 are visible in Figure 1. The defects in asphalt pavements are categorized to assess the quality of the investigated pavement surface.

Figure 1 – Types of defects found in asphalt pavements

Source: CNT (2017).
01-Fissure; 02-Transverse cracks; 3-Longitudinal cracks; 4-Alligator crack; 5-Block crack; 6-Plastic deformation; 7-Consolidation deformation; 8-Corrugation; 9-Shoving; 10-Bleeding; 11-Ravelling; 12-Pothole; 13-Patch.

Global Severity Index (acronym in Portuguese is IGG – *Índice de Gravidade Global*) is the evaluation of the pavement according to the DNIT regulations, DNIT 005/2003 (IPR; DNIT, 2003a) and DNIT 006/2003 (IPR; DNIT, 2003b), which deal with the categorization of defects and the systematic survey of deficiencies, respectively. The DNIT 006/2003-PRO norm permits the calculation of a pavement quality indicator based on an inventory of pavement distress, using sampling with a specified area as opposed to the complete

road area. Defects are inventoried within the samples by filling out the form in Figure 2. Defects are grouped into 8 categories according to the defect criticality weighting factors.

Figure 2 – Asphalt Layer Defect Inventory Form

Asphalt Layer Defect Inventory																							
Sample	OK	Isolated cracks						Connected cracks				Other Defects										Depth of Wheel Rut	
		FC-1						FC-2		FC-3		Deformation				Corrugation	Pothole	Shoving	Bleeding	Raveling	Pacth	Internal	External
		FI	TTC	TTL	TLC	TLL	TRR	J	TB	JE	TBE	ALP	ATP	ALC	ATC	O	P	E	EX	D	R	TRI (mm)	TRE (mm)
1	No	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	1	1	-	-
2	No	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-
3	No	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-
4	Yes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	Yes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	No	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-
7	No	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	1	-	-	-
8	No	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
9	No	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	No	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-
Total		5						1		1		2				3			0	6	2	0	0

Source: elaborated by the author.

After registering the defects, the weighting factors (fp) are assigned according to Table 3 for each type of distress. The weighted value is a function of the severity of the defects.

Table 3 – Value of the weighting factor to calculate the IGG

Category	Defects	Weighting Factor
1	Isolated cracks	0.2
2	Connected cracks without erosion	0.5
3	Connected cracks with erosion	0.8
4	Deformation	0.9
5	Corrugation, Shoving, Pothole and Slippage	1.0
6	Bleeding	0.5
7	Raveling	0.3
8	Patch	0.6

Source: DNIT 005/2003 (IPR; DNIT, 2003a).

For this methodology, when occurrences of categories 1, 2, and 3 are found in the same sample, only those in category 3 are considered to calculate the relative frequency in percentage and Individual Severity Index (IGI); similarly, when occurrences of categories 1 and 2 are verified in the same sample, only those in category 2 are considered. Always select the most severity.

The absolute and relative frequencies of the inventoried occurrences in the worksheet must be determined to process the collected data (calculating the individual and global severity index). The calculations are performed in the Equations 1, 2, and 3:

$$fr = \frac{fa * 100}{n}$$

(1)

Where:

fr – relative frequency.

fa – absolute frequency.

n – number of inventoried samples.

$$IGI = fr * fp \qquad (2)$$

Where:

fr – relative frequency.

fp – weighting factor.

IGI – individual severity index.

$$IGG = \Sigma\ IGI \qquad (3)$$

Where:

IGI – individual severity index.

IGG – global severity index.

The DNIT 006/2003-PRO norm presents a worksheet (Table 4) to be filled in with information on the number of samples for each defect category, the IGIs and IGG calculations and the final state of the asphalt layer.

Table 4 – Global Severity Index (IGG) Calculation Worksheet

Category	Defects	Absolute Frequency	Absolute Frequency Considered	Relative Frequency	Weighting Factor	Individual Severity Index (IGI)
1	FI, TTC, TTL, TLC, TLL, TRR	5	4	40	0,2	8
2	(FC – 2) J, TB	1	1	10	0,5	5
3	(FC – 3) JE, TBE	1	1	10	0,8	8
4	ALP, ATP, ALC, ATC	2	⨯	20	0,9	18
5	O, P, E	3	⨯	30	1	30
6	EX	0	⨯	0	0,5	0
7	D	6	⨯	60	0,3	18
8	R	2	⨯	20	0,6	12
9	The arithmetic mean of the mean values of the depths measured in mm in the TRI and	TRE= 0	TRI= 0	F= 0	1A ( )	0
					1B ( )	0
10	The arithmetic mean of the variances of the depths measured in both wheel ruts	TREv= 0	TRIv= 0	FV= 0	2A ( )	0
					2B ( )	0
TOTAL NUMBER OF SAMPLES		10	Sum of IGI = IGG =			99
1A) IGI = F x 4/3 when F ≤ 30		2A) IGI = FV when FV ≤ 50			State	Bad
1B) IGI = 40 when F > 30		2B) IGI = 50 when FV > 50				

Source: elaborated by the author.

Table 5 presents the asphalt layer quality expressed in the state as excellent, good, regular, bad, and terrible according to the IGG values.

Table 5 – State of degradation of the asphalt layer according to the IGG

State	IGG
Excellent	0 – 20
Good	21 – 40
Regular	41 – 80
Bad	81 – 160
Terrible	>160

Source: elaborated by the author.

Simplified Pavement State Index (SPS) is based on a methodology for calculating the Global Severity Index (acronym in Portuguese is IGG – *Índice de Gravidade Global*), which uses sampling, identification of defects from norm IPR and DNIT (2003a), application of weights by type and severity of defects and frequency of appearance of defects. The IGG methodology is presented below and in the method/results the simplification proposal for the (SPS) will be presented.

6.3 Methods

The sequence of operations pertaining to the implementation of the proposal for the Simplified Pavement State Index (SPS) adheres to the depicted flowchart in Figure 3. The research process commences with the identification of the specific street or avenue to be examined. Subsequently, the pavement samples are demarcated, and a comprehensive record is made of all defects present within the designated sampling area. This procedure is then replicated for subsequent samples, and the second and third stages are iterated until the entirety of the predetermined segment, as established in the initial stage, is thoroughly assessed. After the completion of the inventory, the subsequent stage involves the computation of the IGG and SPS. It is important to note that the IGG encompasses a comprehensive assessment of 20 distinct flaws, which are further categorized into 8 different types of defect classification. The SPS examines a comprehensive set of 16 flaws, categorizing them into four distinct forms of defect classification.

Figure 3 – A flowchart outlining the process for determining pavement indices
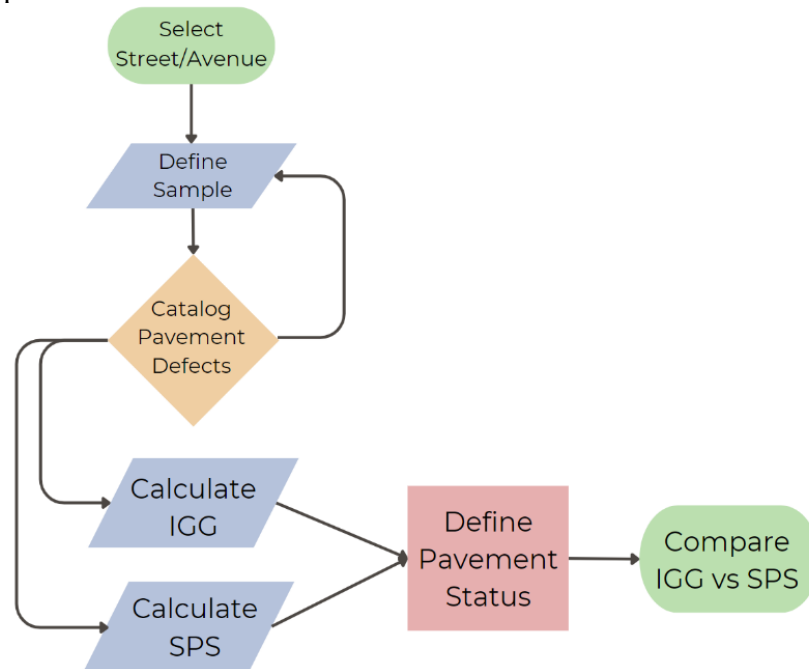
Source: elaborated by the author.

Once the IGG and SPS values have been determined, the notions of pavement quality are established, enabling the assessment of whether the indicators exhibit divergence or convergence. The process of generating road network maps using these concepts is also covered, with an emphasis on the most crucial areas that necessitate restoration and maintenance. The provision of crucial information for managerial decision-making about investment location determination.

6.3.1 Case study: Alagoas-Brazil

The transport study group at the Federal University of Alagoas (UFAL) regularly selects streets or roads in the metropolitan region of Maceió/Alagoas to provide practical training opportunities for undergraduate researchers who are enrolled in the program. The instructions primarily prioritize the development of knowledge related to several methods of pavement assessment and the capacity to distinguish and classify defects in asphalt pavement. Therefore, the inventory data analysed in this study corresponded to the data utilized in the training phase of the UFAL transport study group.

A comprehensive analysis was conducted on a total of 30 streets, avenues, and highways, all of which were single-lane roads with asphalt coating and varying hierarchy, as show in Table 6 and Figure 4. There are two highways (ID: 3 - AL-101 and ID: 22 - AL-110)

that are not encompassed within the metropolitan territory of Maceió but are as linking routes inside the inland areas of the state of Alagoas. It should be noted that certain roads were assessed multiple times between 2015 and 2023, resulting in a total of 40 inventoried segments with the procedure of DNIT 006/2003-PRO norm (IPR; DNIT, 2003b). The segments combined covered a total distance of 16.86 km, as evidenced in Table 6. During certain time intervals, the condition of streets was evaluated many times to ascertain the progression of pavement deterioration and to identify any interventions implemented by the authorities.

Table 6 – The compilation of streets and avenues inventory

Date	ID	Location	Extension (m)
May, 2016	3	AL-101	220
May, 2018	22	AL-110	520
October, 2016	4	Av. Almirante Álvaro Calheiros	300
October, 2016	9	Av. Cachoeira do Mirim	520
November, 2017	15	Av. Empresário Carlos da Silva Nogueira	520
May, 2018	23	Av. Empresário Nelson Oliveira Menezes	520
December, 2022	35	Av. Empresário Nelson Oliveira Menezes	200
February, 2022	28	Av. Eng. Paulo Brandão Nogueira	200
May, 2016	1	Av. Francisco Freire Ribeiro	200
October, 2016	8	Av. Governador Luis Cavalcante	520
October, 2016	2	Av. Governador Luis Cavalcante	220
May, 2017	11	Av. Governador Luís Cavalcante	520
November, 2017	19	Av. José Airton Gondim Lamenha	520
November, 2017	16	Av. José Haílton dos Santos	200
June, 2022	31	Av. Menino Marcelo	220
June, 2018	21	Av. Paulo Holanda	500
July, 2022	32	Av. Paulo Holanda	600
December, 2022	34	Av. Paulo Holanda	500
October, 2023	40	Av. Paulo Holanda	520
April, 2023	39	Av. Pierre Chalita	420
December, 2022	36	Av. Presidente Getúlio Vargas	500
May, 2017	14	Rua Adelaíde Melo Mota	520
February, 2022	29	Rua Desportista Cláudio da Rocha Lima	240
May, 2017	12	Rua Dona Alzira Aguiar	500
November, 2017	17	Rua Empresário Jorge Montenegro	200
February, 2022	25	Rua General Hermes	440
June, 2022	37	Rua General Hermes	520
October, 2016	5	Rua Hugo Corrêa Paes	520
October, 2016	7	Rua 01	520
December, 2022	33	Rua 01	500
April, 2023	38	Rua 01	200
May, 2017	13	Rua João Norberto de Lima	520
February, 2022	26	Rua José Henrique do Nascimento	200
July, 2022	30	Rua Paulina Maria de Mendonça	500
November, 2017	18	Rua Paulina Maria de Mendonça	500
October, 2016	6	Rua Professor Antônio Nemésio de Albuquerque	540
November, 2017	20	Rua Professor Manuel Coelho Neto	500
February, 2022	27	Rua Santos Ferraz	500
May, 2017	10	Rua Senador Firmino de Vasconcelos	500
May, 2018	24	Rua Senador Rui Palmeira	500
Total			**16,860.0**

Source: elaborated by the author.

* In 2016, two assessments were conducted on Av. Governador Luis Cavalcante, focusing on separate segments.

Figure 4 – Segments assessed within the Maceió-AL metropolitan area

Source: elaborated by the author.

6.3.2 Data collection

The data that was gathered pertained to the inventory of defects found in the 40 segments indicated in Table 32. The guidelines outlined in the DNIT 006/2003-PRO norm were adhered to, which permits the computation of a pavement quality metric by utilizing an assessment of pavement distresses. The IGG is determined on a sampling basis for samples with a predetermined area, spacing between them of 20 m, and when switching between tracks rather than for the total road area. The evaluation is specified in each sample by a 6-meter-long region, 3 meters before and 3 meters after the road stakes, as illustrated in Figure 5 (IPR; DNIT, 2003b).

Figure 5 – Demarcation of areas to register defects
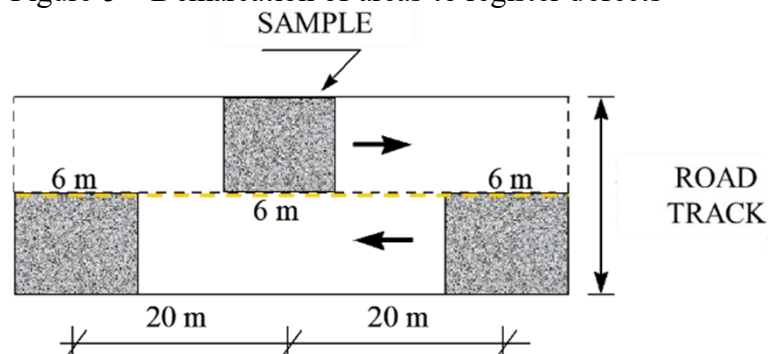
During the process of inventorying the road samples, a comprehensive photographic report was generated, documenting the many problems observed. For illustrative purposes, Figure 6 displays a compilation of recorded flaws observed along ID: 35 - Av. Empresário Nelson Oliveira Menezes in December 2022.

Figure 6 – Images from the Photographic Report of the ID assessment: 35 - Av. Empresário Nelson Oliveira Menezes. (A) Sample 5 – Connected crack (alligator crack) and patch; (B) Sample 4 – Connected crack (alligator crack) and patch; (C) Sample 7 – Raveling and patch; (D) Sample 11 – Isolated crack (longitudinal).
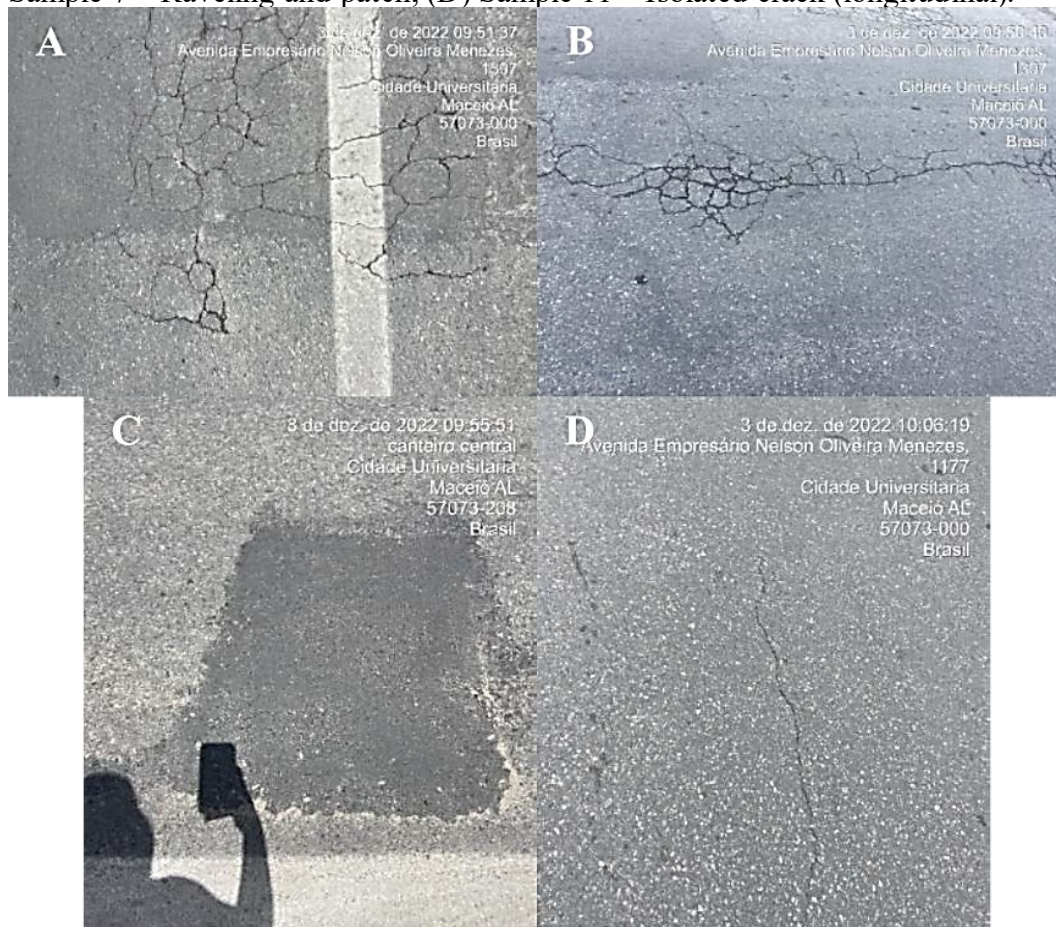
6.3.3 Proposal method for determining the Simplified Pavement State Index (SPS)

The developed deep learning model can identify 5 different types of pavement surface defects. The defects that were not included in the model were Deformation, Corrugation and Shoving due to the dataset available during the development of the model of multiple classifications of defects that did not present enough images to compose it with accuracy and sufficient precision. Raveling (was also excluded due to the model still not being able to differentiate open texture asphalt mixtures and chip seal of the pavement with raveling. Then the defects inventory table was reduced as shown in Figure 7.

Figure 7 – Reductions on pavement defects inventory

Samples	OK	Isolated cracks	Connected cracks without erosion	Connected cracks with erosion	Deformation	Corrugation and Shoving	Pothole	Bleeding	Raveling	Patch
1	No	-	1	-	1	-	-	-	1	1
2	No	1	-	-	-	1	-	-	1	-
3	No	-	-	-	-	-	1	-	-	-
4	No	1	-	-	-	-	1	-	1	-
5	Yes	-	-	-	-	-	-	-	-	-
6	No	1	-	-	-	-	1	-	1	-
7	No	1	-	-	-	-	1	-	1	1
8	No	-	-	1	-	-	-	-	1	1
9	No	1	-	-	-	-	-	-	1	-
10	No	-	-	-	1	-	-	-	-	1
Total		5	1	1	2	1	4	0	7	4

Samples	OK	Isolated cracks	Connected cracks	Pothole	Bleeding	Patch
1	No	-	1	-	-	1
2	No	1	-	-	-	-
3	No	-	-	1	-	-
4	No	1	-	1	-	-
5	Yes	-	-	-	-	-
6	No	1	-	1	-	-
7	No	1	-	1	-	1
8	No	-	1	-	-	1
9	No	1	-	-	-	-
10	No	-	-	-	-	1
Total		5	2	4	0	4

Source: elaborated by the author.

An important information to make is that the model required the junction of the types of interconnected cracks to occur. In the index that the proposal was based on, the severity level of the interconnected crack was taken into account. However, the available dataset did not show interconnected cracks with erosion (FC-03) sufficient to compose one of the model classes with accuracy and pressure during prediction. Therefore, it was decided to merge the two types of interconnected cracks (FC-2 and FC-03 with a single class), as shown in Figure 8.

Figure 8 – Adjustments to weight factor to address defect reduction

Index	Defect	Absolute frequency	Relative frequency	Weighting Factor	Individual Quality Index (IQI)
1	Isolated cracks	5	50.00	0.20	10.00
2	Connected cracks without erosion	1	10.00	0.50	5.00
3	Connected cracks with erosion	1	10.00	0.80	8.00
4	Deformation	2	20.00	0.90	18.00
6	Corrugation, Shoving and Pothole	5	50.00	1.00	50.00
7	Bleeding	0	0.00	0.50	0.00
8	Raveling	7	70.00	0.30	21.00
9	Patch	4	40.00	0.60	24.00
	Total number of samples	10	Global Severity Index (IGG)		136.00
			Concept		Bad

Defect	Absolute frequency	Relative frequency	Weighting Factor	Individual Quality Index (IQI)
Isolated cracks	5	50.00	0.20	10.00
Connected cracks	2	20.00	0.65	13.00
Pothole	4	40.00	1.00	40.00
Bleeding	0	0.00	0.50	0.00
Patch	4	40.00	0.60	24.00
Total number of samples	0		SPS	87.00
			Concept	Bad

Source: elaborated by the author.

Table 7 presents the asphalt layer quality expressed in the state as excellent, good, regular, bad, and terrible according to the SPS values. The IGG and SPS quality ranges were modified due to the omission of four defects in the original method's model.

Table 7 – Status of degradation of the asphalt layer according to the SPS

Concept	SPS limits
Excellent	0 – 15
Good	16 – 30
Regular	31 – 60
Bad	61 – 130
Terrible	>130

Source: elaborated by the author.

6.4 Results and discussion

This section presents an examination of the evaluation outcomes of the IGG and SPS, along with comparisons between the two. Furthermore, road network administrators are furnished with information in the form of maps containing a ranking of the most and least favourable segments.

Table 8 presents the calculated IGG and SPS data, along with the information about pavement quality (Figure 9) and the numerical difference between the IGG and SPS values. An analysis of the disparities in pavement quality, as measured by two indexes, showed that 67.5% of the observations in both the IGG and SPS concepts remained constant. In addition, 15.0% of the observations indicated an improvement in status position, whereas 17.5% indicated a decline in status position.

Table 8 – IGG and SPS summary results

ID	Location	IGG	Status – IGG	SPS	Status – SPS	Disparity Among Index
1	Av. Francisco Freire Ribeiro	115.0	Bad	61.50	Bad	53.50
2	Av. Governador Luis Cavalcante	145.77	Bad	120.58	Bad	25.19
3	AL-101	100.91	Bad	63.18	Bad	37.73
4	Av. Almirante Álvaro Calheiros	75.33	Regular	63.33	Bad	12.00
5	Rua Hugo Corrêa Paes	129.00	Bad	88.50	Bad	40.50
6	Rua Prof. Antônio Nemésio de Albuquerque	58.80	Regular	41.40	Regular	17.40
7	Rua 01	137.50	Bad	93.75	Bad	43.75
8	Av. Governador Luis Cavalcante	74.80	Regular	38.80	Regular	36.00
9	Av. Cachoeira do Mirim	117.69	Bad	88.85	Bad	28.85
10	Rua Senador Firmino de Vasconcelos	27.62	Good	11.43	Excellent	16.19
11	Av. Governador Luis Cavalcante	140.40	Bad	94.20	Bad	46.20
12	Rua Dona Alzira Aguiar	44.40	Regular	37.80	Regular	6.60
13	Rua João Norberto de Lima	32.33	Good	21.33	Good	11.00
14	Rua Adelaí de Melo Mota	46.92	Regular	19.81	Good	27.12
15	Av. Empresário Carlos da Silva Nogueira	43.33	Regular	30.93	Regular	12.41
16	Av. José Haílton dos Santos	73.46	Regular	24.23	Good	49.23
17	Rua Empresário Jorge Montenegro	103.64	Bad	69.09	Bad	34.55
18	Rua Paulina Maria de Mendonça	120.00	Bad	105.00	Bad	15.00
19	Av. José Airton Gondim Lamenha	105.77	Bad	73.65	Bad	32.12
20	Rua Professor Manuel Coelho Neto	46.15	Regular	41.54	Regular	4.62
21	Av. Paulo Holanda	248.00	Terrible	213.50	Terrible	34.50
22	AL-110	19.09	Excellent	20.45	Good	-1.36
23	Av. Empresário Nelson Oliveira Menezes	123.08	Bad	49.23	Regular	73.85
24	Rua Senador Rui Palmeira	140.38	Bad	59.04	Regular	81.35
25	Rua General Hermes	77.00	Regular	68.00	Bad	9.00
26	Rua José Henrique do Nascimento	64.40	Regular	48.80	Regular	15.60
27	Rua Santos Ferraz	233.00	Terrible	152.50	Terrible	80.50
28	Av. Eng. Paulo Brandão Nogueira	52.31	Regular	44.81	Regular	7.50
29	Rua Desportista Cláudio da Rocha Lima	66.15	Regular	42.50	Regular	23.65
30	Rua Paulina Maria de Mendonça	37.20	Good	36.00	Regular	1.20
31	Av. Menino Marcelo	73.08	Regular	63.27	Bad	9.81
32	Av. Paulo Holanda	200.00	Terrible	162.50	Terrible	37.50
33	Rua 01	174.80	Terrible	117.20	Bad	57.60
34	Av. Paulo Holanda	86.80	Bad	58.80	Regular	28.00
35	Av. Empresário Nelson Oliveira Menezes	68.85	Regular	56.15	Regular	12.69
36	Av. Presidente Getúlio Vargas	47.60	Regular	39.20	Regular	8.40
37	Rua General Hermes	146.00	Bad	96.20	Bad	49.80
38	Rua 01	64.55	Regular	60.45	Bad	4.09
39	Av. Pierre Chalita	85.60	Bad	63.00	Bad	22.60
40	Av. Paulo Holanda	44.23	Regular	50.58	Regular	-6.35

Source: elaborated by the author.

Figure 9 – Map of pavement quality concepts using IGG and SPS methods
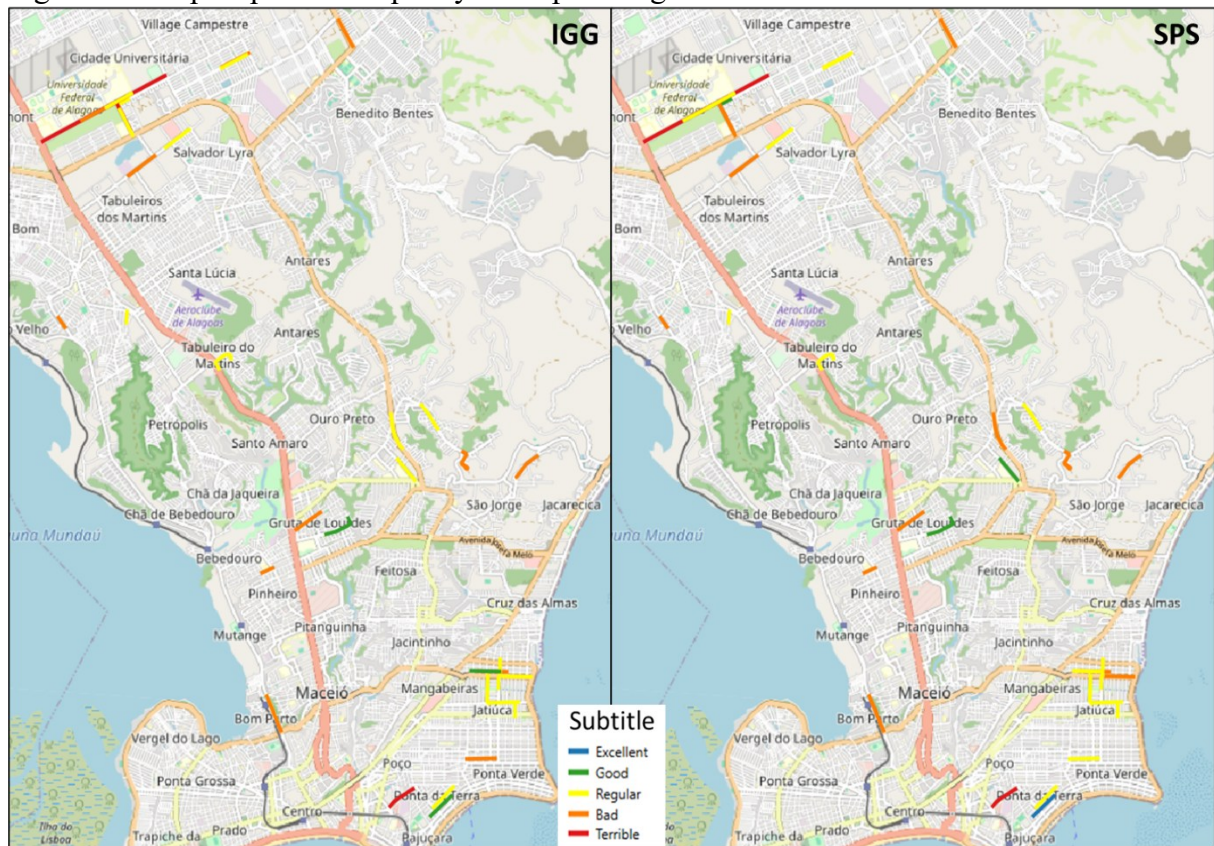
Source: elaborated by the author.

Significant variations are observed when the difference between the IGG and SPS values for each segment is analysed. The data points exhibit a range of values from minimum 1.20 to maximum 81.35, mean 27.50, and high standard deviation 21.91. The locations where the values of the two indices diverged most significantly were ID 24 - Rua Senador Rui Palmeira, ID 27 - Rua Santos Ferraz, and ID 23 - Av. Empresário Nelson Oliveira Menezes.

Segment ID 8 along Av. Governador Luis Cavalcante, consisted of twelve samples that experienced deformation. All these samples (26 in total) displayed indications of raveling, which led to an elevation in the IGG value. However, these distresses were excluded from the SPS method.

As anticipated, the SPS values exhibit significant variation for segments characterized by a substantial number of samples displaying plastic deformations, owing to their considerable weight factor (0.9). Deformation is recognized as a major factor contributing to the widening gap between the IGG and SPS indexes. To address this issue, it is feasible to coordinate pavement assessment using two-dimensional (2D) images in conjunction with data collection on pavement irregularity/roughness through smartphone-based sensors. Evaluating pavement degradation, quantified by the Surface Pavement Score (SPS), along with assessing

user comfort levels, indicated by roughness, provides additional criteria for analyzing and identifying critical segments.

The raveling defect has a significant impact on the outcomes, primarily attributed to the substantial number of samples present rather than the weight factor. The presence of the raveling defect was observed in all segments examined, and in some segments in all samples analysed, such as ID 8 - Av. Governador Luis Cavalcante and ID 40 - Av. Paulo Holanda.

By utilizing the data presented in Table 8, it becomes feasible to ascertain the segments that exhibit the most unfavourable conditions of the pavement (as indicated in Table 9), as well as those that demonstrate the most favourable conditions (as indicated in Table 10), employing both approaches. When comparing the variation of the IGG quality concepts to the SPS, it was observed that the number of paved segments in terrible condition decreased from 4 to 3, while the number of segments in excellent condition was the same. Another interesting information is that it becomes that 8 out of the 10 segments are present in both the IGG and SPS lists. This holds true for both the segments ranked as the worst and those ranked as the best in terms of pavement surface quality. Although certain segments may experience a shift in their ranking, they would remain a priority for maintenance and rehabilitation interventions.

Table 9 – The worst segments according to each index

ID	Location	IGG	Status – IGG	ID	Location	SPS	Status – SPS
21	Av. Paulo Holanda	248.0	Terrible	21	Av. Paulo Holanda	213.5	Terrible
27	Rua Santos Ferraz	233.0	Terrible	32	Av. Paulo Holanda	162.5	Terrible
32	Av. Paulo Holanda	200.0	Terrible	27	Rua Santos Ferraz	152.5	Terrible
33	Rua 01	174.8	Terrible	2	Av. Governador Luis Cavalcante	120.6	Bad
37	Rua General Hermes	146.0	Bad	33	Rua 01	117.2	Bad
2	Av. Governador Luis Cavalcante	145.8	Bad	18	Rua Paulina Maria de Mendonça	105.0	Bad
11	Av. Governador Luis Cavalcante	140.4	Bad	37	Rua General Hermes	96.2	Bad
24	Rua Senador Rui Palmeira	140.4	Bad	11	Av. Governador Luis Cavalcante	94.2	Bad
7	Rua 01	137.5	Bad	7	Rua 01	93.8	Bad
5	Rua Hugo Corrêa Paes	129.0	Bad	9	Av. Cachoeira do Mirim	88.8	Bad

*The segments that are highlighted in bold are the segments that are present in both assessment methods.
Source: elaborated by the author.

Table 10 – The best segments according to each index

ID	Location	IGG	Status – IGG	ID	Location	SPS	Status – SPS
22	AL-110	19.1	Excellent	10	Rua Senador Firmino de Vasconcelos	11.4	Excellent
10	Rua Senador Firmino de Vasconcelos	27.6	Good	14	Rua Adelaí de Melo Mota	19.8	Good
13	Rua João Norberto de Lima	32.3	Good	22	AL-110	20.5	Good
30	Rua Paulina Maria de Mendonça	37.2	Good	13	Rua João Norberto de Lima	21.3	Good
15	Av. Empresário Carlos da Silva Nogueira	43.3	Regular	16	Av. José Haílton dos Santos	24.2	Good
40	Av. Paulo Holanda	44.2	Regular	15	Av. Empresário Carlos da Silva Nogueira	30.9	Regular
12	Rua Dona Alzira Aguiar	44.4	Regular	30	Rua Paulina Maria de Mendonça	36.0	Regular
20	Rua Professor Manuel Coelho Neto	46.2	Regular	12	Rua Dona Alzira Aguiar	37.8	Regular
14	Rua Adelaí de Melo Mota	46.9	Regular	8	Av. Governador Luis Cavalcante	38.8	Regular
36	Av. Presidente Getúlio Vargas	47.6	Regular	36	Av. Presidente Getúlio Vargas	39.2	Regular

*The segments that are highlighted in bold are the segments that are present in both assessment methods.
Source: elaborated by the author.

It is important to emphasize that in all 19 segments, where the IGG exhibited a terrible or bad status, the SPS also presented a terrible or bad status, not necessarily identical statuses, but always negative condition. It is important to emphasize that in all 4 segments, where the IGG exhibited an excellent or good status, the SPS also presented an excellent or good status, inclusive identical statuses.

The segments with regular statuses are characterized by the largest frequency of distinct statuses. The minor deterioration pavements would not significantly affect the formulation of investment programs or intervention strategies when considering the entire network at a macro level. These segments would be given priority for intervention only when the agency in charge has ample financial resources, which is uncommon in developing nations like Brazil.

6.5 Conclusions

The following conclusions have been drawn from the research that was conducted:

□ Despite the variation between the IGG and SPS values, utilizing data derived from the SPS index can support Pavement Management System (PMS) initiatives at the network level. This is achieved by facilitating the identification of roads exhibiting significant pavement surface degradation and the presence of defects. Consequently, this enables the prioritization of maintenance activities by selecting segments that require immediate attention.

□ Due to insufficient data, it is advisable to utilize the available information for the purpose of planning annual interventions, as the development of performance prediction models based on this index is currently not feasible.

□ Deformation is recognized as a major factor that contributes to the observed separation between the IGG and SPS indices. To address this problem, it is feasible to coordinate the assessment of the pavement with the collection of data on pavement irregularity/roughness using smartphone-based devices. The examination of pavement degradation, quantified by the Surface Pavement Score (SPS), along with the assessment of user comfort levels, as indicated by roughness, offers further criteria for analysing and identifying critical parts. Both smartphones and action cameras are affordable and readily available devices in the commercial market, offering cost-effective alternatives to traditional assessment methods.

6.6 References

CAUSIM, Patrícia Bolsonaro. **Estudo de um sistema de gerência de pavimentos para cidades de pequeno e medio porte**. 2001. 128 f. Universidade Estadual de Campinas, Campinas, 2001.

CHA, Young-Jin; CHOI, Wooram; BÜYÜKÖZTÜRK, Oral. Deep learning-based crack damage detection using convolutional neural networks. **Computer-Aided Civil and Infrastructure Engineering**, [*s. l.*], v. 32, p. 361-378, 2017.

CNT; SEST SENAT. **Pesquisa CNT de rodovias 2023**. Brasília: [*s. n.*], 2023.

COENEN, Tom B. J.; GOLROO, Amir. A review on automated pavement distress detection methods. **Cogent Engineering**, [*s. l.*], v. 4, n. 1, p. 1-23, 2017. Available from: http://doi.org/10.1080/23311916.2017.1374822. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; RAHMAN, Mujib; MATHAVAN, Senthan; NOBRE JÚNIOR, Ernesto Ferreira. Comparing different deep learning architectures as vision-based multi-label classifiers for identification of multiple distresses on asphalt pavement. **Transportation Research Record**: Journal of the Transportation Research Board, [New York], v. 2677, n. 5, p. 24-39, May 2023. DOI: https://doi.org/10.1177/036119812227273. Available from: https://journals.sagepub.com/doi/abs/10.1177/03611981221127273. Access in: 15 June 2024.

FERNANDES JUNIOR, Jose Leomar. **Sistemas de gerência de pavimentos urbanos para cidades de médio porte**. São Carlos, SP: Universidade de São Paulo, 2001.

GOPALAKRISHNAN, Kasthurirangan. Deep learning in data-driven pavement image analysis and automated distress detection: a review. **Data**, [*s. l.*], v. 3, n. 3, 2018.

IKRAM, Chaudhry Rehan. **A benchmark for evaluating deep learning based image analytics**. Oslo: University of Oslo, 2019.

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **NORMA DNIT 005/2003 - TER**: defeitos nos pavimentos flexíveis e semi-rígidos: terminologia. Rio de Janeiro: DNIT, 2003a.

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **NORMA DNIT 006/2003 - PRO**: avaliação objetiva da superfície de pavimentos flexíveis e semi-rígidos: procedimento. Rio de Janeiro: DNIT, 2003b.

LI, Qingquan *et al.* FoSA: F * Seed-growing Approach for crack-line detection from pavement images. **Image and Vision Computing journal**, [*s. l.*], v. 29, n. 12, p. 861-872, 2011. Available from: https://doi.org/10.1016/j.imavis.2011.10.003. Access in: 15 June 2024.

ZALAMA, Eduardo; JAIME, G.; MEDINA, Roberto. Road crack detection using visual features extracted by gabor filters. **Computer-Aided Civil and Infrastructure Engineering**, [*s. l.*], v. 29, p. 342-358, 2014.

7 MANUSCRIPT PAPER 04: ECONOMICAL AUTOMATED PAVEMENT DIAGNOSTIC SYSTEM TO NETWORK PAVEMENT MANAGEMENT SYSTEMS (PMS)

Abstract

The thorough and ongoing assessment of the pavement in a road network is crucial for the functioning of Pavement Management Systems (PMS) and the upkeep of the standard of pavement quality. Through pavement diagnosis, it is feasible to determine priorities and take effective action on the pavement, resulting in reduced costs for restoration projects and an extended lifespan for the roads. Nevertheless, the expenses associated with evaluating the PMS can render its deployment impractical due to the need for significant expertise, costly equipment, and lengthy execution periods. One option is to transition to completely automated pavement evaluations, utilizing road photos and employing Convolutional Neural Network to categorize flaws in the photographs. The primary findings outlined in this paper include: 1) organizing the inexpensive diagnostic system using affordable survey equipment, 2) utilizing image processing on a cost-free platform to automatically generate reports and maps of defects, and 3) confirming the necessity for calibration or preliminary updates to enhance the applicability and accuracy of the developed multi-label classification model, minimizing false-positive results. The new generalized model is anticipated to serve as a beneficial tool for practicing engineers to utilize in their normal investigations.

Keywords: Defect Classification; Pavement Diagnostic System; Pavement Management Systems; Simplified Pavement State Index.

7.1 Introduction

Development can be defined as the process of enhancing the well-being of a society by establishing suitable social, political, and economic circumstances. The anticipated results encompass both quantitative and qualitative enhancements in human capital, such as increased income and education levels, as well as improvements in physical capital, including infrastructure such as utilities, transportation, and telecommunications.

The growth of transportation networks is intricately linked to the socioeconomic environment. It relies on the interplay between physical and human resources, as infrastructure cannot function optimally without adequate management, operations, and maintenance. Simultaneously, economic activity is reliant on the presence of an infrastructure foundation. Effective logistics depend on the presence of infrastructure and the application of managerial expertise.

Due to its extensive utilization of infrastructure, the transportation industry plays a crucial role in the economy. There is a clear correlation between the amount and quality of

transportation infrastructure and the level of economic development. Development is often linked to the presence of well-developed transport infrastructure and interconnected networks. Efficient transport networks yield positive multiplier effects by enhancing accessibility to markets, employment, and new investments, so generating economic and social possibilities and benefits. Insufficient capacity or reliability in transport systems can result in economic consequences, such as wasted opportunities and a decreased quality of life.

Therefore, the management of this critical asset must be adequate to allow progress, mainly in countries with a transport matrix focused on highways. Nevertheless, due to limited resources available for maintenance versus the tremendous demand for investments, infrastructure suffers from continuous degradation of roads. For example, according to data from the Management Report of the National Transport Confederation (CNT) 2023, 67.5% of Brazilian highways managed by public management have unsatisfactory conditions, showing deficiencies in geometry, signage, or pavements (CNT; SEST SENAT, 2023).

When there are not enough resources available for maintenance, one possible solution to improve the performance of pavements is to adopt Pavement Management Systems (PMS). The Pavement Management System (PMS) has long been utilized to assist managers in making informed decisions regarding the preservation and expansion of the lifespan of pavements, as well as to strategize interventions and allocate resources in the near, medium, and far future. The objective of pavement management is to effectively allocate financial resources in order to provide users with pavements that are comfortable, safe, and cost-effective (Fernandes Junior, 2001). Put simply, it enables more efficient distribution of resources by implementing a prioritized system based on pavement evaluation, hence reducing expenses related to maintenance and rehabilitation. An essential component of a successful Pavement Management System (PMS) is the availability of dependable data regarding the current condition of the pavement. This data is used to assess the status of the pavement and to identify crucial sections and defects within the road network.

The evaluation of the pavement can be conducted using three different stages: manual, semi-automated, and automatic. Nevertheless, the manual examination is becoming unfeasible due to the necessary magnitude and regularity, its subjectiveness, exorbitant financial expenses, time constraints, and reliance on human resources (Gopalakrishnan, 2018). The capture of pavement images and position information was achieved by a semi-automatic evaluation method. This method utilized high-speed line scanning digital cameras, resulting in significant advancements over time (Zalama; Jaime; Medina, 2014). Nevertheless, these methods require the manual recognition of distress through the use of indoor interpretation

devices. In addition, the implementation of automated inspection systems was encouraged to facilitate the efficient and accurate detection of defects. However, these systems relied on costly sensors such as laser scanners (Coenen; Golroo, 2017).

Although these systems can obtain precise data, their utilization in high-frequency data collecting is hindered by the intricate operational and financial expenses involved. Primarily in small and medium-sized cities, as well as developing countries, where limited human and financial resources hinder the feasibility of establishing a Pavement Management System (PMS) for road assessment operations (Causim, 2001).

These algorithms can efficiently perform feature extraction and object identification by autonomously learning from a vast dataset of examples. Advancements in technology have led to the improvement of automatic systems that use image analysis to extract information and features from images. This is achieved through the process of automatic learning, which involves analyzing a large number of samples (Cha; Choi; Büyüköztürk, 2017). Various research works on various computer vision-based techniques for automatic detection of pavement deterioration can be found in the existing literature. Various techniques, including image processing methods like intensity-thresholding, edge detection, and seed-based methods, as well as machine learning techniques such as unsupervised learning, supervised learning, and deep learning, have been extensively employed for the purpose of identifying and quantifying defects (He *et al.*, 2016; Li *et al.*, 2011; Long *et al.*, 2020; Zhou *et al.*, 2012).

The multi-defect classification model of pavement developed by Espíndola *et al.* (2023) and improved by Paper 02 identified six different types of defects: bleeding, isolate crack, connected crack, revaling, patch and pothole. In addition, they examined several Convolutional Neural Network (CNN) structures and determined that the ResNet50-based model had the most favorable performance outcomes for predicting pavement problems. Also provided details on the optimization parameters and picture preprocessing required for this model. This paper aims to assess and evaluate the effectiveness of the model proposed by Paper 02 in an uncontrolled urban street environment with distinct characteristics from the original model. The study will utilize affordable equipment to collect data and georeferenced the model's outcomes, enabling the visualization of areas requiring rehabilitation and identifying the most critical segments.

The convolutional neural network consists of three layers: convolution layers, subsampling layers (max-pooling), and fully connected layers. Figure 1 displays the whole structure of the convolutional neural network. The convolution layer processes the input image by applying a kernel or filter to each pixel, resulting in the production of an output image.

Subsampling layers decrease the size of the input image to enhance the neural network's ability to handle variations and maintain stability. Ultimately, the inclusion of complete connection layers enables the neural network to progress in a forward direction, transforming the input into vectors of a predetermined length that can be classified into distinct categories (Xie; Zhang; Bai, 2017).

Figure 1 – The architecture of convolution neural network
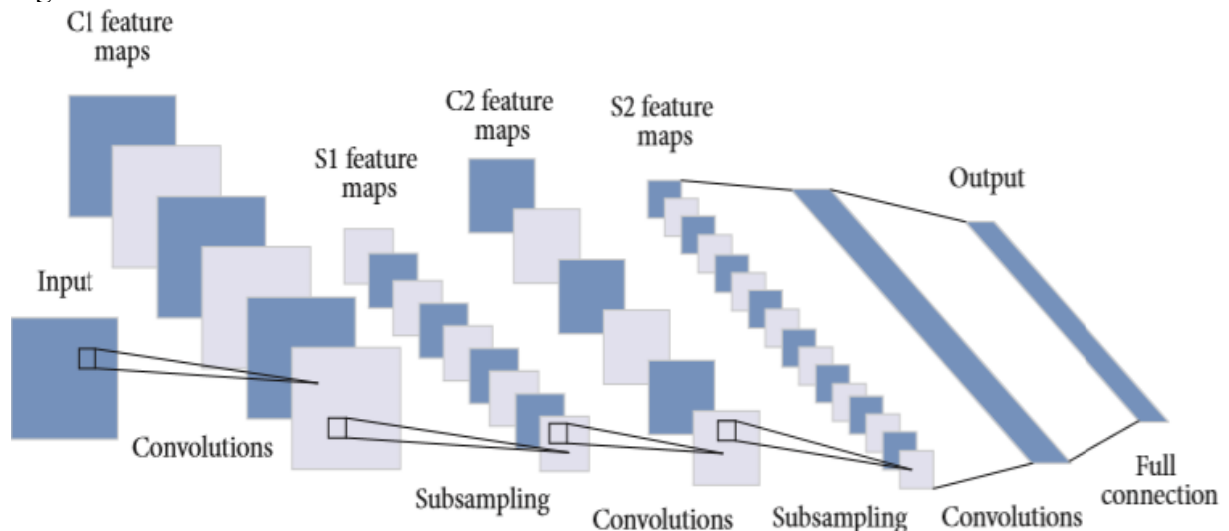
Source: Xie, Zhang and Bai (2017).

To reach a high level of accuracy in its predictions, CNN often relies on many annotated images. Nevertheless, the task of amassing a multitude of photos, which is normally necessary, and thereafter assigning labels to them manually, is unattainable. Transfer learning enables the construction of highly precise models using a reduced amount of input data. Instead of commencing the learning process from the beginning, it opts for a pre-trained model on a more extensive dataset to address a distinct categorization problem (Aparna *et al.*, 2019). Based on a literature survey, the pre-training models that are frequently used are VGG and ResNet (Keras, 2020). The study conducted by Espíndola *et al.* (2023) suggests utilizing ResNet50 as the convolutional neural network (CNN) architecture for the multi-distress classifier.

In 2015, the Residual Learning Network (ResNet) was created to examine the effectiveness of deep neural networks. They accomplished this by combining many layers and introducing CNNs with 18, 34, 50, 101, and 152 layers (He *et al.*, 2016). The network architecture achieved a test accuracy of 94% on the CIFAR-10 dataset and top-5 accuracy scores of 93% on ImageNet (Coleman *et al.*, 2017). Currently, deep learning (DL) networks exhibit excellent performance while demanding relatively modest computational resources. The computational demand of a DL network can be defined by its floating-point operations per

second (FLOPS), which is the number of floating-point operations it needs to do per second. Classifier networks such as ResNet34 and ResNet50 have a computational capacity of 4 Giga Floating Point Operations per Second (GFLOPS) (He *et al.*, 2016) for solving computer vision issues.

7.2 Research of pavement distress data detection algorithms

Research that used computer vision and deep learning to develop CNN to classify pavement defects by binary or multi-class methods. In the binary classification methods, the output indicates whether images contain the defect or not (Cao; Liu; He, 2020). Research such as Zhang *et al.* (2016), Paulya *et al.* (2017), Wang and Hu (2017), Chen and Jahanshahi (2018), Hoang *et al.* (2018), Yusof *et al.* (2018), Ranjbar *et al.* (2020), and Zhou and Song (2020), used a binary method to identify cracks in the pavement, and examples of research aimed at identifying potholes are Aparna *et al.* (2019), An *et al.* (2018), Pereira *et al.* (2018), and Srinidhi and Devi (2020).

Other researchers opted for multi-class classification methods that each object is only associated with a single label among the available classes, examples of research on pavement defects identification are Eisenbach *et al.* (2017), Nie and Wang (2018), Li *et al.* (2018), Xia (2018), and Asian *et al.* (2019). In other words, if the image had more than one defect simultaneously, it was impossible to identify all of them. While in multi-task learning, different tasks may involve different domains and data sets. Traditional two-class and multi-class problems can be cast into multi-label problems as developed by Espíndola *et al.* (2023) that in case the image contained bleeding, crack, patch and pothole simultaneously, the model was able to identify.

This research has shown that using models developed with a single dataset in an automated defect detection system is discouraged because the extrapolated may present wrong predictions or fail to return any label as an output. They presented a model built with three different datasets and showed that multiple defects in a single image could be reliably identified with up to 97% accuracy. The goal of including different datasets was to make it possible to generalize the classification algorithm more likely to be incorporated into network-level pavement management systems (Espíndola *et al.*, 2023). Figure 2 presents some results the authors obtained in identifying the simultaneous defect.

Figure 2 – Identifying the simultaneous pavement defect by model from Espíndola *et al.* (2023)

Dataset 01
Ground Truth: Bleeding Pothole
Model Output: Bleeding Pothole

Dataset 02
Ground Truth: Crack Patch
Model Output: Crack Patch

Dataset 03
Ground Truth: Patch Pothole
Model Output: Patch Pothole

Source: elaborated by the author.

7.3 Methods

7.3.1 Case study: Maceió-AL

An investigation was conducted by using the simultaneous defect classification model to a completely other dataset from the one used to construct the model. The classification model primarily utilized image of rural roadways to demonstrate the necessity of diversifying the dataset and avoiding extrapolation beyond the algorithm's bounds. To test this, the model was then applied to a distinct dataset featuring urban characteristics. The investigation of the extrapolation limit was conducted on urban roadways in the city of Maceió-AL in Brazil, as depicted in Figure 3 on the map. The image depicting pavement problems were isolated for inclusion in the construction of the comprehensive map and for comparison with the output generated by the model.

Figure 3 – Map of analyzed points
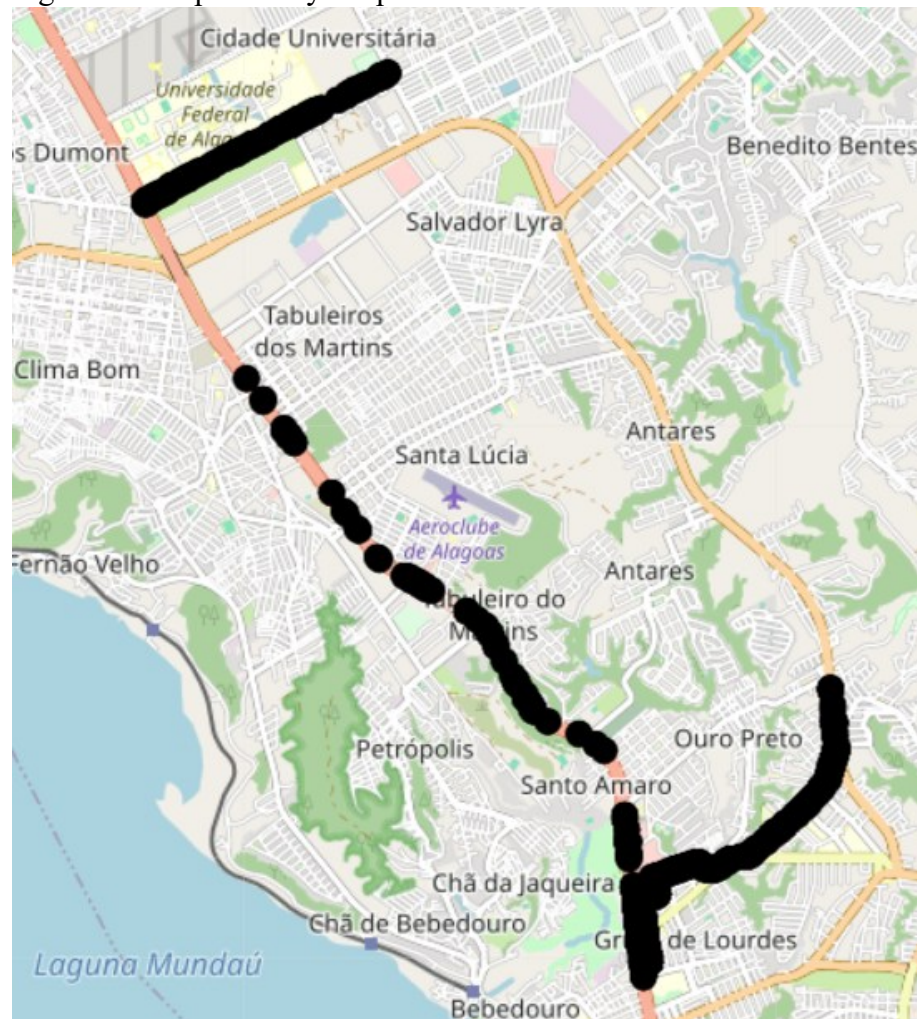
Source: elaborated by the author.

7.3.2 Data Collection

The equipment used for collecting images included a GoPro Hero 7 Black action camera, which was obtained for less than US$ 300. This camera, along with its protective case and suction holder, is readily available and reasonably priced, making it suitable for use in small communities. The camera was mounted in a car moving at a speed of 40-60 km/h (within the legal limit). The data collection was conducted in the JPG image format using the GPS-enabled "Time Lapse Photo" option, which was set to record an image every 0.5 seconds in linear mode. The linear mode prevents distortions at the edges of the image, and the GPS data is used to georeferenced information on pavement defects.

Figure 4 depicts the arrangement of the equipment utilized in this study, which was securely attached to the outer part of the car bonnet / front, Figure 4 (a), and at the rear of the vehicle, Figure 4 (b), configured to predominantly capture the pavement. The original concept

was to exclusively utilize the model on frontal images, which are comparable to the ones employed during the model's training process. Nevertheless, during the investigation, a diminished level of performance was noted compared to the control data. As a result, alternative data collection methods and image processing techniques for application were explored.

Figure 4 – Vehicle cameras placements – front (A) and rear (B) camera positions

Source: elaborated by the author.

The images were captured during daylight hours and under favorable weather conditions. They depict a range of scenarios with varying lighting conditions and shadows cast by items at the edges of the streets, as illustrated in Figure 5 which showcases a selection of collected images.

Figure 5 – Examples of images collected in the case study – front (A) and rear (B) camera positions

Source: elaborated by the author.

7.3.3 Prediction model system

The multi-label classification model utilized was previously developed and described in Paper 02. The method requires a minimal input of data collecting using a camera equipped with GPS, placed on a vehicle at an angle facing the pavement. The system follows the sequence of actions depicted in Figure 6. First, the image files are uploaded. Then, the system extracts the location information (latitude and longitude) and applies a multi-label classifier defects model. This model generates a list of the images and their corresponding defects. Ultimately, using this data, defect maps are produced.

Figure 6 – Flowchart of Economical Automated Pavement Diagnostic System
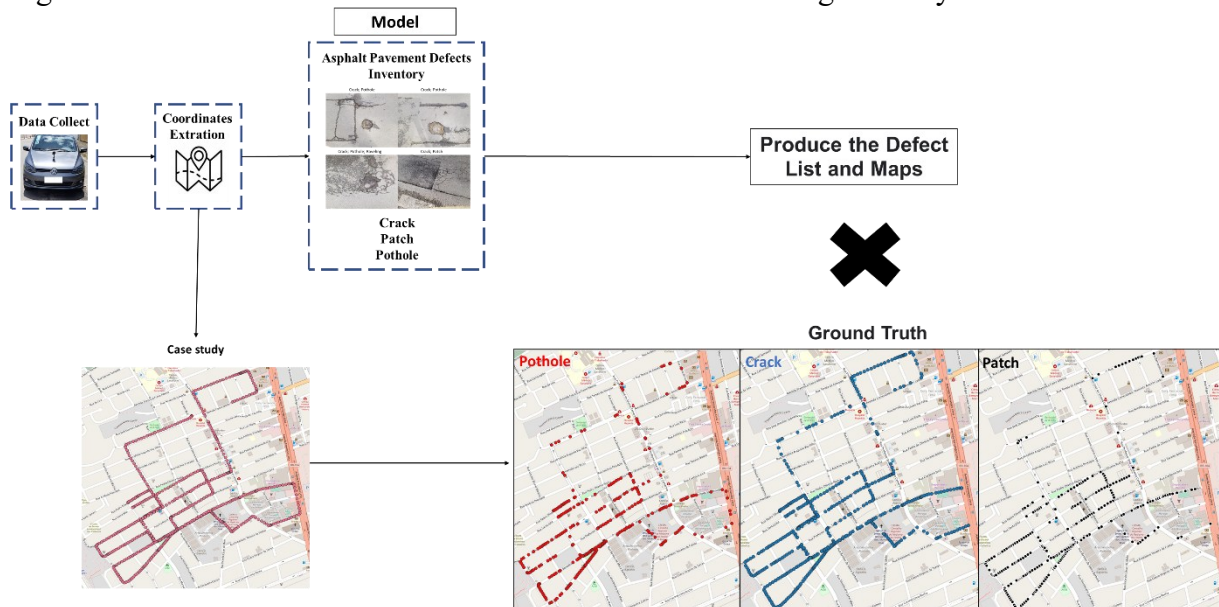
Source: elaborated by the author.

The objective of this study is to demonstrate the system's performance in an actual case study. During the survey of pavement defects, the following defects were identified: isolated crack, connected crack, bleeding, raveling, patch, and pothole. A total of 840 images with distresses from the front camera position and 384 images from the rear camera were extracted. From the collected images, those of streets/sections on cobblestone and concrete pavement were excluded, as well as asphalt pavement without defects. A total of 1224 images were submitted for analysis at the conclusion. Images were categorized based on the Ground Truth of the survey and subsequently compared with the outcomes derived from the model.

The Manuscript Paper 02 and Espíndola *et al.* (2023) model focuses on identifying defects in specific parts of an image rather than the complete image. Therefore, the prediction

of defects in the images of the case study followed a sequence depicted in Figure 7. Initially, the image undergoes pre-processing, during which 2 or 4 parts of the image are trimmed to ensure that the areas to be examined by the model consist mostly of pavement.

Next, the CNN model is utilized to classify each part of the image and assign labels. These labels are then grouped together to identify and analyze the pavement issues in that specific area. Once the labels were grouped, the results were compared to the Ground Truth and the performance metrics were validated.

Figure 7 – Flowchart illustrating the process of identifying distress
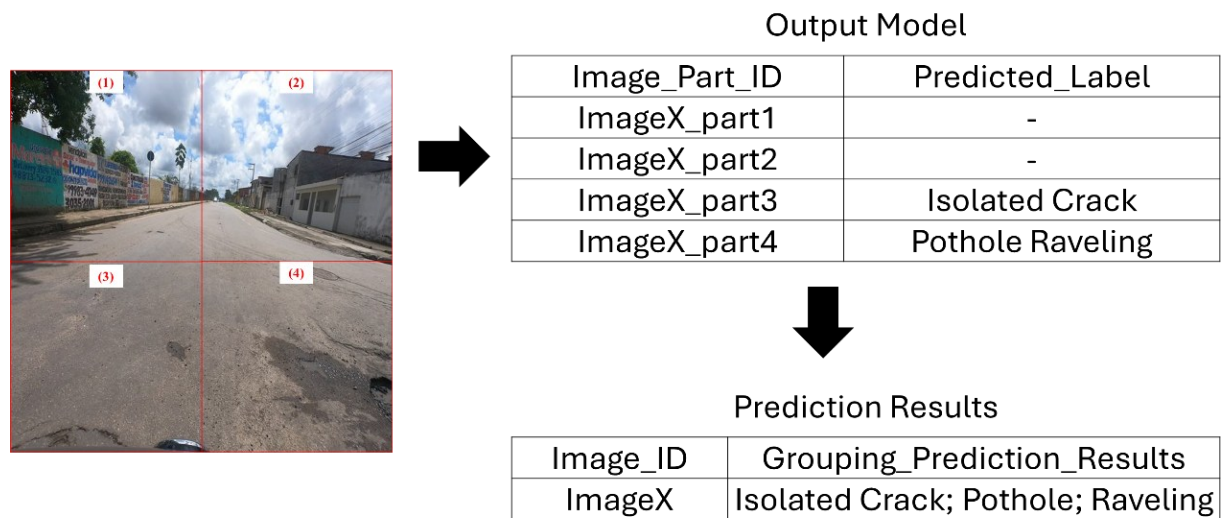
Source: elaborated by the author.

The original proposal was to examine the layout that was cropped in the images of the datasets to construct the model, type 2 of Figure 8. However, it was discovered that there could be advantages in exploring other pre-processing options. This is because the frontal image contains a larger variety of elements within the road's right-of-way, making it challenging to detect pavement defects. Subsequently, two more layouts of preprocessing were examined for the images obtained from the front camera, as shows Figure 8.

Figure 8 – The image preprocessing layouts
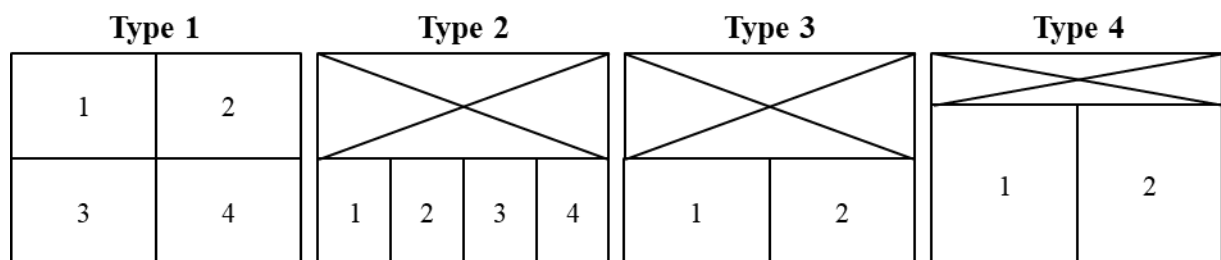
Source: elaborated by the author.

Accuracy, precision, recall, and F1-score were the performance metrics used to evaluate the model's applicability in the case study. Accuracy provides an overall assessment of the model's performance across all classes, making it useful when all classes are equally important. It is calculated as the ratio of correct predictions to the total number of predictions, as illustrated in Equation 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision can be defined as the proportion of true positives (TP) to the total number of predicted positives. This metric indicates the accuracy of the positive predictions made by the model. The formula for precision is provided in Equation 2.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall is calculated as the ratio of the number of true positive predictions to the total number of actual positive instances in the ground truth. The formula for recall is provided in Equation 3.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The F1 score quantifies the overall predictive capability of a model by harmoniously merging two typically conflicting variables: precision and recall. This is mathematically represented by Equation 4.

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} \tag{4}$$

7.4 Results and discussion

Analyzing the results of the real evaluation of the road (Ground Truth), it is possible to verify that the pavement is in an advanced state of degradation. Table 1 presents the results of cataloging the defects during the assessment and identifies which proportion of samples (images) have defects, separately.

Table 1 – Ground Truth Defect proportions in the case study

Distress	Image Quantity	%
Connected Crack	267	21.8%
Isolated Crack	564	46.1%
Patch	687	56.1%
Pothole	266	21.7%
Raveling	228	18.6%
Bleeding	413	33.7%
Total	**1,224**	**100.0%**

Source: elaborated by the author.

Through Table 2, it is possible to verify that 69.5% of the samples presented contain more than one defect simultaneously, reinforcing the importance of improving automatic methods of identifying multiple defects. Figure 9 presents a map displaying the quantity of defects identified in each sample.

Table 2 – Quantity of simultaneous defects

#	6	5	4	3	2	1	0	Total
Images Quantity	0	13	57	221	560	349	24	1224
%	0.0%	1.1%	4.7%	18.1%	45.8%	28.5%	2.0%	100.0%

Source: elaborated by the author.

Figure 9 – Distress Map

Source: elaborated by the author.

The case study images entered the flow of the automatic identification system, were preprocessed, divided into 2 or 4 image parts according to the type of pre-processing (Type 1, 2, or 3), the defect prediction model was applied to each image part, and then the prediction results were grouped to obtain the final label of the images.

According to Out of the four defects analyzed, the connected crack showed the most favorable outcomes. It achieved an accuracy of 91% in correctly identifying the label, both in terms of true-positive and true-negative results. Additionally, it obtained an F1-score of 76%, which indicates a high ability to identify fewer false positives and false negatives. Table 3 demonstrates that following the occurrence of a linked crack, the defects consistently appeared in the same order in terms of performance: connected crack, patch, pothole, and isolated crack. However, in contrast to the control data, only the connected crack yielded accuracy values over 90%, while the accuracy for other defects and metrics ranged from 60 to 85%.

Table 3, the model performed worse than the controlled data. Highlighting the need for calibration to maximize its application. Out of the three forms of pre-processing, type 2 processing, which is like the one utilized during model training, had the lowest performance. It has a high level of accuracy in diagnosing only connected cracks most of the time.

Table 3 – Performance of the model per distress

Preprocessing Layouts	Metric/Distress	Bleeding	Connected Crack	Isolated Crack	Patch	Pothole	Raveling
Type 1	Accuracy	74.2%	92.3%	69.9%	84.0%	78.6%	68.2%
	Precision	74.6%	82.9%	63.2%	86.2%	50.5%	19.4%
	Recall	35.6%	81.6%	83.3%	85.2%	73.3%	22.4%
	F1 score	48.2%	82.3%	71.9%	85.7%	59.8%	20.8%
Type 2	Accuracy	66.7%	89.2%	57.0%	66.4%	75.9%	71.8%
	Precision	69.2%	88.1%	52.1%	74.0%	44.9%	16.9%
	Recall	2.2%	58.4%	85.5%	62.0%	47.7%	13.2%
	F1 score	4.2%	70.3%	64.7%	67.5%	46.3%	14.8%
Type 3	Accuracy	73.6%	91.1%	74.8%	82.8%	82.7%	71.7%
	Precision	95.9%	92.0%	71.9%	88.4%	59.6%	18.6%
	Recall	22.8%	64.8%	74.1%	79.9%	62.8%	15.4%
	F1 score	36.8%	76.0%	73.0%	83.9%	61.2%	16.8%
Type 4	Accuracy	73.8%	91.1%	74.9%	83.0%	82.7%	71.7%
	Precision	96.0%	92.0%	72.0%	88.6%	59.5%	18.6%
	Recall	23.2%	64.8%	74.6%	80.1%	62.6%	15.4%
	F1 score	37.4%	76.0%	73.3%	84.1%	61.0%	16.8%

Source: elaborated by the author.

Due to the limited number and variety of images, the metrics analysis revealed a very low occurrence of raveling and bleeding defects, specifically less than 15%. Consequently, these defects had to be excluded from the Simplified Pavement State Index (SPS), as recommended in manuscript Paper 03.

It is worth noting that the type of pre-processing yielded optimal results but type 2 yielded unsatisfactory results. In terms of the processing of types 1, 3 and, they exhibited comparable performance, with type 4 demonstrating a tiny advantage. The act of manufacturing anything has been established as the standard procedure to be used when the model is being implemented.

Although the occurrence of raveling has been examined in the model, the SPS index no longer considers this defect into account when calculating its value. Therefore, it had a negligible effect on the low Recall and F1-score values.

The low Recall and F1-score values related to bleeding will lead to an increased number of false positives, necessitating a reconsideration of the training and identification methods for this defect. The bleeding is typically observed in greater areas, particularly along the wheel track. During the training, a decision was made to crop the image to conduct a more detailed analysis of the defects, particularly those found in smaller areas such as isolated cracks. This preprocessing choice may have improved the detection of cracks, but it also resulted in a decrease in bleeding performance.

In the ongoing research conducted by the study group, they have added entire images or different preprocessing layout that cover more regions of the image into the training dataset. These preprocessing approaches can include type 3 or 4 preprocessing layouts. Nevertheless, the primary objective of this system module is to ascertain the SPS index, which identifies the most crucial segments and indicates any existing defects. This information is intended to aid in the development of a pavement maintenance plan, with priority given to segments that exhibit subpar or inadequate conditions.

The complexity of images and defects is significant, with several variables that differentiate each dataset and type of defect. Using potholes as an example, Figure 10 displays a range of characteristics like dimensions, colors, format, location, lighting, texture, material of the base layer, and the presence or absence of water, among others. Based on the results, it is evident that calibration is essential until the dataset becomes sufficiently reliable and encompasses every variation in defects and conditions within the road domain range.

The ILSVRC employs 1.2 million images for training, 50 thousand for validation, and 100 thousand for testing purposes. Utilizing transfer learning reduces the need for such a large number of images, while it remains important to augment the training dataset in order to enhance the classifier.

Figure 10 – Mosaic with the pothole diversity of the dataset

Source: elaborated by the author.

Upon utilizing the defect classification module, which employs a CNN to identify defects within the image, a csv file is produced with detailed information regarding the defects present in each section of the image. This step follows the third type of pre-processing. The georeferenced data is merged with the defect list csv. Next, it proceeds to the management module, where the primary objective is to identify the most concerning segments that exhibit the greatest SPS values. The management module provides two common outputs: a visual map displaying the results (Figure 11) and a list of defects identified in the segments (Figure 12), allowing for easy identification of segments requiring maintenance due to their critical state.

Figure 11 – Defects Prediction Map
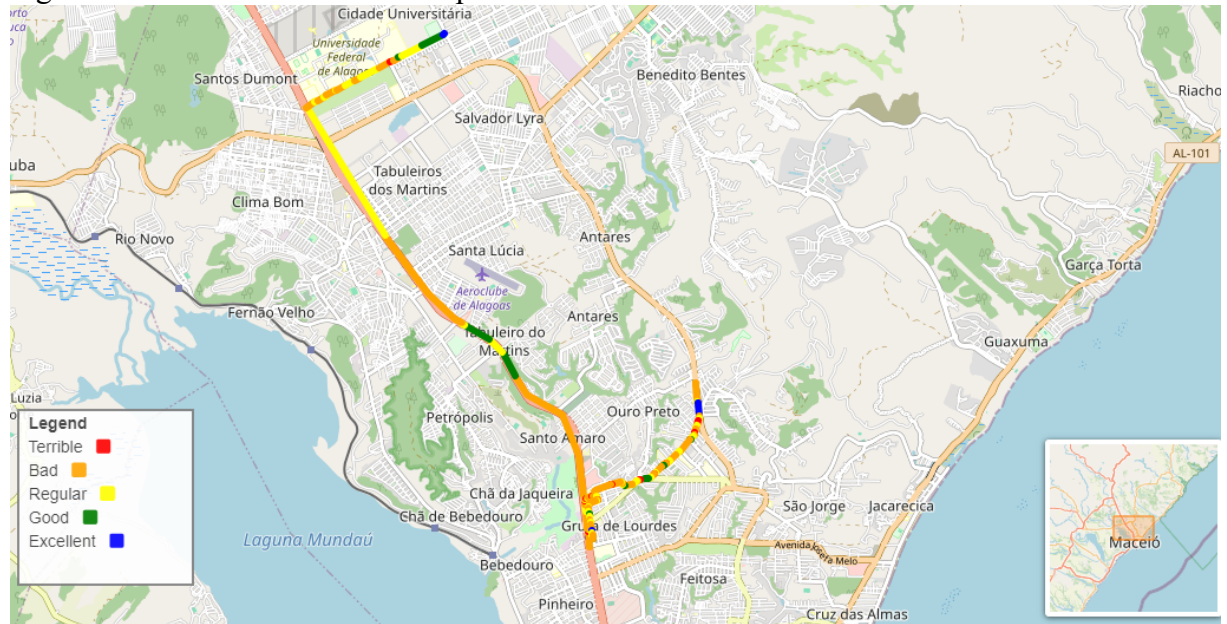
Source: elaborated by the author.

Figure 12 – Pavement quality report

seg_hom	SPS	Concept	Isolated_Crack	Connected_Crack	Patch	Pothole	Bleeding	Raveling	Latitude_Inicial	Longitude_Inicial	Latitude_Final	Longitude_Final
0	120.0	Bad	5.0	0.0	0.0	5.0	0.0	0.0	-9.600402	-35.723140	-9.601802	-35.722973
1	130.0	Bad	3.0	2.0	0.0	4.0	0.0	2.0	-9.601863	-35.722966	-9.603056	-35.722827
2	39.0	Regular	5.0	1.0	0.0	0.0	0.0	1.0	-9.603110	-35.722821	-9.603326	-35.722795
3	8.0	Excellent	2.0	0.0	0.0	0.0	0.0	0.0	-9.603373	-35.722789	-9.605181	-35.722582
4	100.0	Bad	5.0	0.0	0.0	4.0	0.0	0.0	-9.605217	-35.722582	-9.605479	-35.722631
...	...	...	...	...	...	...	...	...	...	...	...	...

Source: elaborated by the author.

Figure 11 presents the map of the defects predicted by the model to show how the system output is. Consistent with the findings in Paper 03 document, the SPS indicators reduce the severity of pavement deterioration. Furthermore, it is crucial to note that the model's performance remains in need of enhancement, as indicated by Table 3.

This improvement is mostly required to decrease the occurrence of false positives. Therefore, the model still requires calibration, resulting in discrepancies between the map depicted in Figure 11.

An advantage of generating this data in a straightforward and automated manner is the ability to rapidly get the most degraded segments, which can then be used to design preventive maintenance system (PMS) operations at the project level. Figure 13 displays the sections that achieved the highest SPS values. It is evident that the portions are brief, with minimal expansions. This is attributed to the methodology used for collecting images, because the camera captures images at a frequency of 0.5s. In urban regions with heavy traffic or

intersections with traffic lights, the recorded images may have minimal or no time gap between them, especially when a vehicle is stationary at a traffic light or corner.

Figure 13 – Map displaying segments with the top 10 highest SPS
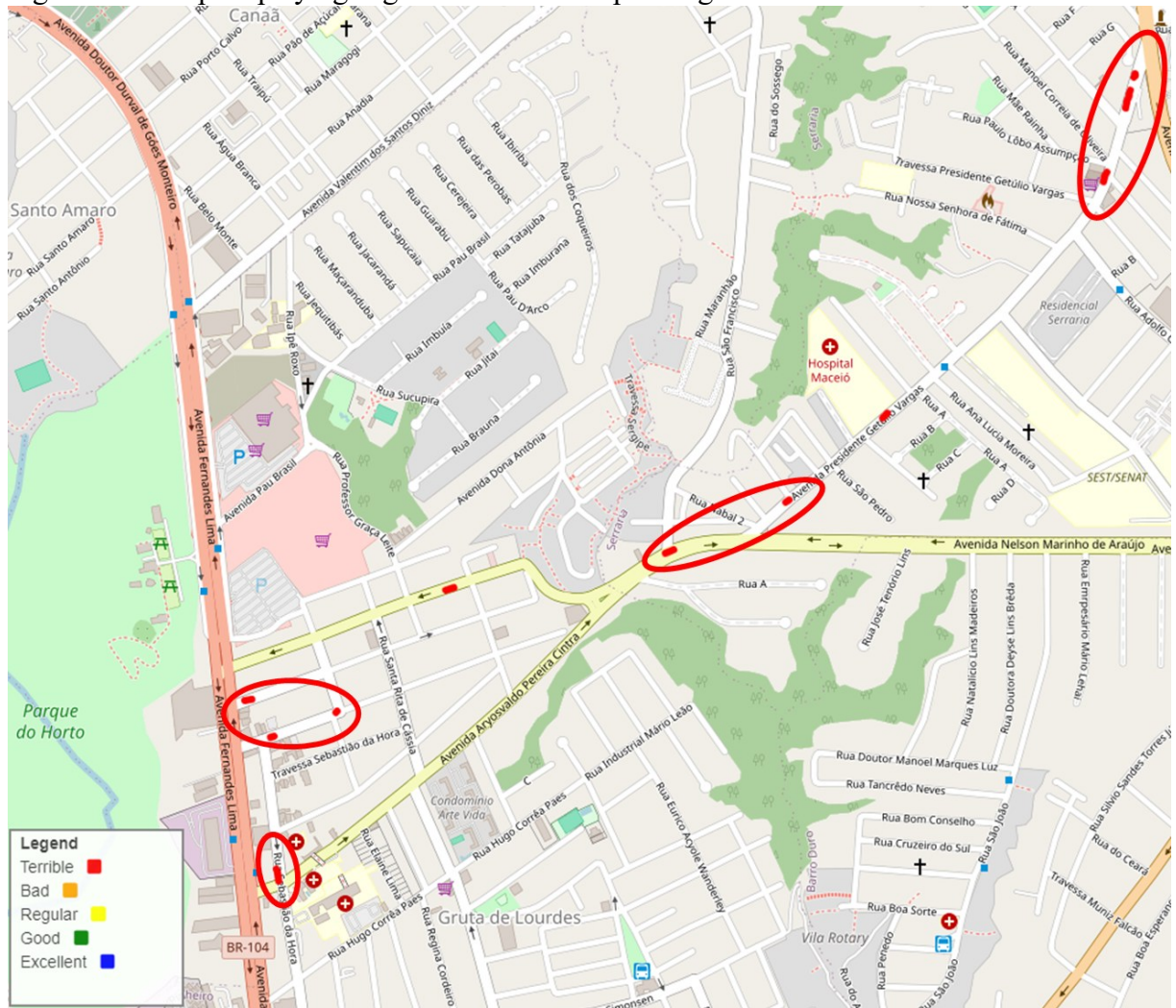
Source: elaborated by the author.

To replicate the IGG determination approach, a sample (image) would be taken every 20 meters. This would need the vehicle to travel at a speed of 36km/h, as commercial cameras can capture an image every 0.5 seconds. Within urban areas, the speed at which images can be captured is typically reduced, with a speed of 36km/h being a feasible option. Nevertheless, in the case of rural roads, the speed limits are elevated. For instance, state highways in Alagoas are primarily engineered to accommodate speeds of 70 and 80km/h. Traveling at a speed of 36km/h is not advisable from a safety perspective.

An area for improvement in the suggested method is to enhance the model itself by increasing the quantity and variety of images used for training. Another measure that can be

implemented is to assess the influence of gathering samples at a velocity of 54km/h on rural highways, with an image or sample acquired every 30m.

Figure 14 displays the visual comparison between the IGG results and the SPS of the case study. There is a noticeable difference between them, particularly in the degradation of the intermediate (regular) pavement, which is similar to what is described in manuscript Paper 03.

Figure 14 – IGG vs. SPS results

Source: elaborated by the author.

Upon comparing the Ground Truth Map (IGG) and the Prediction Map (SPS), it can be verified that the streets allocated for rehabilitation will stay almost unaltered, successfully fulfilling the purpose of the PMS at the network level. Nevertheless, there can be discrepancies in the intensity or regularity of defects, which would directly influence a Project Management System (PMS) at the project level.

7.5 Conclusions

The following conclusions are drawn from the research:

– The action camera utilized in the system tests produced georeferenced images of high quality, enabling the straightforward, reliability, and cost-effective identification of defects.

– Fully automated evaluations offer the benefit of producing reports and maps that include a comprehensive list of defects and their respective locations within a few minutes. This minimizes human intervention, accelerates the evaluation process, and lowers labor expenses. The obtained results can assist managers in the process of decision-making, which is one of the primary goals of the Performance Management System (PMS).

– The study proposed an economically feasible alternative to conventional evaluations for assessing the surface state of the pavement. This alternative can be used at a network level to aid in pavement management. It provides technical information in the form of a map and report.

– The pavement evaluation conducted in the case study highlights the necessity of automated assessments that encompass several types of defects. Approximately half of the samples exhibited numerous defects simultaneously.

– It is worth mentioning that despite encouraging outcomes, it is crucial to fine-tune the model in order to minimize labeling errors as the database grows and becomes more diverse.

Several measures are being taken to enhance the diagnostic system, with the primary ones being:

– Examine image preprocessing techniques that enhance the optimal classification of bleeding defects, considering the utilization of entire images and/or images obtained by preprocessing methods of layout type 3 or 4.

– When incorporating a new dataset with diverse features into a generalized model, it is important to determine the minimum number of images required. These characteristics include pavement type, lighting, location, climate, camera specifications, angulation, and camera orientation.

– Research the establishment of a standardized method for adjusting the angle and position of the camera to ensure that the image primarily captures the pavement, so minimizing any disruptions caused by elements in the region around it.

7.6 References

AN, Kwang Eun *et al.* Detecting a pothole using deep convolutional neural network models for an adaptive shock observing in a vehicle driving. **2018 IEEE International Conference on Consumer Electronics, ICCE 2018**, [*s. l.*], 2018.

APARNA *et al.* Convolutional neural networks based potholes detection using thermal imaging. **Journal of King Saud University – Computer and Information Sciences**, [*s. l.*], 2019. Available from: https://doi.org/10.1016/j.jksuci.2019.02.004. Access in: 15 June 2024.

ASLAN, O. D. *et al.* Using artifical intelligence for automating pavement condition assessment. **International Conference on Smart Infrastructure and Construction 2019, ICSIC 2019: Driving Data-Informed Decision-Making**, [*s. l.*], v. 2019, n. 2018, p. 337-341, 2019.

CAO, Wenming; LIU, Qifan; HE, Zhiquan. Review of pavement defect detection methods. **IEEE Access**, [*s. l.*], v. 8, p. 14531-14544, 2020.

CAUSIM, Patrícia Bolsonaro. **Estudo de um sistema de gerência de pavimentos para cidades de pequeno e medio porte**. 2001. 128 f. Universidade Estadual de Campinas, Campinas, 2001.

CHA, Young-Jin; CHOI, Wooram; BÜYÜKÖZTÜRK, Oral. Deep learning-based crack damage detection using convolutional neural networks. **Computer-Aided Civil and Infrastructure Engineering**, [*s. l.*], v. 32, p. 361-378, 2017.

CHEN, Fu-Chen; JAHANSHAHI, Mohammad R. NB-CNN: deep learning-based crack detection using convolutional neural network and naive bayes data fusion. **IEEE Transactions on Industrial Electronics**, [*s. l.*], v. 65, n. 5, p. 4392-4400, 2018.

CNT; SEST SENAT. **Pesquisa CNT de rodovias 2023**. Brasília: [*s. n.*], 2023.

COENEN, Tom B. J.; GOLROO, Amir. A review on automated pavement distress detection methods. **Cogent Engineering**, [*s. l.*], v. 4, n. 1, p. 1-23, 2017. Available from: http://doi.org/10.1080/23311916.2017.1374822. Access in: 15 June 2024.

COLEMAN, Cody *et al.* DAWNBench: an end-to-end deep learning benchmark and competition. *In*: Long Beach, CA, USA. **31st Conference on Neural Information Processing Systems**. Long Beach, CA, USA: [*s. n.*], 2017.

EISENBACH, Markus *et al.* How to get pavement distress detection ready for deep learning? A systematic approach. **Proceedings of the International Joint Conference on Neural Networks**, [*s. l.*], v. 2017-May, p. 2039-2047, 2017.

ESPÍNDOLA, Aline Calheiros; RAHMAN, Mujib; MATHAVAN, Senthan; NOBRE JÚNIOR, Ernesto Ferreira. Comparing different deep learning architectures as vision-based multi-label classifiers for identification of multiple distresses on asphalt pavement. **Transportation Research Record**: Journal of the Transportation Research Board, [New York], v. 2677, n. 5, p. 24-39, May 2023. DOI: https://doi.org/10.1177/036119812227273.

Available from: https://journals.sagepub.com/doi/abs/10.1177/03611981221127273. Access in: 15 June 2024.

FERNANDES JUNIOR, Jose Leomar. **Sistemas de gerência de pavimentos urbanos para cidades de médio porte**. São Carlos, SP: Universidade de São Paulo, 2001.

GOPALAKRISHNAN, Kasthurirangan. Deep learning in data-driven pavement image analysis and automated distress detection: a review. **Data**, [*s. l.*], v. 3, n. 3, 2018.

HE, Kaiming *et al.* Deep residual learning for image recognition. *In*: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [*S. l.*: *s. n.*], 2016. p. 770-778. Available from: http://image-net.org/challenges/LSVRC/2015/. Access in: 15 June 2024.

HOANG, Nhat-duc; NGUYEN, Quoc-lam; TRAN, Van-duc. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. **Automation in Construction**, [*s. l.*], v. 94, n. July, p. 203-213, 2018.

KERAS. **Keras Documentation**. [*S. l.*], 2020.

LI, Baoxian *et al.* Automatic classification of pavement crack using deep convolutional neural network. **International Journal of Pavement Engineering**, [*s. l.*], v. 0, n. 0, p. 1-7, 2018. Available from: https://doi.org/10.1080/10298436.2018.1485917. Access in: 15 June 2024.

LI, Qingquan *et al.* FoSA: F * Seed-growing Approach for crack-line detection from pavement images. **Image and Vision Computing journal**, [*s. l.*], v. 29, n. 12, p. 861-872, 2011. Available from: https://doi.org/10.1016/j.imavis.2011.10.003. Access in: 15 June 2024.

LONG, Xiang *et al.* PP-YOLO: An effective and efficient implementation of object detector. **arXiv**, [*s. l.*], 2020.

NIE, Mingxin; WANG, Kun. Pavement distress detection based on transfer learning. **2018 5th International Conference on Systems and Informatics, ICSAI 2018**, [*s. l.*], n. Icsai, p. 435-439, 2018.

PAULYA, Leo *et al.* Deeper networks for pavement crack detection. *In*: Taipei, Taiwan. **International Symposium on Automation and Robotics in Construction**. Taipei, Taiwan: [*s. n.*], 2017. p. 479-485.

PEREIRA, Vosco *et al.* Classification of paved and unpaved road image using convolutional neural network for road condition inspection system. **ICAICTA 2018 – 5th International Conference on Advanced Informatics: Concepts Theory and Applications**, [*s. l.*], p. 165-169, 2018.

RANJBAR, Sajad; NEJAD, Fereidoon Moghadas; ZAKERI, H. An image-based system for pavement crack evaluation using transfer learning and wavelet transform. **International Journal of Pavement Research and Technology**, [*s. l.*], p. 1-13, 2020.

SRINIDHI, G.; DEVI, Renuka. **Pothole Detection using CNN and AlexNet**. [*S. l.*], 2020.

WANG, Xianglong; HU, Zhaozheng. Grid-based pavement crack analysis using deep learning. **2017 4th International Conference on Transportation Information and Safety, ICTIS 2017 – Proceedings**, [*s. l.*], p. 917-924, 2017.

XIA, Wei. An approach for extracting road pavement disease from HD camera videos by deep convolutional networks. **ICALIP 2018 – 6th International Conference on Audio, Language and Image Processing**, [*s. l.*], p. 418-422, 2018.

XIE, Danfeng; ZHANG, Lei; BAI, Li. Deep learning in visual computing and signal processing. **Applied Computational Intelligence and Soft Computing**, [*s. l.*], v. 2017, 2017.

YUSOF, N. A. M. *et al.* Crack detection and classification in asphalt pavement images using deep convolution neural network. **Proceedings – 8th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2018**, [*s. l.*], p. 227-232, 2018.

ZALAMA, Eduardo; JAIME, G.; MEDINA, Roberto. Road crack detection using visual features extracted by gabor filters. **Computer-Aided Civil and Infrastructure Engineering**, [*s. l.*], v. 29, p. 342-358, 2014.

ZHANG, Lei *et al.* Road crack detection using deep convolutional neural network. **Proceedings – International Conference on Image Processing, ICIP**, [*s. l.*], v. 2016-Augus, p. 3708-3712, 2016.

ZHOU, Shanglian; SONG, Wei. Deep learning–based roadway crack classification with heterogeneous image data fusion. **Structural Health Monitoring**, [*s. l.*], 2020.

ZHOU, Zhi Hua *et al.* Multi-instance multi-label learning. **Artificial Intelligence**, [*s. l.*], v. 176, n. 1, p. 2291-2320, 2012. Available from: https://doi.org/10.1016/j.artint.2011.10.002. Access in: 15 June 2024.

8 CONCLUSIONS

Automating the collection of pavement surface condition data offers several significant advantages. Primarily, the integration of deep learning, particularly through CNNs, enhances the accuracy and efficiency of data collection processes. Automated systems can continuously monitor and evaluate pavement conditions with precision, reducing human error and subjectivity. The use of CNNs enables the identification of surface defects, such as cracks, potholes, and patches, facilitating timely maintenance and rehabilitation activities.

Additionally, automated systems can cover larger areas in less time compared to manual inspections, leading to cost savings and improved resource allocation. This real-time data collection capability ensures that Pavement Management Systems are updated promptly, aiding in proactive decision-making and extending the service life of pavements. Despite these benefits, the initial setup and deployment of automated systems involve substantial investments in equipment and training. Ensuring robustness and reliability in diverse environmental conditions is a significant challenge.

The accuracy of the model is directly linked to the quality and size of the dataset used for CNN development, which may not adapt to all real-life scenarios. Moreover, the complexity of deep learning models like CNNs requires considerable computational power and expertise, necessitating collaboration between academia and highway management agencies. Additionally, there may be resistance to adopting new technologies among traditional pavement management professionals, necessitating comprehensive change management strategies.

Enhancing efficiency in data collection for pavement surface conditions involves a comprehensive methodology that includes several key stages. The process begins with meticulous planning and objective definition, where the specific pavement conditions to be assessed and desired outcomes are clearly identified. It is crucial to determine whether the assessment is for a network-level PMS or a project-level evaluation. This stage also involves defining the scope of data collection, including the geographical area and frequency.

The next stage is technology selection, which involves choosing appropriate technologies such as high-resolution cameras and LiDAR sensors for project-level assessments. For network-level evaluations, low-cost action cameras and smartphones, as proposed in this study, can be effective. Data acquisition is conducted using systems mounted on vehicles to efficiently cover large areas. Ensuring the reliability of these methods under various environmental conditions is essential for obtaining trustworthy data.

The collected data is then processed and analyzed using deep learning models, involving data cleaning, feature extraction, and defect classification. CNNs play a pivotal role in identifying pavement conditions. The final stage integrates the processed data into a PMS for visualization, reporting, and informed decision-making regarding maintenance and resource allocation.

To develop a network-level PMS using CNNs for defect identification, a robust and comprehensive dataset structure is essential. The dataset should include pavement inventory data such as identifiers for road segments, GPS coordinates, road type, surface type, and lane information. Additionally, condition data is crucial and should comprise inspection dates, detailed records of surface distresses (e.g., cracks, potholes, patches), and roughness measurements.

A rich repository of imagery and associated metadata, including images with capture date, time, and location, is also necessary. To facilitate the development and training of CNN models, the dataset must feature a diverse array of labeled data, covering various pavement conditions and defect types to ensure effective model generalization. Furthermore, geographical information system (GIS) data is required, with layers for the road network and pavement conditions, accompanied by metadata on data sources and accuracy.

Implementing image preprocessing techniques significantly enhances the quality of input data for CNN models in automated pavement data collection systems. Depending on how the image was captured, portions of it may not contain the pavement, necessitating preprocessing. This could be minimized by using a camera focused on the pavement. However, if only one camera is used for the entire road inventory, preprocessing is essential to exclude non-pavement elements like the sky, sidewalks, or vegetation.

Defining the dimensions and optimal location for cropping the image aids in the analysis and extraction of defect information. As studies have shown, at a minimum, the upper part of the image, which predominantly contains the sky, should be excluded to reduce image complexity and facilitate defect detection.

The study proposed using action camera-based devices and/or smartphones, which are accessible and economical, for data collection, offering low-cost alternatives to traditional methods. Despite variations between IGG and SPS values, using data derived from the SPS index supports network-level PMS initiatives by identifying roads with significant degradation and defects, thus enabling prioritized maintenance. Automated assessments produce comprehensive defect reports and maps quickly, aiding decision-making processes and providing a viable alternative to conventional evaluations.

The integration of automated systems using CNNs for pavement condition assessment offers substantial benefits in terms of accuracy, efficiency, and cost-effectiveness. By employing action cameras and smartphones for data collection, the study demonstrates a viable, low-cost alternative to traditional methods. The continuous improvement of these systems through regular updates, model recalibration, and the inclusion of diverse datasets will further enhance their reliability and applicability.

Collaborations between academic and highway management agencies are crucial for developing and maintaining these advanced systems. The research underscores the potential of automated assessments to revolutionize pavement management by providing rapid, accurate, and comprehensive evaluations. Ultimately, these advancements aid in the effective maintenance and extension of pavement service life.

REFERENCES

AI SINGAPORE. **Illustrated: 10 CNN Architectures**. [*S. l.*], 2019. Available from: https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d. Access in: 15 June 2024.

AN, Kwang Eun *et al.* Detecting a pothole using deep convolutional neural network models for an adaptive shock observing in a vehicle driving. **2018 IEEE International Conference on Consumer Electronics, ICCE 2018**, [*s. l.*], 2018.

ANTHONY, Evania Joycelin; KUSNADI, Regina Anastasia. Computer Vision for Supporting Visually Impaired People: A Systematic Review. **Engineering, MAthematics and Computer Science (EMACS) Journal**, [*s. l.*], v. 3, n. 2, p. 65-71, 2021.

ANTONIO, Jorge; ALAVA, Párraga-´. Computer Vision and Medical Image Processing: A brief survey of application areas. [*s. l.*], p. 152-159, 2015.

ANTT. **PROGRAMA DE EXPLORAÇÃO DA RODOVIA (PER)**. Brasilia: [*s. n.*], 2018.

APARNA *et al.* Convolutional neural networks based potholes detection using thermal imaging. **Journal of King Saud University – Computer and Information Sciences**, [*s. l.*], 2019. Available from: https://doi.org/10.1016/j.jksuci.2019.02.004. Access in: 15 June 2024.

APEAGYEI, Alex; ADEMOLAKE, Toyosi Elijah; ADOM-ASAMOAH, Mark. Evaluation of deep learning models for classification of asphalt pavement distresses. **International Journal of Pavement Engineering**, [*s. l.*], v. 24, n. 1, 2023.

ASCE. 2021 **Report Card for America's Infrastructure**. [*s. l.*], 2021. Available from: https://infrastructurereportcard.org/. Access in: 15 June 2024.

ASLAN, O. D. *et al.* Using artifical intelligence for automating pavement condition assessment. **International Conference on Smart Infrastructure and Construction 2019, ICSIC 2019: Driving Data-Informed Decision-Making**, [*s. l.*], v. 2019, n. 2018, p. 337-341, 2019.

ASTM - E 1777 – 96. **Standard Guide for Prioritization of Data Needs for Pavement Management 1**. [*S. l.: s. n.*], 1990.

BARLOW, H. B. Unsupervised Learning. **Neural Computation**, [*s. l.*], v. 1, n. 3, p. 295-311, 1989.

BELTZUNG, Benjamin *et al.* Deep learning for studying drawing behavior: A review. **Frontiers in Psychology**, [*s. l.*], v. 14, n. February, p. 1-13, 2023.

BERNUCCI, Liedi Bariani *et al.* **Pavimentação Asfáltica**: Formação Básica para Engenheiros. 2. ed. Rio de Janeiro: [*s. n.*], 2022.

BISHOP, Christopher M. **Pattern Recognition and Machine Learning**. [*S. l.*]: Springer Science+Business Media, LLC, 2006-. ISSN 27887669.v. 4

BOCHKOVSKIY, Alexey; WANG, Chien Yao; LIAO, Hong Yuan Mark. YOLOv4: Optimal Speed and Accuracy of Object Detection. **arXiv**, [*s. l.*], n. May, 2020.

BROSNAN, Tadhg; SUN, Da Wen. Improving quality inspection of food products by computer vision: a review. **Journal of Food Engineering**, [*s. l.*], v. 61, n. 1 spec., p. 3-16, 2004.

CANZIANI, Alfredo; CULURCIELLO, Eugenio; PASZKE, Adam. ANANALYSIS OF DEEP NEURAL NETWORK MODELS. [*s. l.*], p. 1-7, 2017.

CAO, Wenming; LIU, Qifan; HE, Zhiquan. Review of pavement defect detection methods. **IEEE Access**, [*s. l.*], v. 8, p. 14531-14544, 2020.

CAUSIM, Patrícia Bolsonaro. **Estudo de um sistema de gerência de pavimentos para cidades de pequeno e medio porte**. 2001. 128 f. Universidade Estadual de Campinas, Campinas, 2001.

CHA, Young-Jin; CHOI, Wooram; BÜYÜKÖZTÜRK, Oral. Deep learning-based crack damage detection using convolutional neural networks. **Computer-Aided Civil and Infrastructure Engineering**, [*s. l.*], v. 32, p. 361-378, 2017.

CHEN, Fu-Chen; JAHANSHAHI, Mohammad R. NB-CNN: deep learning-based crack detection using convolutional neural network and naive bayes data fusion. **IEEE Transactions on Industrial Electronics**, [*s. l.*], v. 65, n. 5, p. 4392-4400, 2018.

CHEN, Liang Chieh *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [*s. l.*], v. 40, n. 4, p. 834-848, 2018.

CHINCHOR, Nancy. MUC-4 evaluation metrics. **4th Message Understanding Conference, MUC 1992 – Proceedings**, [*s. l.*], p. 22-29, 1992.

CHOLLET, François. Xception: Deep learning with depthwise separable convolutions. **Proceedings – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017**, [*s. l.*], v. 2017-Janua, p. 1800-1807, 2017.

CNT. **Infraestrutura de transporte**: investimento e financiamento de longo prazo. Brasília: [*s. n.*], 2021. Available from: https://www.cnt.org.br/pesquisas. Access in: 15 June 2024.

CNT. **Plano CNT de transporte e logística 2018**. Brasília: [*s. n.*], 2018.

CNT; SEST SENAT. **Pesquisa CNT de rodovias 2022**. Brasília: [*s. n.*], 2022.

CNT; SEST SENAT. **Pesquisa CNT de rodovias 2023**. Brasília: [*s. n.*], 2023.

COENEN, Tom B. J.; GOLROO, Amir. A review on automated pavement distress detection methods. **Cogent Engineering**, [*s. l.*], v. 4, n. 1, p. 1-23, 2017. Available from: http://doi.org/10.1080/23311916.2017.1374822. Access in: 15 June 2024.

COLEMAN, Cody *et al.* DAWNBench: an end-to-end deep learning benchmark and competition. *In*: Long Beach, CA, USA. **31st Conference on Neural Information Processing Systems**. Long Beach, CA, USA: [*s. n.*], 2017.

DADASHOVA, Bahar; DOBROVOLNY, Chiara Silvestri; TABESH, Mahmood. **Detecting pavement distresses using crowdsourced dashcam Camera ImagesSafe-D National UTC**. [*S. l.: s. n.*], 2021.

DASH, Tirtharaj; NAYAK, Tanistha. English Character Recognition using Artificial Neural Network. **CoRR**, [*s. l.*], v. abs/1306.4, p. 7-9, 2013.

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **NORMA DNIT 006/2003 - PRO**: avaliação objetiva da superfície de pavimentos flexíveis e semi-rígidos: procedimento. Rio de Janeiro: DNIT, 2003b.

DEPARTAMENTO NACIONAL DE INFRAESTRUTURA DE TRANSPORTES (Brasil). **Instrução de Serviço/DG nº 10, de 11 de setembro de 2017**: Dispõe sobre a rotina de procedimentos para a realização do levantamento de campo do Índice de Condição da Manutenção (ICM) das rodovias federais brasileiras. Rio de Janeiro: DNIT, 2017. Available from: https://www.gov.br/dnit/pt-br/central-de-conteudos/instrucoes-normativas/instrucoes-de-servicos/2017/instrucao-de-servico-no-10-2017-colegiada-indice-de-condicao-da-manutencao-icm.pdf. Access in: 15 June 2024.

DING, Lei *et al.* MP-ResNet: Multi-path residual network for the semantic segmentation of high-resolution PolSAR images. **arXiv**, [*s. l.*], p. 1-5, 2020.

DU, Yuchuan *et al.* Pavement distress detection and classification based on YOLO network. **International Journal of Pavement Engineering**, [*s. l.*], v. 22, n. 13, p. 1659-1672, 2021. Available from: https://doi.org/10.1080/10298436.2020.1714047.

EISENBACH, Markus *et al.* How to get pavement distress detection ready for deep learning? A systematic approach. **Proceedings of the International Joint Conference on Neural Networks**, [*s. l.*], v. 2017-May, p. 2039-2047, 2017.

ESPÍNDOLA, Aline Calheiros; FREITAS, Gabriel Tavares de Melo; NOBRE JÚNIOR, Ernesto Ferreira. Pothole and patch detection on asphalt pavement using deep convolutional neural network. *In*: IBERO-LATIN-AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 42.; PAN-AMERICAN CONGRESS ON COMPUTATIONAL MECHANICS, 3., 2021, Rio de Janeiro. **Proceedings** [...]. Rio de Janeiro: ABMEC-IACM, 2021. Available from: https://repositorio.ufc.br/handle/riufc/63680. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; NOBRE JÚNIOR, Ernesto Ferreira; SILVA JÚNIOR, Elias Teodoro da. Pavement surface type classification based on deep learning to the automatic pavement evaluation system. *In*: IBERO-LATIN-AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 42.; PAN-AMERICAN CONGRESS ON COMPUTATIONAL MECHANICS, 3., 2021, Rio de Janeiro. **Proceedings** [...]. Rio de Janeiro: ABMEC-IACM, 2021. Available from: https://repositorio.ufc.br/handle/riufc/63676. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; RAHMAN, Mujib; MATHAVAN, Senthan; NOBRE JÚNIOR, Ernesto Ferreira. Comparing different deep learning architectures as vision-based multi-label classifiers for identification of multiple distresses on asphalt pavement. **Transportation Research Record**: Journal of the Transportation Research Board, [New York], v. 2677, n. 5, p. 24-39, May 2023a. DOI: https://doi.org/10.1177/036119812227273. Available from: https://journals.sagepub.com/doi/abs/10.1177/03611981221127273. Access in: 15 June 2024.

ESPÍNDOLA, Aline Calheiros; RAHMAN, Mujib; MATHAVAN, Senthan; NOBRE JÚNIOR, Ernesto Ferreira. Low-cost automated pavement diagnostic system to network pavement management systems. *In*: ANNUAL MEETING OF THE TRANSPORTATION REVIEW BOARD, 102., 2023, Washington, DC. **Proceedings** [...]. Washington, DC: Transportation Research, 2023b. Available from: https://annualmeeting.mytrb.org/Online ProgramArchive/Details/19122. Access in: 15 June 2024.

ESPOSITO, Floriana; MALERBA, Donato; LISI, Francesca A. Machine learning for intelligent processing of printed documents. **Journal of Intelligent Information Systems**, [*s. l.*], v. 14, n. 2-3, p. 175-198, 2000.

FERNANDES JUNIOR, Jose Leomar. **Sistemas de gerência de pavimentos urbanos para cidades de médio porte**. São Carlos, SP: Universidade de São Paulo, 2001.

GAMAGE, Deshan; PASINDU, H. R.; BANDARA, Saman. Pavement roughness evaluation method for low volume roads. **8th International Conference on Maintenance and Rehabilitation of Pavements, MAIREPAV 2016**, [*s. l.*], n. July, p. 976-985, 2016.

GARG, Shree *et al.* Behaviour analysis of machine learning algorithms for detecting P2P botnets. **2013 15th International Conference on Advanced Computing Technologies, ICACT 2013**, [*s. l.*], p. 1-4, 2013.

GEIGER, A. *et al.* Vision meets robotics: the KITTI dataset. **The International Journal of Robotics Research**, [*s. l.*], n. October, p. 1-6, 2013.

GHOSH, Rohit; SMADI, Omar. Automated detection and classification of pavement distresses using 3D pavement surface images and deep learning. **Transportation Research Record**, [*s. l.*], v. 2675, n. 9, p. 1359-1374, 2021.

GIRSHICK, Ross *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, [*s. l.*], p. 580-587, 2014.

GONG, Haitao; WANG, Feng. Pavement image data set for deep learning: a synthetic approach. **International Airfield and Highway Pavements Conference 2021**, [*s. l.*], p. 253-263, 2021.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento digital de imagens**. 3. ed. [*S. l.*]: Pearson Prentice Hall, 2010.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [*S. l.*]: MIT Press, 2016. Available from: https://www.deeplearningbook.org/. Access in: 15 June 2024.

GOPALAKRISHNAN, Kasthurirangan *et al.* Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. **Construction and Building Materials**, [*s. l.*], v. 157, p. 322-330, 2017. Available from: https://doi.org/10.1016/j.conbuildmat.2017.09.110. Access in: 15 June 2024.

GOPALAKRISHNAN, Kasthurirangan. Deep learning in data-driven pavement image analysis and automated distress detection: a review. **Data**, [*s. l.*], v. 3, n. 3, 2018.

HAAS, R.; HUDSON, W. R.; ZANIEWSKI, J. **Modern pavement management**. Malabar, FL: Krieger Publishing Company, 1994.

HAAS, Ralph; HUDSON, W. Ronald; FALLS, Lynne Cowe. **Pavement asset management**. New Jersey: [*s. n.*], 2015.

HARA, Sunao; KOBAYASHI, Shota; ABE, Masanobu. Sound collection systems using a crowdsourcing approach to construct sound map based on subjective evaluation. **2016 IEEE International Conference on Multimedia and Expo Workshop, ICMEW 2016**, [*s. l.*], p. 1-6, 2016.

HARTMANN, A.; DEWULF, G. Contradictions in infrastructure management – the introduction of performance-based contracts at the Dutch Highways and Waterways Agency. **2009 Second International Conference on Infrastructure Systems and Services: Developing 21st Century Infrastructure Networks (INFRA)**, [*s. l.*], p. 1-5, 2009.

HE, Kaiming *et al.* Deep residual learning for image recognition. *In*: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [*S. l.: s. n.*], 2016. p. 770-778. Available from: http://image-net.org/challenges/LSVRC/2015/. Access in: 15 June 2024.

HE, Kaiming *et al.* Mask R-CNN. *In*: **Proceedings of the IEEE international conference on computer vision**. [*S. l.: s. n.*], 2017. p. 2961-2969.

HEATON, Jeff. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: deep learning. **Genetic Programming and Evolvable Machines**, [*s. l.*], v. 19, n. 1-2, p. 305-307, 2018.

HECHT-NIELSEN, Robert. **III.3 – Theory of the backpropagation neural network**. [*S. l.*]: Academic Press, 1992. Available from: https://doi.org/10.1016/B978-0-12-741252-8.50010-8. Access in: 15 June 2024.

HOANG, Nhat-duc; NGUYEN, Quoc-lam; TRAN, Van-duc. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. **Automation in Construction**, [*s. l.*], v. 94, n. July, p. 203-213, 2018.

HUANG, Gao *et al.* Densely connected convolutional networks. **Proceedings – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017**, [*s. l.*], v. 2017-Janua, p. 2261-2269, 2017.

HUVAL, Brody *et al.* **An empirical evaluation of deep learning on highway driving**. [*s. l.*], p. 1-7, 2015. Available from: http://arxiv.org/abs/1504.01716. Access in: 15 June 2024.

IKRAM, Chaudhry Rehan. **A benchmark for evaluating deep learning based image analytics**. Oslo: University of Oslo, 2019.

IOFFE, Sergey; SZEGEDY, Christian. Batch normalization: accelerating deep network training by reducing internal covariate shift. **32nd International Conference on Machine Learning, ICML 2015**, [*s. l.*], v. 1, p. 448-456, 2015.

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **Manual de gerência de pavimentos**. Rio de Janeiro: DNIT, 2011. (IPR. Publ. 745).

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **NORMA DNIT 005/2003 - TER**: defeitos nos pavimentos flexíveis e semi-rígidos: terminologia. Rio de Janeiro: DNIT, 2003a.

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **NORMA DNIT 006/2003 - PRO**: avaliação objetiva da superfície de pavimentos flexíveis e semi-rígidos: procedimento. Rio de Janeiro: DNIT, 2003b.

INSTITUTO DE PESQUISAS RODOVIÁRIAS; DEPARTAMENTO NACIONAL DE INFRA-ESTRUTURA DE TRANSPORTES (Brasil). **NORMA DNIT 008/2003 - PRO**: levantamento visual contínuo para avaliação da superfície de pavimentos flexíveis e semi-rígidos: procedimento. Rio de Janeiro: DNIT, 2003c.

JABBAR, Haider Khalaf; KHAN, Rafiqul Zaman. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *In*: **Computer Science, Communication and Instrumentation Devices**. [*S. l.: s. n.*], 2014. p. 163-172.

JUNG, Franck. Detecting building changes from multitemporal aerial stereopairs. **ISPRS Journal of Photogrammetry and Remote Sensing**, [*s. l.*], v. 58, n. 3-4, p. 187-201, 2004.

KAELBLING, Leslie Pack; LITTMAN, Michael L.; MOORE, Andrew W. Reinforcement learning: a survey. **Journal of Artificial Intelligence Research**, [*s. l.*], n. 4, p. 237-285, 1996. Available from: https://www.jair.org/index.php/jair/article/view/10166. Access in: 15 June 2024.

KARIM, Raimi. Illustrated: 10 CNN Architectures – a compiled visualisation of the common convolutional neural networks. *In*: **Towards Data Science**. [*S. l.: s. n.*], 2019.

KERAS. **Keras Documentation**. [*S. l.*], 2020.

KHAN, Asharul Islam; AL-HABSI, Salim. Machine learning in computer vision. **Procedia Computer Science**, [*s. l.*], v. 167, n. 2019, p. 1444-1451, 2020. Available from: https://doi.org/10.1016/j.procs.2020.03.355. Access in: 15 June 2024.

KOEHN, Philipp. **Combining genetic algorithms and neural networks**: the encoding problem. 1994. 167 f. University of Tennessee, [*s. l.*], 1994.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. **Advances in Neural Information Processing Systems 25**, [*s. l.*], p. 1097-1105, 2012. Available from: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural- networks.pdf. Access in: 15 June 2024.

KULKARNI, Amruta; PARIKH, Devi; ABBOTT, A Lynn. **Classification of faults in railway ties using computer vision and machine learning**. [*S. l.*], 2017. Available from: https://vtechworks.lib.vt.edu/handle/10919/86522. Access in: 15 June 2024.

LECLERC, Maxime *et al.* Ship classification using deep learning techniques for maritime target tracking. **2018 21st International Conference on Information Fusion, FUSION 2018**, [*s. l.*], p. 737-744, 2018.

LECUN, Y.; BOTTOU, L.; ORR, G. B.; MÜLLER, K.-R. Efficient BackProp. *In*: ORR, G. B.; MÜLLER, K.-R. (ed.). **Neural networks**: tricks of the trade. Heidelberg: Springer, 1998. p. 9-50.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, [New York], v. 86, n. 11, p. 2278-2324, Nov. 1998. Available from: https://ieeexplore.ieee.org/document/726791. Access in: 15 June 2024.

LI, Baoxian *et al.* Automatic classification of pavement crack using deep convolutional neural network. **International Journal of Pavement Engineering**, [*s. l.*], v. 0, n. 0, p. 1-7, 2018. Available from: https://doi.org/10.1080/10298436.2018.1485917. Access in: 15 June 2024.

LI, Fei-Fei. **CS231n Convolutional Neural Networks for Visual Recognition**. [*S. l.*], 2024. Available from: https://cs231n.github.io/convolutional-networks/. Access in: 15 June 2024.

LI, Fei-Fei; GAO, Ruohan; LI, Yunzhu. **CS231n**: deep learning for computer vision – Lecture 1: Overview. [*s. l.*], p. 1-68, 2022.

LI, Qingguang *et al.* A real-time 3D scanning system for pavement distortion inspection. **Measurement Science and Technology**, [*s. l.*], v. 21, n. 1, 2010.

LI, Qingquan *et al.* FoSA: F * Seed-growing Approach for crack-line detection from pavement images. **Image and Vision Computing journal**, [*s. l.*], v. 29, n. 12, p. 861-872, 2011. Available from: https://doi.org/10.1016/j.imavis.2011.10.003. Access in: 15 June 2024.

LI, Yishun *et al.* RoadID: a dedicated deep convolutional neural network for multipavement distress detection. **Journal of Transportation Engineering, Part B**: Pavements, [*s. l.*], v. 147, n. 4, p. 1-12, 2021.

LIN, Andrew *et al.* Artificial intelligence in cardiovascular imaging for risk stratification in coronary artery disease. **Radiology: Cardiothoracic Imaging**, [*s. l.*], v. 3, n. 1, 2021.

LIN, Tsung Yi *et al.* Focal Loss for Dense Object Detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [*s. l.*], v. 42, n. 2, p. 318-327, 2020.

LIU, Wei *et al.* SSD: Single shot multibox detector. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*], v. 9905 LNCS, p. 21-37, 2016.

LONG, Xiang *et al.* PP-YOLO: An effective and efficient implementation of object detector. **arXiv**, [*s. l.*], 2020.

MA, Na *et al.* Wheat seed detection and counting method based on improved YOLOv8 Model. **Sensors**, [*s. l.*], v. 24, n. 5, 2024.

MAGALHÃES, Fabrício Helder Mareco. **Procedimentos para estimar a irregularidade longitudinal do pavimento por meio de veículos calibrados, utilizando-se dados de aceleração vertical obtidos de smartphones**. 2019. 105 f. Universidade Federal do Ceará, [*s. l.*], 2019.

MATIAS, André Victória *et al.* What is the state of the art of computer vision-assisted cytology? A Systematic Literature Review. **Computerized Medical Imaging and Graphics**, [*s. l.*], v. 91, 2021.

MENEGAZZO, Jeferson; WANGENHEIM, Aldo Von. Multi-contextual and multi-aspect analysis for road surface type classification through inertial sensors and deep learning. **IEEE X Brazilian Symposium on Computing Systems Engineering (SBESC)**, [*s. l.*], 2020.

MURPHY, Kevin P. **Machine learning**: a probabilistic perspective. [*S. l.: s. n.*], 2012.

NG, Andrew Y. Feature selection, L1 vs. L2. **Twenty-first international conference on Machine learning – ICML '04**, [*s. l.*], p. 78-85, 2004. Available from: http://portal.acm.org/citation.cfm?doid=1015330.1015435. Access in: 15 June 2024.

NG, Andrew. **Machine learning**. [*S. l.: s. n.*], 2012. Available from: https://www.coursera.org/learn/machine-learning. Access in: 15 June 2024.

NGUYEN, Y. H. **A beginner's guide to gradient descent in machine learning**. [*S. l.*], 2023. Available from: https://medium.com/@yennhi95zz/4-a-beginners-guide-to-gradient-descent-in-machine-learning-773ba7cd3dfe. Access in: 15 June 2024.

NHAT-DUC, Hoang; NGUYEN, Quoc Lam; TRAN, Van Duc. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. **Automation in Construction**, [*s. l.*], v. 94, n. July, p. 203-213, 2018.

NIE, Mingxin; WANG, Kun. Pavement distress detection based on transfer learning. **2018 5th International Conference on Systems and Informatics, ICSAI 2018**, [*s. l.*], n. Icsai, p. 435-439, 2018.

NIELSEN, M. **Neural networks and deep learning**. [*S. l.*]: Determination Press, 2015.

NOGUEIRA, Ana Filipa Rodrigues *et al.* Transformers for urban sound classification: a comprehensive performance evaluation. **Sensors**, [*s. l.*], v. 22, n. 8874, p. 1-17, 2022.

O'MAHONY, Niall *et al.* Deep Learning vs. Traditional Computer Vision. **Advances in Intelligent Systems and Computing**. [*s. l.*], v. 943, n. Cv, p. 128-144, 2019.

O'MAHONY, Niall *et al.* **Deep Learning vs. Traditional Computer Vision**. [*S. l.*]: Springer International Publishing, 2020. Available from: https://doi.org/10.1007/978-3-030-17795-9_10. Access in: 15 June 2024.

O'SHEA, Keiron; NASH, Ryan. An introduction to convolutional neural networks. **International Journal for Research in Applied Science and Engineering Technology**, [*s. l.*], 2015.

PAN, Sinno Jialin; YANG, Qiang. A survey on transfer learning. **IEEE Transactions on Knowledge and Data Engineering**, [*s. l.*], v. 22, n. 10, p. 1345-1359, 2010.

PAPAGEORGIOU, Constantine; POGGIO, Tomaso. Trainable system for object detection. **International Journal of Computer Vision**, [*s. l.*], v. 38, n. 1, p. 15-33, 2000.

PATEL, S. **Understanding Convolutional Neural Networks (CNN)**. [*S. l.*], 2023. Available from: https://learnopencv.com/understanding-convolutional-neural-networks-cnn/. Access in: 15 June 2024.

PATERSON, William D. O.; SCULLION, Thomas. Information systems for road management: draft guidelines on system design and data issues. **Infrastructure and Urban Development Department Report**, [*s. l.*], v. INU 77, n. Washingto, p. The World Bank, 1990.

PAULYA, Leo *et al.* Deeper networks for pavement crack detection. *In*: Taipei, Taiwan. **International Symposium on Automation and Robotics in Construction**. Taipei, Taiwan: [*s. n.*], 2017. p. 479-485.

PEREIRA, Vosco *et al.* Classification of paved and unpaved road image using convolutional neural network for road condition inspection system. **ICAICTA 2018 – 5th International Conference on Advanced Informatics: Concepts Theory and Applications**, [*s. l.*], p. 165-169, 2018.

PEREZ, Luis; WANG, Jason. **The effectiveness of data augmentation in image classification using deep learning**. [*S. l.*], 2017. Available from: http://arxiv.org/abs/1712.04621. Access in: 15 June 2024.

PHUNG, Van Hiep; RHEE, Eun Joo. A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. **Applied Sciences (Switzerland)**, [*s. l.*], v. 9, n. 21, 2019.

PIERCE, Linda M.; MCGOVERN, Ginger; ZIMMERMAN, Kathryn A. Practical guide for quality management of pavement condition data collection. **Practical Guide for Quality Management of Pavement Condition Data Collection**, [*s. l.*], p. 170, 2013. Available from: https://www.fhwa.dot.gov/pavement/management/qm/data_qm_guide.pdf. Access in: 15 June 2024.

POWERS, David M. W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. **International Journal of Machine Learning Technology**, [*s. l.*], p. 37-63, 2011. Available from: http://arxiv.org/abs/2010.16061. Access in: 15 June 2024.

RAGNOLI, Antonella; DE BLASIIS, Maria Rosaria; DI BENEDETTO, Alessandro. Pavement distress detection methods: a review. **Infrastructures**, [*s. l.*], v. 3, n. 4, p. 1-19, 2018.

RANJBAR, Sajad; NEJAD, Fereidoon Moghadas; ZAKERI, H. An image-based system for pavement crack evaluation using transfer learning and wavelet transform. **International Journal of Pavement Research and Technology**, [*s. l.*], p. 1-13, 2020.
RATEKE, Thiago; JUSTEN, Karla Aparecida; WANGENHEIM, Aldo von. Road surface classification with images captured from low-cost camera-road traversing knowledge (RTK) dataset. **Revista de Informatica Teorica e Aplicada**, [*s. l.*], v. 26, n. 3, p. 50-64, 2019.

REDMON, Joseph; FARHADI, Ali. YOLOv3: An incremental improvement. **arXiv**, [*s. l.*], 2018.

REN, Shaoqing *et al.* Faster R-CNN: towards real-time object detection with region proposal networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [*s. l.*], v. 39, n. 6, p. 1137-1149, 2017.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: convolutional networks for biomedical image segmentation. **International Conference on Medical image computing and computer-assisted intervention**, [*s. l.*], p. 234-241, 2015.

RUSSAKOVSKY, Olga *et al.* ImageNet large scale visual recognition challenge. **International Journal of Computer Vision**, [*s. l.*], p. 211-252, 2015.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência artificial**: uma abordagem moderna. 4. ed. [*S. l.*: *s. n.*], 2022.

SAPONARA, Sergio; ELHANASHI, Abdussalam. Impact of image resizing on deep learning detectors for training time and model performance. *In*: SAPONARA, S.; DE GLORIA, A. (ed.). **Applications in electronics pervading industry, environment and society**. ApplePies 2021. Lecture Notes in Electrical Engineering, [*s. l.*], v. 866, 2022.

SCHERER, D.; MULLER, A.; BEHNKE, S. Evaluation of pooling operations in convolutional architectures for object recognition. *In*: **Artificial Intelligence and Lecture Notes in Bioinformatics**. [*S. l.*: *s. n.*], 2010. p. 92-101.

SHAHIN, M. Y. **Pavement management for airports, roads and parking lots**. New York: [*s. n.*], 2002.

SHINZATO, Patrick Y. *et al.* CaRINA dataset: An emerging-country urban scenario benchmark for road detection systems. **IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC**, [*s. l.*], p. 41-46, 2016.

SHOLEVAR, Nima; GOLROO, Amir; ESFAHANI, Sahand Roghani. Machine learning techniques for pavement condition evaluation. **Automation in Construction**, [*s. l.*], v. 136, n. March, p. 104190, 2022. Available from: https://doi.org/10.1016/j.autcon.2022.104190. Access in: 15 June 2024.

SIMONYAN, Karen; ZISSERMAN, Andrew. Two-stream convolutional networks for action recognition in videos. **Advances in Neural Information Processing Systems**, [*s. l.*], v. 1, n. January, p. 568-576, 2014.

SIRHAN, Mai; BEKHOR, Shlomo; SIDESS, Arieh. Multilabel CNN Model for Asphalt Distress Classification. **Journal of Computing in Civil Engineering**, [*s. l.*], v. 38, n. 1, p. 1-7, 2024.

SITARZ, Mikołaj. Extending F1 Metric, Probabilistic Approach. **Advances in Artificial Intelligence and Machine Learning**, [*s. l.*], v. 3, n. 2, p. 1025-1038, 2023.

SMITH, Leslie N. Cyclical learning rates for training neural networks. **Proceedings – 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017**, [*s. l.*], n. April, p. 464-472, 2017.

SOKOLOVA, Marina; LAPALME, Guy. A systematic analysis of performance measures for classification tasks. **Information Processing and Management**, [*s. l.*], v. 45, n. 4, p. 427-437, 2009. Available from: https://doi.org/10.1016/j.ipm.2009.03.002. Access in: 15 June 2024.

SONG, Liang; WANG, Xuancang. Faster region convolutional neural network for automated pavement distress detection. **Road Materials and Pavement Design**, [*s. l.*], v. 22, n. 1, p. 23-41, 2021. Available from: https://doi.org/14680629.2019.1614969. Access in: 15 June 2024.

SRINIDHI, G.; DEVI, Renuka. **Pothole Detection using CNN and AlexNet**. [*S. l.*], 2020.

SRIVASTAVA, Nitish *et al.* Dropout: a simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, [*s. l.*], v. 15, p. 1929-1958, 2014.

SZEGEDY, Christian *et al.* Rethinking the inception architecture for computer vision. *In*: **IEEE Conference on Computer Vision and Pattern Recognition**. [*S. l.: s. n.*], 2015. p. 2818-2826.

SZELISK, Richard. **Computer vision**: algorithms and applications. London: Springer, 2010. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20549881. Access in: 15 June 2024.

TAJBAKHSH, Nima *et al.* Convolutional neural networks for medical image analysis: full training or fine tuning? **IEEE Transactions on Medical Imaging**, [New York], v. 35, n. 5, p. 1299-1312, May 2016. DOI: https://doi.org/10.1109/TMI.2016.2535302. Available from: https://ieeexplore.ieee.org/document/7426826. Access in: 15 June 2024.

TAMAGUSKO, Tiago; GOMES CORREIA, Matheus; FERREIRA, Adelino. Machine learning applications in road pavement management: a review, challenges and future directions. **Infrastructures**, [*s. l.*], v. 9, n. 12, 2024.

TUMEN, Vedat; YILDIRIM, Ozal; ERGEN, Burhan. Recognition of road type and quality for advanced driver assistance systems with deep learning. **Elektronika ir Elektrotechnika**, [*s. l.*], v. 24, n. 6, p. 67-74, 2018.

UBIAI. **Data augmentation techniques**. [*S. l.*], 2023. Available from: https://ubiai.tools/what-are-the-advantages-anddisadvantages-of-data-augmentation-2023-update/. Access in: 15 June 2024.

VARONA, Braian; MONTESERIN, Ariel; TEYSEYRE, Alfredo. A deep learning approach to automatic road surface monitoring and pothole detection. **Personal and Ubiquitous Computing**, [*s. l.*], v. 24, n. 4, p. 519-534, 2020.

VOULODIMOS, Athanasios *et al.* Deep learning for computer vision: a brief review. **Computational Intelligence and Neuroscience**, [*s. l.*], v. 2018, 2018.

WANG, Jinjiang *et al.* Deep learning for smart manufacturing: methods and applications. **Journal of Manufacturing Systems**, [*s. l.*], v. 48, p. 144-156, 2018. Available from: https://doi.org/10.1016/j.jmsy.2018.01.003. Access in: 15 June 2024.

WANG, Xianglong; HU, Zhaozheng. Grid-based pavement crack analysis using deep learning. **2017 4th International Conference on Transportation Information and Safety, ICTIS 2017 – Proceedings**, [*s. l.*], p. 917-924, 2017.

WHITE, David *et al.* Error rates in users of automatic face recognition software. **PLoS ONE**, [*s. l.*], v. 10, n. 10, p. 1-14, 2015.

XIA, Wei. An approach for extracting road pavement disease from HD camera videos by deep convolutional networks. **ICALIP 2018 – 6th International Conference on Audio, Language and Image Processing**, [*s. l.*], p. 418-422, 2018.

XIE, Danfeng; ZHANG, Lei; BAI, Li. Deep learning in visual computing and signal processing. **Applied Computational Intelligence and Soft Computing**, [*s. l.*], v. 2017, 2017.

XIE, Saining *et al.* Aggregated residual transformations for deep neural networks. **Proceedings – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017**, [*s. l.*], v. 2017-Janua, p. 5987-5995, 2017.

YAMASHITA, Rikiya *et al.* Convolutional neural networks: an overview and application in radiology. **Insights Imaging Springer**, [*s. l.*], v. 9, p. 611-629, 2018.

YE, Wanli *et al.* Convolutional neural network for pothole detection in asphalt pavement. **Road Materials and Pavement Design**, [*s. l.*], v. 0, n. 0, p. 1-17, 2019. Available from: https://doi.org/14680629.2019.1615533. Access in: 15 June 2024.

YUSOF, N. A. M. *et al.* Crack detection and classification in asphalt pavement images using deep convolution neural network. **Proceedings – 8th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2018**, [*s. l.*], p. 227-232, 2018.

ZALAMA, Eduardo; JAIME, G.; MEDINA, Roberto. Road crack detection using visual features extracted by gabor filters. **Computer-Aided Civil and Infrastructure Engineering**, [*s. l.*], v. 29, p. 342-358, 2014.

ZEILER, Matthew D.; FERGUS, Rob. Visualizing and understanding convolutional Networks. **Springer International Publishing Switzerland**, [*s. l.*], v. 8689 LNCS, n. PART 1, p. 818-833, 2014.

ZHANG, Allen A. *et al.* Intelligent pavement condition survey: overview of current researches and practices. **Journal of Road Engineering**, [*s. l.*], v. 4, n. 3, p. 257-281, 2024. Available from: https://doi.org/10.1016/j.jreng.2024.04.003. Access in: 15 June 2024.

ZHANG, Lei *et al.* Road crack detection using deep convolutional neural network. **Proceedings – International Conference on Image Processing, ICIP**, [*s. l.*], v. 2016-Augus, p. 3708-3712, 2016.

ZHAO, Hengshuang *et al.* Pyramid scene parsing network. **Proceedings – 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017**, [*s. l.*], v. 2017-Janua, p. 6230-6239, 2017.

ZHENG, Yi. Evaluation and implementation of convolutional neural networks in image recognition. **Journal of Physics**: Conference Series, [*s. l.*], v. 1087, n. 6, 2018.

ZHOU, Shanglian; SONG, Wei. Deep learning–based roadway crack classification with heterogeneous image data fusion. **Structural Health Monitoring**, [*s. l.*], 2020.

ZHOU, Zhi Hua *et al.* Multi-instance multi-label learning. **Artificial Intelligence**, [*s. l.*], v. 176, n. 1, p. 2291-2320, 2012. Available from: https://doi.org/10.1016/j.artint.2011.10.002. Access in: 15 June 2024.