

People's Democratic Republic of Algeria  
Ministry of Higher Education and Scientific Research



ⵜⴰⵎⴰⵔⵜ ⵏ ⵙⵉⵔⵉⵜ  
جامعة بجاية  
Université de Béjaïa

University A. Mira of Bejaia  
Faculty of Exact Sciences  
Computer Science Department

# THESIS

*For obtaining master's degree in computer science*  
*Project presented in setting of ministerial decree no. 1275*

**Domain:** Mathematics and Informatics      **Specialty:** Computer Science  
**Option:** Artificial Intelligence / Software Engineering

Presented by  
**Rayane AGGOUNE & Chahinez AMRANE**

## Theme

---

**A new deep learning based data pipeline and  
cloud architecture for cervical spine fracture  
detection**

---

Defended on October 1<sup>st</sup>, 2023 in front of the jury composed of:

Mr Hachem SLIMANI	Professor	Univ. of Bejaia	Chair
Mr Fatah BOUCHEBBAH	M.C.B	Univ. of Bejaia	Supervisor
Mr Yani-Athmane BENNAI	M.C.B	Univ. of Bejaia	Examinator
Mr Mourad MAHMOUDI	M.C.B	Univ. of Bejaia	Examinator

Academic Year: 2022/2023

# Acknowledgments

I would like to express my sincere gratitude to all those who have contributed to the successful completion of this research project and the preparation of this report.

First and foremost, I am deeply indebted to my advisor, for his unwavering support, guidance, and invaluable insights throughout this journey. His expertise and encouragement have been instrumental in shaping the direction of this study.

I would also like to extend my appreciation to the faculty members of the Computer Science at Abderrahmane Mira University of Béjaïa for their dedication to imparting knowledge and their willingness to engage in insightful discussions.

My heartfelt thanks go to my friends and classmates who provided moral support, encouragement, and occasional study sessions during late nights. Your camaraderie made this academic endeavour more enjoyable.

I am grateful to my family for their continuous support, understanding, and belief in my abilities. Their unwavering faith in me has been a constant source of motivation.

Lastly, I would like to acknowledge the research participants who generously contributed their time and insights to this study. Without their cooperation, this research would not have been possible.

This report is the culmination of countless hours of hard work and collaboration. I am profoundly thankful to everyone who has been a part of this journey.

Thank you all for your support and encouragement.

Chanez...

# Acknowledgments

I wish to extend my deepest gratitude to my esteemed supervisor, Mr. Bouchebbah, whose invaluable guidance, unwavering encouragement, and insightful critiques have been instrumental in the successful completion of this research. His willingness to invest his expertise and time into my academic and professional development is a gift for which I am profoundly thankful.

To my family—my bedrock of support and love—I owe an immeasurable debt of gratitude. Your unwavering love and support have been my strength throughout this journey and beyond. I would like to specifically acknowledge my sister, whose constant encouragement has been a beacon of light during moments of challenge and doubt. In memoriam, I pay homage to my late grandparents, whose legacy of wisdom and love continues to guide and inspire me. Although they are no longer with us, the principles they instilled in me remain an indelible part of who I am today. I wish to express heartfelt thanks to my BigO and dearest friends, who have stood by me as pillars of support and inspiration. Your words of encouragement and unwavering belief in my capabilities have enriched this experience beyond measure. Your friendship is a treasure that has made this arduous journey not only bearable but truly rewarding.

Lastly, I extend my warm thanks to all those whose contributions may not be overt but have been no less significant. While your roles may not be enumerated here, please know that your impact has been profoundly felt and is deeply appreciated.

Anna...

# Dedications

To my loving family,  
For their unwavering support, encouragement, and belief in my dreams.  
And to all those who inspire me daily,  
This work is dedicated to you.

AMRANE Chahinez

# Dedications

To me and my loving family mom, dad, Ayoub, Souhaib, Aness, the perfect Soundousse, the best Grandma and those who left us...

To dearest teachers and friends Sara, Chanez, Sorore, Feryal, Ziko, Zou, Achref and Noureddine, for their unwavering support and encouragement, and for being a constant source of inspiration.

To my spirit and Ilaf...You have been my rock throughout this journey, and I am forever grateful for your presence in my life.

This work is dedicated to all of you.

Anna

# Contents

<b>Contents</b>	<b>i</b>
<b>Liste of Figures</b>	<b>iii</b>
<b>Liste of Tables</b>	<b>iv</b>
<b>Liste of Abbreviations</b>	<b>v</b>
<b>General introduction</b>	<b>1</b>
<b>1 Generalities on cervical spine fraction and medical imaging</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Anatomy and disease . . . . .	3
1.2.1 Anatomy of a normal cervical spine . . . . .	4
1.2.2 Common cervical spine fractures . . . . .	5
1.3 Statistics on cervical spine fractures . . . . .	6
1.4 Medical imaging techniques for the diagnosis of cervical spine fractures	7
1.5 Treatment of cervical spine fractures . . . . .	8
1.6 Early detection and accurate diagnosis of cervical spine fractures . . .	9
1.7 Current approaches to cervical spine fracture detection . . . . .	9
1.8 Artificial intelligence applications in e-medicine . . . . .	10
1.9 Performance of AI-based cervical spine fracture detection systems . .	11
1.10 Image analysis and computer vision . . . . .	11
1.11 Automatic detection: principle and methods . . . . .	11
1.12 Potential benefits and limitations of AI in cervical spine fracture de- tection . . . . .	12
1.13 Conclusion . . . . .	12
<b>2 State-of-the-art on cervical spine fracture detection</b>	<b>13</b>
2.1 Introduction . . . . .	14
2.2 Description of related work . . . . .	14
2.3 Comparaison of reviewed methods . . . . .	19
2.4 Summary of comparison and discussion . . . . .	20
2.5 Future work perspectives for cervical spine fracture detection . . . . .	21
2.6 Conclusion . . . . .	22
<b>3 A new deep learning model and cloud-based architecture for cer- vical spine fracture detection</b>	<b>23</b>
3.1 Introduction . . . . .	23
3.2 General overview of ViT . . . . .	23
3.2.1 ViT model architecture . . . . .	24

3.2.2	ViT model variants	26
3.3	The object detection model Faster R-CNN	28
3.4	Attention maps	29
3.4.1	Attention rollout	29
3.4.2	Rollout matrices equation	30
3.5	Personal contributions	30
3.5.1	A multifaceted computational pipeline for the detection and visualization of cervical spine fractures	30
3.5.2	Proposed cloud-based system for cervical spine fracture detection	34
3.6	Conclusion	36
<b>4</b>	<b>Tests and evaluations of the proposed data pipeline</b>	<b>38</b>
4.1	Introduction	38
4.2	Dataset description and exploration	39
4.2.1	Exploratory data analysis	39
4.2.2	Dataset's visualization	43
4.2.3	Observations and implications	46
4.3	Data pipeline implementation and training	47
4.3.1	Volumetric image slicing	47
4.3.2	Vertebrae detection using Faster R-CNN and region cropping	48
4.3.3	Data augmentation on cropped vertebrae	55
4.3.4	Next-ViT model implementation and tuning	56
4.4	Model validation	58
4.5	Fracture presentation using attention map	60
4.6	Summary and discussion	61
4.7	Conclusion	62
	<b>General Conclusion</b>	<b>63</b>

# List of Figures

1.1	Anatomy of cervical spine [5]. . . . .	5
1.2	Examples of cervical spine fractures [24]. . . . .	6
1.3	A fracture in the cervical spine detected with x-ray [3]. . . . .	7
1.4	The fracture in cervical spine detected with CT scan [5]. . . . .	8
1.5	A fracture in cervical spine detected with MRI [31]. . . . .	8
3.1	A general representation of the architecture of a transformer [9]. . . .	24
3.2	Transformer encoder and decoder internal layers [23]. . . . .	25
3.3	Multi-head attention mechanism [37]. . . . .	26
3.4	Global architecture of the ViT [14]. . . . .	26
3.5	Faster R-CNN block architecture [13]. . . . .	29
3.6	A representation of the proposed data pipeline. . . . .	34
3.7	A representation of the proposed cloud-based system. . . . .	37
4.1	Kaggle and Python logos. . . . .	39
4.2	NIFTI segmentation reveals cervical spine fractures. . . . .	40
4.3	One of the train images for a patient. . . . .	41
4.4	Train data frame of the cervical spine dataset . . . . .	41
4.5	Train bounding boxes dataframe of the cervical spine dataset . . . . .	42
4.6	Presented fracture in a vertebral fracture for a patient. . . . .	43
4.7	Test dataframe of the cervical spine dataset. . . . .	43
4.8	The resulting correlation matrix. . . . .	44
4.9	Visual representation of relationships between C1,...,C7 exploration. .	45
4.10	A comprehensive representation of the size and shape of the data. . .	46
4.11	An extracted slice of a CT scan alongside its corresponding slice after resize operation. . . . .	47
4.12	A slice after windowing operation. . . . .	48
4.13	Bounding box around vertebrae segmented region. . . . .	48
4.14	Test to see if saved bounding box coordinates are correct. . . . .	50
4.15	Downloading the Faster R-CNN model. . . . .	50
4.16	Detected vertebrae using Faster R-CNN. . . . .	53
4.17	Cropped vertebrae. . . . .	55
4.18	Validation results of Next-ViT. . . . .	60



# List of Tables

2.2	Comparative table of the studied works according to criteria. . . . .	20
3.1	A comparison of the main ViT variants according to some features. .	27
4.1	Performance metrics of the Faster R-CNN model on the training dataset. . . . .	52
4.2	Training Loss results. . . . .	53
4.3	Data augmentation techniques applied in the proposed data pipeline.	55
4.4	Next-ViT model parameter tuning. . . . .	56

# List of Abbreviations

	<i>2D</i>	Two-Dimensional.
	<i>3D</i>	Three-Dimensional.
<b>A</b>	<i>AI</i>	Artificielle Intelligence.
	<i>ASNR</i>	American Society of Neuroradiology.
	<i>ASSR</i>	American Society of Spine Radiology.
	<i>AUC</i>	Area Under the Curve.
<b>B</b>	<i>BLSTM</i>	Bidirectional Long Short-Term Memory.
<b>C</b>	<i>CAM</i>	Class Activation Maps.
	<i>COCO</i>	Common Objects in Context.
	<i>CNN</i>	Convolutional Neural Networks.
	<i>CSWin</i>	Causal Swin Transformer.
	<i>CSPDark – net53</i>	Cross-Stage Partial Networks.
	<i>CT</i>	Computed Tomography.
<b>D</b>	<i>DCM</i>	Degenerative Cervical Myelopathy.
	<i>DCNN</i>	Deep Convolutional Neural Network.
	<i>DeiT</i>	Data-efficient Image Transformers.
	<i>DICOM</i>	Digital Imaging and Communications in Medicine.
<b>E</b>	<i>EDA</i>	Exploratory Data Analysis.
<b>F</b>	<i>FDA</i>	Food and Drug Administration.
	<i>Faster R – CNN</i>	Faster Region-based Convolutional Neural Network.
<b>G</b>	<i>GDCM</i>	Grassroots DICOM.
	<i>GPU</i>	Graphics Processing Unit.
<b>I</b>	<i>ID</i>	Identification.
<b>J</b>	<i>JPEG</i>	Joint Photographic Experts Group.
<b>L</b>	<i>LSTM</i>	Long Short-Term Memory.
<b>M</b>	<i>mAPs</i>	Mean Average Precisions.
	<i>mJOA</i>	Modified Japanese Orthopaedic Association.
	<i>MLP</i>	Multilayer Perceptrons.
	<i>MRI</i>	Magnetic Resonance Imaging.
<b>N</b>	<i>Next – ViT</i>	Next-generation Vision Transformer.
	<i>NIFTI</i>	Neuroimaging Informatics Technology Initiative.
	<i>NLP</i>	Natural Language Processing.
	<i>NPV</i>	Negative Predictive Values.
<b>P</b>	<i>PPV</i>	Positive Predictive Values.
	<i>PVT</i>	Patches for Vision Transformers.
<b>R</b>	<i>RAdam</i>	Rectified Adam.
	<i>R – CNN</i>	Region Convolutional Neural Network.
	<i>ResNet – 50</i>	Residual Network 50.
	<i>ROC</i>	Receiver Operating Characteristic.
	<i>RPN</i>	Region Proposal Network.
	<i>RSNA</i>	Radiological Society of North America.

<b>S</b>	<i>SPPF</i>	Spatial Pyramid Pooling.
<b>T</b>	<i>TP</i>	number of True Positive cases.
	<i>TN</i>	number of True Negative cases.
<b>V</b>	<i>ViT</i>	Vision Transformer.
<b>U</b>	<i>UID</i>	Unique IDentifier.

# General introduction

## Scientific context

In the current age of rapid technological advancement, Artificial Intelligence (AI) is making profound inroads into various industries. In the field of medical science, AI holds immense untapped potential. Among its most promising applications is the transformation of cervical spine fracture diagnosis and treatment. This medical challenge, marked by complex diagnostics and the potential for severe neurological consequences if mishandled, beckons AI as a solution.

Cervical spine fractures, often caused by accidents or falls, pose a unique medical dilemma. These injuries, occurring in a delicate part of the human skeletal structure, demand swift and precise identification to prevent serious neurological damage. AI, with its computational capabilities and ability to decipher intricate medical data patterns, emerges as a key player.

By leveraging technologies like deep learning algorithms and computer vision, AI can not only detect cervical spine fractures but also provide crucial details about their type, location, and severity. This fusion of AI with medical expertise promises to enhance diagnostic accuracy and revolutionize medical imaging and patient care [38].

This research delves into the anatomy of the cervical spine, current AI techniques for fracture detection, data collection and algorithm training, and the development of a visionary AI model. Beyond academia, this work has far-reaching implications, potentially improving patient outcomes through faster and more accurate diagnoses.

## Problematic

Every year, over 1.5 million people in the United States alone sustain spine fractures, a significant proportion of which affect the delicate architecture of the cervical spine [27]. For the elderly and those with pre-existing conditions like osteoporosis, such fractures could be catastrophic. The situation is further complicated by the fact that cervical spine fractures often necessitate immediate attention, yet rapid and accurate diagnosis remains elusive. This paradox forms the crucible in which this research is forged.

## Objective

The primary objective of this research is to develop and validate an advanced AI-driven system for the early and accurate detection of cervical spine fractures, utilizing a combination of deep learning algorithms and cloud computing. This system

aims to enhance diagnostic precision, reduce detection times, and ultimately improve patient outcomes.

## Research methodology

This research project journeys into the confluence of AI and medical science to answer these questions. Rooted in cutting-edge technologies like Vision Transformer (ViT) and Faster R-CNN, and fortified by the computational prowess of cloud-based systems, this work aims to offer an innovative solution for the diagnosis of cervical spine fractures.

## Organization of the manuscript

This manuscript is structured in four chapters. In what follows, we give a brief description of the content of the four chapters.

The first chapter offers an exhaustive analysis of cervical spine anatomy, establishing a fundamental comprehension of the subject matter. This foundation then informs an evaluative discussion of contemporary diagnostic procedures for cervical spine fractures, effectively framing the context for the research at hand.

The second chapter undertakes a rigorous review of extant literature, zeroing in on cutting-edge artificial intelligence methodologies for fracture detection. This literature review serves to delineate the current state of the field and identifies the gaps that our research aims to fill.

The third chapter serves as a detailed exposition of the system architecture and proposed modelling techniques, including innovations such as the Vision Transformer (ViT) and Faster R-CNN. This chapter elucidates the underlying mathematical frameworks and original contributions that substantiate our research, thus forming a pivotal segment of our broader endeavour to advance cervical spine fracture detection.

The fourth chapter marks a pivotal stage in our research trajectory, providing a meticulous evaluation of the model's performance using both quantitative and qualitative metrics. This assessment not only validates the effectiveness of the proposed solutions but also lays the groundwork for future refinements to further augment the accuracy of cervical spine fracture detection.

Finally, we conclude this manuscript with a general conclusion in which we present an all-encompassing recapitulation of the study's significance, findings, limitations, and contributions. This final discussion articulates the transformative capacity of artificial intelligence in the realm of cervical spine fracture detection and serves as the capstone of our research journey in medical imaging.

# Chapter 1

## Generalities on cervical spine fracture and medical imaging

### 1.1 Introduction

This introductory chapter is designed as a comprehensive primer about our research, with the essential concepts and terminologies that are central to the scope of this study. The principal aim of the chapter is to offer an exhaustive exploration of the anatomy of the cervical spine, the various types of fractures associated with it, the current diagnostic methods, and the revolutionary potential of artificial intelligence (AI) in cervical spine fracture detection.

Structurally, the chapter begins with Section 1.2, which focuses on the foundational anatomy of the cervical spine and common fracture types. This is followed by Section 1.3, which presents vital statistics that highlight the far-reaching public health implications of cervical spine fractures, thereby emphasizing the urgent need for precise diagnostic techniques. Section 1.4 provides an overview of existing diagnostic methods such as X-rays, CT scans, and MRIs, while Section 1.5 discusses prevalent treatment options. Section 1.6 underscores the critical nature of timely and accurate diagnosis, setting the stage for the introduction of AI solutions in Section 1.7, which reviews traditional methods and speculates on future AI-based innovations. The transformative impact of AI in the medical diagnosis domain, especially concerning cervical spine fractures, is elaborated in Section 1.8. Section 1.9 offers an evaluation of different AI systems for fracture detection, analyzing their respective strengths and weaknesses. The role of image analysis and computer vision techniques in enhancing diagnostic accuracy is covered in Section 1.10. Section 1.11 demystifies the core principles and algorithms of AI applicable in this context.

The chapter concludes with an assessment of the potential benefits and limitations of AI in cervical spine fracture detection in Section 1.12, and Section 1.13 amalgamates the insights gained, aiming to contribute to the broader medical science discourse and highlight the transformative role of AI in cervical spine fracture diagnosis.

### 1.2 Anatomy and disease

The cervical spine, also known as the neck region of the spine, is a critical part of the human body. It consists of seven vertebrae, denoted as C1 to C7, which are stacked on top of each other (see Figure 1.1 for illustration). These vertebrae are

separated by inter-vertebral discs, acting as cushions, and connected by ligaments, providing stability and facilitating movement. The first two cervical vertebrae, C1 and C2, are particularly distinctive. C1, also called the atlas, holds the weight of the head and allows nodding movements, while C2, known as the axis, permits rotation of the head. Together, they form the foundation for the wide range of motion exhibited by the neck, enabling us to turn, tilt, and flex our heads in various directions. The cervical spine's primary functions include supporting the weight of the head, protecting the spinal cord, and ensuring flexibility. The spinal cord, an essential part of the central nervous system, runs through the vertebral foramen of the cervical vertebrae. It carries nerve signals between the brain and the rest of the body, playing a vital role in coordinating voluntary and involuntary movements. The unique design of the cervical spine allows us to carry out everyday activities with ease, such as looking around, maintaining balance, and engaging in activities that require precise head movements. However, this intricate structure also makes the cervical spine susceptible to injuries and disorders.

One significant concern related to the cervical spine is the occurrence of fractures. Cervical spine fractures can result from traumatic events, such as falls, motor vehicle accidents, or sports-related injuries. These fractures can lead to severe consequences, including spinal cord injuries, nerve damage, and paralysis [21].

### 1.2.1 Anatomy of a normal cervical spine

The cervical spine, also known as the neck region of the spine, consists of seven vertebrae, namely C1 (atlas), C2 (axis), and C3 – C7. These vertebrae are essential for maintaining the structure and function of the neck. C1, also referred to as the atlas, is the uppermost vertebra and plays a crucial role in supporting the head. It forms the connection between the skull and the spine, allowing for smooth movement of the head in various directions. C2, known as the axis, is the second vertebra of the cervical spine. Its unique structure includes a bony projection called the odontoid process or dens, which acts as a pivot point. This feature enables the head to rotate from side to side, facilitating essential movements like turning the head. The remaining vertebrae, C3 to C7, complete the cervical spine. Each of these vertebrae is interconnected by discs and ligaments, providing flexibility and mobility to the neck. They work together to support the weight of the head and maintain the stability of the neck region.

In summary, the seven vertebrae of the cervical spine play distinct roles, from supporting the head (C1) to enabling pivotal head movements (C2) and providing stability and support to the neck and head (C3-C7). The interconnected and coordinated functioning of these vertebrae is vital for maintaining proper posture, movement, and overall neck health [21].

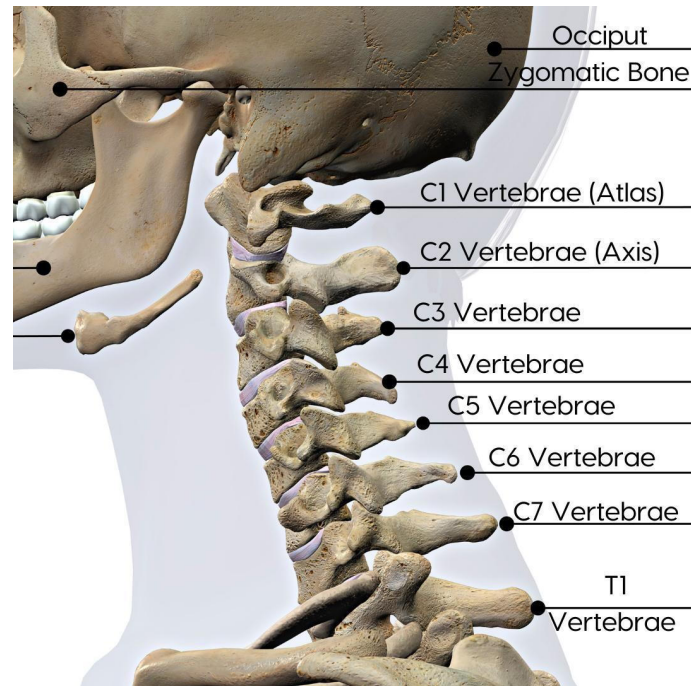


Figure 1.1: Anatomy of cervical spine [5].

### 1.2.2 Common cervical spine fractures

Common cervical spine fractures encompass three main types: compression, flexion-distraction, and burst fractures. Each type presents distinct characteristics and potential complications [22] :

- **Compression fractures:** they occur when the vertebrae are subjected to a compressive force, leading to a decrease in height. This type of fracture may result in pain and difficulty moving the neck. The vertebrae lose their structural integrity due to compression, which can lead to instability in the spine. Additionally, nerve damage may occur, leading to further complications.
- **Flexion-distraction fractures:** they happen when the vertebrae are pulled apart, causing a widening of the spinal canal. This type of fracture can be quite severe, often resulting from high-impact accidents or sudden forceful movements. As with compression fractures, flexion-distraction fractures can lead to pain, reduced neck mobility, and spinal instability.
- **Burst fractures:** they are characterized by a fragmented and displaced vertebra, causing a loss of height. The vertebrae are typically broken into multiple pieces, which can exert pressure on the spinal cord and nerve roots. This type of fracture is particularly concerning due to the potential for significant nerve damage, paralysis, and spinal cord injury.

It is important to note that all three types of fractures can have serious implications and require prompt and accurate diagnosis. Early detection and appropriate treatment are crucial to prevent further damage and ensure the best possible outcomes for patients. Moreover, upper cervical levels, particularly the cranio-cervical junction (C1), are particularly vulnerable and may lead to more severe injuries, making accurate detection and assessment vital in these cases [15]. Figure 1.2 illustrates two kinds of fractures in a cervical spine.



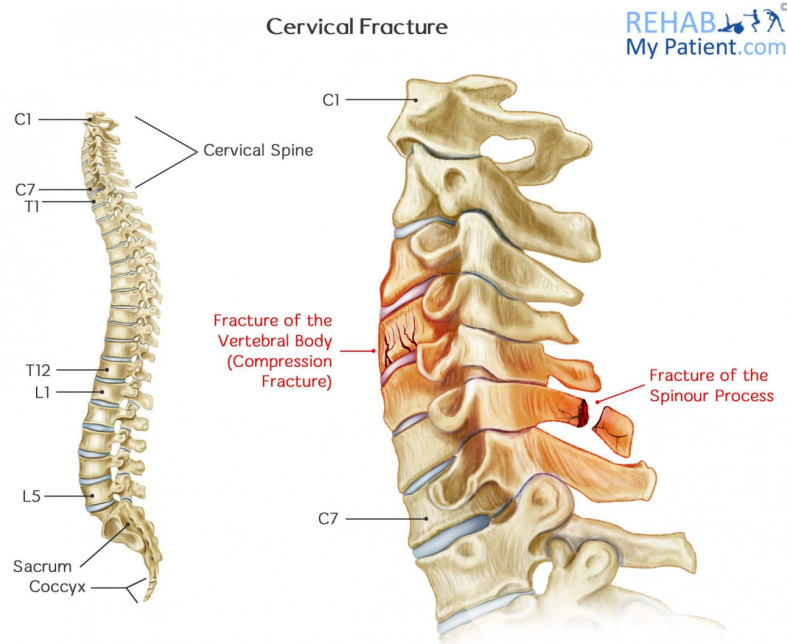


Figure 1.2: Examples of cervical spine fractures [24].

Cervical spine fractures are a subject of extensive research, driven by their diverse causes, such as motor vehicle accidents, falls, and sports-related injuries. The severity of these fractures can vary widely, leading to various symptoms, including pain, instability, and potential nerve damage.

### 1.3 Statistics on cervical spine fractures

As reported by the National Institute of Neurological Disorders and Stroke [20], over 1.5 million spine fractures occur annually in the United States, leading to more than 17,730 spinal cord injuries. Among all spinal fractures, the cervical spine is the most common site of occurrence. Notably, the incidence of spinal fractures has risen in the elderly population, presenting unique challenges in detection due to the presence of superimposed degenerative disease and osteoporosis. To diagnose adult spine fractures, computed tomography (CT) has become the primary imaging method, replacing traditional radiographs (x-rays). Swift detection and precise localization of vertebral fractures are crucial in trauma cases to prevent neurologic deterioration and paralysis. According to the American Academy of Orthopaedic Surgeons, approximately 5% of all fractures in the United States are cervical spine fractures. Motor vehicle accidents are the leading cause of these fractures, accounting for around 60% of all cases. Among cervical spine fractures, compression fractures are the most prevalent type, constituting approximately 50% of all occurrences. Understanding the epidemiology and characteristics of cervical spine fractures is essential for effective preventive measures and treatment strategies [20].

## 1.4 Medical imaging techniques for the diagnosis of cervical spine fractures

Cervical spine fractures are a common injury that can occur due to various factors, such as trauma, osteoporosis, and degenerative conditions. To diagnose cervical spine fractures, medical imaging techniques play a crucial role, including X-rays, CT scans, and MRIs. X-rays are a form of electromagnetic radiation that can pass through soft tissues, enabling visualization of the bones. They are frequently employed as a primary imaging method for cervical spine fractures due to their widespread availability and cost-effectiveness. However, x-rays do have certain limitations, as they may not offer sufficient detail to accurately diagnose specific types of fractures. In such cases, more advanced imaging modalities like CT scans and MRIs may be required to provide a comprehensive assessment of the cervical spine and its fractures.



Figure 1.3: A fracture in the cervical spine detected with x-ray [3].

Figure 1.3 demonstrates the detection of cervical spine fractures using Magnetic Resonance Imaging (MRI). However, in addition to MRI, Computed Tomography (CT) scans are another widely used imaging tool for diagnosing cervical spine fractures. CT scans utilize X-rays and advanced computer processing to create detailed cross-sectional images of the cervical spine. This imaging technique offers a higher level of accuracy in diagnosing fractures and allows for a comprehensive assessment of the surrounding soft tissues and organs. Moreover, CT scans are commonly employed in planning surgical interventions for the treatment of cervical spine fractures.

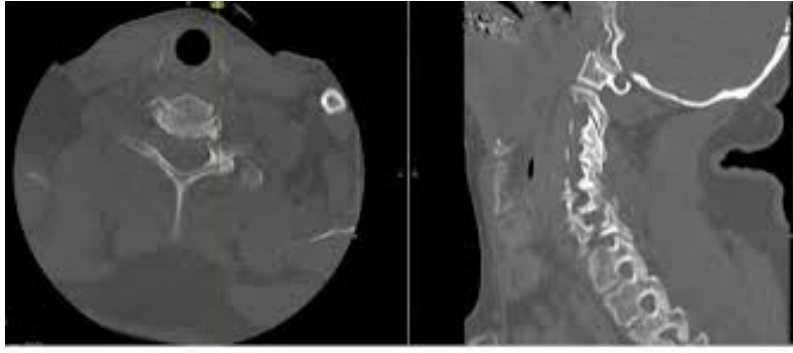


Figure 1.4: The fracture in cervical spine detected with CT scan [5].

Figure 1.4 depicts the detection of cervical spine fractures using Computed Tomography (CT) scans, which utilize X-rays and advanced computer processing to create detailed cross-sectional images. In addition to CT scans, Magnetic Resonance Imaging (MRI) is another non-invasive imaging method used to visualize the cervical spine. MRI employs a strong magnetic field and radio waves to produce high-resolution images, offering excellent soft tissue contrast for accurate evaluation of ligaments, nerves, and surrounding tissues in the cervical spine. Although MRI provides valuable diagnostic information, it is generally more expensive and less readily available compared to X-rays or CT scans.



Figure 1.5: A fracture in cervical spine detected with MRI [31].

In conclusion, the world of medical imaging plays a pivotal role in the detection and management of cervical spine fractures. Within this realm, X-rays, CT scans, and MRIs serve as indispensable tools, each with its own set of advantages and limitations. The choice of which imaging technique to employ hinges on the unique circumstances of each case and the specific objectives of the diagnostic study. Figure 1.5 visually emphasizes the importance of MRI, showcasing its remarkable ability to detect fractures.

## 1.5 Treatment of cervical spine fractures

Fractures are managed differently depending on where they are located and what symptoms you are experiencing. This may include surgery or a neck collar/brace,

but some people do not require either of these. You should take painkillers to control your pain and enable you to move around and resume your normal activities [31].

In more severe cases, surgery may be necessary to stabilize the spine and reduce the risk of further injury. Surgery may involve the use of plates, screws, or rods to hold the vertebrae in place, or the removal of bone fragments or herniated discs. Additionally, medications may be prescribed to reduce pain and inflammation [31].

## **1.6 Early detection and accurate diagnosis of cervical spine fractures**

Early detection and accurate diagnosis of cervical spine fractures are of paramount importance for several reasons: Improved patient outcomes: Early detection and accurate diagnosis enable prompt treatment, leading to better outcomes for patients. This can result in reduced pain and disability, improved quality of life, and lower healthcare costs. Prevention of further injury: Delayed or incorrect diagnosis of cervical spine fractures can lead to additional injuries, such as spinal cord damage, paralysis, or even death. Accurate and timely diagnosis can prevent these serious complications. Effective treatment: Accurate diagnosis allows for appropriate treatment planning, including the selection of the most effective treatment approach. This can contribute to faster recovery and better long-term outcomes. Reduced healthcare costs: Delayed or incorrect diagnosis may lead to prolonged hospital stays, repeat imaging studies, and unnecessary treatments, resulting in higher healthcare costs. Early and accurate diagnosis can reduce these expenses by avoiding unnecessary interventions.

In conclusion, early detection and accurate diagnosis of cervical spine fractures are crucial for ensuring the best possible outcomes for patients and reducing the burden on the healthcare system [4].

## **1.7 Current approaches to cervical spine fracture detection**

Cervical spine fracture detection involves a combination of physical examination, imaging studies, and laboratory tests to ensure accurate and early diagnosis. During the physical examination, healthcare professionals carefully assess patients for specific signs and symptoms indicative of a cervical spine fracture, including pain, tenderness, and restricted range of motion. These clinical findings provide essential initial clues for further investigation. Imaging studies, such as X-rays, CT scans, and MRI scans, play a crucial role in confirming the diagnosis and evaluating the extent of the fracture [4]. X-rays offer a quick and accessible method to visualize the bones, while CT scans provide detailed cross-sectional images. MRI scans provide excellent soft tissue contrast and are particularly useful for assessing the surrounding ligaments, nerves, and other tissues. Together, these imaging techniques enable a comprehensive understanding of the fracture and its impact. Laboratory tests, including blood and urine tests, serve to identify any underlying conditions that might have contributed to the fracture. Additionally, genetic tests can be utilized to detect any hereditary factors linked to the fracture. Understanding these underlying factors can inform the overall management plan for the patient. Early and accurate diagnosis is paramount as it facilitates timely treatment, leading to improved patient

outcomes. This includes reduced pain and disability, enhanced quality of life, and lower healthcare costs. Furthermore, precise and timely diagnosis is instrumental in preventing further injury, such as spinal cord damage, paralysis, or even fatal consequences. Accurate diagnosis also enables healthcare providers to devise effective treatment plans, resulting in faster recovery and better long-term outcomes.

Ultimately, early and accurate diagnosis contributes to cost savings within the healthcare system by avoiding unnecessary interventions. By employing a multifaceted approach to cervical spine fracture detection, healthcare professionals can ensure the best possible care for their patients.

## 1.8 Artificial intelligence applications in e-medicine

Artificial Intelligence (AI) is a rapidly advancing technology with the potential to transform the medical industry. By leveraging data and experiences, AI enables machines to learn, make decisions, and take actions without direct human intervention [2]. The applications of AI in healthcare are vast and hold promise to enhance medical diagnosis, treatment, and overall patient care. One of the key advantages of AI in medicine lies in its ability to enhance the accuracy and efficiency of medical processes. By analyzing extensive datasets, AI can identify intricate patterns and trends that may not be readily apparent to human clinicians. This empowers medical professionals to make more informed decisions and predictions about potential outcomes, ultimately leading to improved patient care. AI can also streamline administrative tasks, such as medical record-keeping and billing, by automating routine processes. By reducing the burden of administrative work, healthcare providers can allocate more time and attention to patient care, resulting in better overall outcomes [2]. Personalized medicine is another area where AI excels. By processing individual patient data, AI algorithms can tailor treatment plans to meet the unique needs of each person. This personalized approach holds the potential to optimize treatment outcomes and enhance patient satisfaction. Moreover, AI can play a pivotal role in early disease detection and diagnosis. By analyzing vast amounts of data, AI systems can identify subtle indications of diseases at their nascent stages, enabling timely interventions and improved prognoses. The potential impact of AI extends to cost reduction as well. By streamlining processes, enhancing efficiency, and reducing medical errors, AI can help lower healthcare costs, making medical services more accessible and affordable for patients. However, while AI holds immense promise, its integration into the medical industry requires careful consideration and ethical considerations. Ensuring the privacy and security of patient data, validating AI algorithms, and maintaining human oversight are critical aspects of AI adoption in healthcare.

In conclusion, AI has the potential to revolutionize the medical industry, revolutionizing patient care, optimizing treatment outcomes, and reducing costs. By leveraging the power of AI, the healthcare sector can usher in a new era of precision medicine, improved patient experiences, and overall advancement in medical practices [4].

## **1.9 Performance of AI-based cervical spine fracture detection systems**

The assessment of AI-based cervical spine fracture detection systems' performance is a crucial step in determining their accuracy and reliability. Performance evaluation involves gauging the system's ability to detect fractures accurately while minimizing false positives and false negatives. Various methods, such as accuracy metrics, receiver operating characteristic (ROC) curves, and sensitivity and specificity tests, can be employed to evaluate performance. Real-world data, including patient records and imaging studies, can be used for this evaluation. Ensuring the accuracy and reliability of AI-based cervical spine fracture detection systems through thorough performance evaluation is vital in delivering the best possible outcomes for patients [39].

## **1.10 Image analysis and computer vision**

Image analysis and computer vision are closely related fields that involve using computers to analyze and interpret digital images. Image analysis focuses on extracting meaningful information from digital images, while computer vision uses algorithms to interpret and understand the content of these images. Both disciplines find applications in various areas, including medical imaging, robotics, and autonomous vehicles. In the field of e-medicine, image analysis plays a vital role in advancing medical research and improving healthcare. This technology enables the detection of hidden diseases, leading to early diagnosis and saving numerous lives. For instance, in the case of cervical spine fractures, approximately 3,000 CT studies were analyzed by spine radiology specialists from renowned organizations. They expertly annotated the images to identify the presence, vertebral level, and location of any cervical spine fractures. This extensive imaging data was collected from twelve sites across six continents, enhancing the understanding and diagnosis of such fractures. The combination of image analysis and computer vision holds immense potential for revolutionizing the medical field and contributing to better patient care and outcomes [2].

## **1.11 Automatic detection: principle and methods**

Various AI methods have been employed for cervical spine fracture detection, offering promising avenues for improving medical diagnosis. These methods encompass machine learning algorithms, deep learning algorithms, natural language processing (NLP), decision support systems, and more. Deep learning algorithms, particularly Convolutional Neural Networks (CNNs), have demonstrated their effectiveness in analyzing medical images to detect cervical spine fractures. By training on extensive medical image datasets, CNNs learn to identify patterns indicative of fractures with remarkable accuracy. They not only detect fractures but also provide additional insights, such as fracture type, location, and severity, enriching the diagnostic process. Furthermore, NLP technology can be leveraged to analyze radiology reports, ensuring that any missed or inadequately documented fractures are duly recognized. Decision support systems prove valuable in cervical spine fracture detection by integrating data from multiple sources, such as medical images and patient history. AI



algorithms analyze this data and provide valuable recommendations to aid radiologists in making more accurate diagnoses and treatment decisions. It is essential to emphasize that while AI methods hold immense promise, they should complement and support trained medical professionals rather than replace them. The combination of AI techniques with clinical expertise ensures optimal patient care and improved outcomes in cervical spine fracture diagnosis and treatment.

## 1.12 Potential benefits and limitations of AI in cervical spine fracture detection

The integration of AI in cervical spine fracture detection offers numerous potential benefits. Firstly, it can significantly improve the accuracy and efficiency of diagnosis, leading to more precise and timely identification of fractures. This, in turn, can result in reduced costs and improved patient outcomes, as early detection allows for prompt treatment and better management. AI-based systems excel in analyzing extensive datasets, enabling them to identify subtle patterns and trends that might otherwise go unnoticed. Moreover, they can predict potential future events, aiding in proactive and preventive medical approaches. Automating routine tasks, such as medical record-keeping and billing, streamlines administrative processes, freeing up healthcare professionals' time for more critical tasks. AI also contributes to the development of personalized treatments, tailoring medical interventions to individual patients' specific needs. By facilitating early and accurate disease detection, AI can improve patient prognosis and overall health outcomes. However, it is essential to acknowledge potential limitations. AI systems can be susceptible to bias if the data used to train them are not diverse or representative. This may lead to inaccuracies in the system's predictions, especially for certain patient populations. Moreover, AI may face challenges in detecting subtle fractures or fractures in atypical locations, where human expertise remains crucial. Additionally, certain medical conditions, such as osteoporosis, might present challenges for AI-based systems in accurately detecting fractures. As with any technology, it is vital to exercise caution and use AI as a complementary tool, working alongside medical professionals to ensure the best possible patient care and outcomes [39].

## 1.13 Conclusion

This chapter has served as an introductory support to initiate the reader to the problem addressed in the context of our work. Specifically, it has provided an overview of the anatomy of the cervical spine, the types of fractures that it can be subject to, and the techniques used to diagnose and treat these fractures. Moreover, the chapter has exhibited the potential risks associated with cervical spine fractures and the importance of their early detection and treatment, by paying particular attention to the emergence of computer vision and artificial intelligence-based methods and their capacity to improve the diagnosis quality.

Prior to presenting personal contributions, a comprehensive review of the essential works presented in the literature in the context of our study's topic is necessary to understand the main limits of the existing related works. Hence, a synthesis study of the recent state-of-the-art cervical spine fracture detection methods will be the object of the next chapter of this manuscript.

## Chapter 2

### State-of-the-art on cervical spine fracture detection



## 2.1 Introduction

In this chapter, we delve into the field of cervical spine fracture detection and provide a comprehensive overview of the main approaches existing in the literature. The primary focus is to explore state-of-the-art methods that aim to enhance the accuracy of cervical spine detection while reducing the time required for diagnosis. By reviewing and analyzing these works, we aim to gain insights into the advancements made in this domain.

In Section 2.2, we present an in depth review of the current state-of-the-art methods utilized for cervical spine fracture detection. These methods have been developed to leverage the power of Artificial Intelligence (AI) techniques, particularly deep learning approaches, to improve the accuracy and efficiency of detection.

Moving forward, in Section 2.3, we conduct a comparative study of the analyzed works based on carefully chosen criteria. This study allows us to evaluate the strengths and weaknesses of different approaches and identify the key factors that contribute to their performance.

Section 2.4 is dedicated to the summary and discussion of the findings and results derived from our comprehensive review. We highlight significant insights gained from the analyzed works and discuss their implications in the context of cervical spine fracture detection.

Lastly, in Section 2.5, we present potential perspectives and ideas for future research in this area. These suggestions encompass exploring new techniques, such as novel deep learning architectures, incorporating additional AI paradigms like transfer learning or reinforcement learning, and investigating the combination of multiple methods to improve accuracy and detection speed.

Through this thorough examination of the literature, we aim to contribute to the advancement of cervical spine fracture detection and inspire further research in this critical field at the intersection of AI and medical imaging.

## 2.2 Description of related work

Cervical spine fracture detection has been a hot research topic during the last decade. Therefore, many research papers have been published in the literature as attempts to present effective solutions to the problem. With the aim to discern the research progress registered recently in the context of the topic, we present and succinctly discuss in this section some of the recent and relevant state-of-the-art methods presented in the setting.

### **CT cervical spine fracture detection using a convolutional neural network**

In their study, Small et al. [30] have evaluated the performance of a Convolutional Neural Network (CNN) developed by Aidoc, known as FDA-approved CNN, for the detection of cervical spine fractures on computed tomography scans. The researchers utilize two datasets, one consisting of retrospective blinded data from 47 clinical sites with approximately 8000 examinations, and the other comprising cervical spine CT studies with short interval follow-up MR imaging.

The evaluation focuses on estimating the positive predictive values (PPVs) and negative predictive values (NPVs) for both radiologists and the CNN in detecting cervical fractures, particularly in a population with a lower incidence of such fractures. The results indicate that the radiologists achieve an estimated PPV of

32% and an NPV of 99.9%, while the CNN achieves an estimated PPV of 30% and an NPV of 99.5%. It is noteworthy that the CNN demonstrates a sensitivity of 79%, slightly lower than that of the radiologists. Additionally, CNN identifies fractures missed by radiologists in seven examinations, including four cases of chronic fractures. This suggests the potential of the CNN to enhance diagnostic accuracy in detecting cervical fractures, particularly in high-volume practices where efficient worklist prioritization is critical.

However, the study also acknowledges certain limitations of CNN. It may struggle to detect areas of gross bony translation and fractures characterized by distraction rather than linear bony features. Despite these limitations, the findings emphasize the potential of the CNN to augment cervical fracture detection and encourage further exploration in this field.

Overall, this study provides valuable insights into the performance and limitations of a CNN-based approach for CT cervical spine fracture detection, offering significant implications for improving diagnostic accuracy in clinical settings [30].

#### **Detecting intertrochanteric hip fractures with orthopedist level accuracy using a deep convolutional neural network**

In their research, Urakawa et al. [35] conducted a comparative analysis between artificial intelligence and orthopedic surgeons in detecting intertrochanteric hip fractures from anterior-view proximal femoral radiographs. The chosen AI system for this study was the Visual Geometry Group 16-layer (VGG-16) network [35]. The results revealed a comparable accuracy to that of diagnoses made by radiologists.

For the experiments, the researchers utilized a dataset comprising anterior-view hip radiographs from 1773 patients who had undergone treatment by an orthopedic surgeon. These images were subsequently cropped to a matrix size of  $300 \times 300$  pixels for further analysis, resulting in 3346 hip images (1773 with fractures and 1573 without fractures). The dataset was then divided into training, validation, and test sets. To expedite training time and reduce the number of images required for effective classification, the authors employed the TensorFlow deep learning framework and the pre-trained VGG-16 model for transfer learning. Three regularization techniques, namely data augmentation, L2 regularization, and early stopping, were employed in the CNN architecture to mitigate overfitting. The training process involved 2650 iterations, or 132,500 augmented images, with an initial learning rate of 0.0001, decay steps of 265, and a decay rate of 0.8. Adam optimization and exponential learning rate scheduling were utilized. The final network parameters were restored, and each image in the test set was classified as fractured or non-fractured.

This study highlights the successful application of a deep convolutional neural network, specifically the VGG-16 model, in detecting intertrochanteric hip fractures with accuracy on par with radiologists. The use of AI in this context has significant potential to enhance diagnostic capabilities and expedite fracture identification, providing valuable support to orthopedic surgeons.

#### **Deep learning model for detecting cervical spinal cord compression in MRI scans**

Merali et al. [18] conducted a study with the objective of developing a deep-learning model capable of detecting cervical spinal cord compression in patients diagnosed with Degenerative Cervical Myelopathy (DCM). Their proposed model was tested on T2-weighted MRI scans obtained from patients with diverse demographics and disease characteristics, as well as images acquired from various MRI scanners. The researchers employed a range of analytical techniques to gain insights into the model's functioning and performance.

For the study, a retrospective analysis was conducted on prospectively collected MRI studies from patients enrolled in two clinical studies focusing on DCM. Patients meeting the eligibility criteria and excluding specific conditions were included. MRI images of both compressed and non-compressed spinal cords were obtained from this patient cohort for training the model. Two distinct patient cohorts were established for model development and validation, with 75% of patients assigned to the training/validation dataset and the remaining 25% allocated to the holdout dataset. Baseline clinical data, mJOA score, and MRI image parameters were compared between the two datasets using t-tests and X2 tests.

In the model training process, the researchers employed the ResNet-50 convolutional neural network architecture, which consisted of fully connected layers and dropout layers. Multiple network configurations were tested to determine the optimal setup for the dataset. This included variations such as a single fully connected layer with different neuron counts (256, 512, 1024, and 2048) and two fully connected layers with different neuron counts (256, 512, and 1024), each accompanied by two dropout layers with a dropout rate of 30% each. The training and validation datasets were shuffled and split, utilizing the Adam optimizer with a learning rate of 0.0001 and a batch size of 16. To address imbalanced classes, the weighted binary cross-entropy loss was utilized. Each model configuration was evaluated based on binary cross-entropy loss and accuracy on the validation set, and the best-performing model was selected for further testing.

To gain insights into the features influencing correct and incorrect classifications, class activation maps (CAMs) were generated for correctly classified images (true positives) and incorrectly classified images (false negatives). The Keras-vis package in Python was employed to generate these CAMs. The results of the study demonstrated that the deep learning-based model achieved the accurate classification of spine MRIs as either compressed or non-compressed. While the model exhibited high sensitivity, its specificity was relatively lower. However, the model’s ability to identify clinically relevant features associated with spinal compression in MR images provided valuable insights into its decision-making process and could potentially contribute to enhancing its accuracy [18].

### **Faster R-CNN-Based detection of cervical spinal cord injury and disc degeneration**

Shaolong et al. [29] conducted a comprehensive investigation into the utilization of deep learning techniques applied to Magnetic Resonance Imaging (MRI) for the classification and detection of lesions associated with cervical spinal cord diseases. The researchers conducted a retrospective review of MRI scans from a dataset comprising 1,500 patients, spanning the period from January 2013 to December 2018. The MRI data were randomly divided into three distinct groups: disc group (800 datasets), injured group (200 datasets), and normal group (500 datasets).

To facilitate lesion detection during MRI scans, the researchers implemented a deep neural network specifically designed for MRI analysis. They employed the Faster R-CNN (Region Convolutional Neural Network) framework, which combines a backbone convolutional feature extractor utilizing both the ResNet-50 and VGG-16 networks. This integration of Faster R-CNN with ResNet-50 and VGG-16 networks yielded promising results in terms of prediction accuracy and speed for lesion detection and recognition within cervical spinal cord MRIs.

The research findings indicated that the proposed architecture of Faster R-CNN, in conjunction with ResNet-50 and VGG-16 networks, demonstrated commendable recognition capabilities for identifying lesions in cervical spinal cord MRIs. The

mean average precisions (mAPs) achieved by Faster R-CNN with ResNet-50 and VGG-16 were reported as 88.6% and 72.3%, respectively. These outcomes provide compelling evidence that deep learning techniques can effectively contribute to the identification and detection of lesions in cervical MRIs. Consequently, such advancements in lesion recognition can greatly assist radiologists and spine surgeons in making accurate diagnoses. The experimental results obtained through this approach demonstrated promising recognition performance [29].

### **Artificial intelligence-based fracture recognition on computed tomography: a comprehensive literature review and recommendations**

Dankelman et al. [6] conducted a meticulous review of the literature regarding the application of Convolutional Neural Networks (CNNs) for the accurate detection and classification of fractures on computed tomography (CT) scans. The study highlighted the immense potential of CNNs in this domain but underscored the necessity for further research to assess their practicality and effectiveness in clinical settings.

In their analysis, the researchers collected crucial data points from each study, including the author, publication year, anatomical location of the fracture, type of AI models employed, CT slice imaging direction, output classes, ground truth label assignment, number of patients, and performance metrics. The performance of AI models was evaluated using metrics such as accuracy, F1-score, and area under the curve (AUC). Among the 1140 studies initially identified, a comprehensive assessment was performed on 17 relevant studies.

The reported accuracy of AI models ranged from 69% to 99%, indicating varying levels of precision across different studies. The F1-score, which considers both precision and recall, exhibited a range of 0.35 to 0.94. Additionally, the AUC, which measures the model's ability to discriminate between fracture and non-fracture cases, ranged from 77% to 95%. Notably, based on the analysis of ten studies, CNNs demonstrated either comparable or superior diagnostic accuracy when compared to clinical evaluation alone.

These findings affirm the practical applicability of CNNs for the detection and classification of fractures on CT scans. However, to establish their clinical utility, further investigation is warranted. Future research endeavours should focus on validating the performance of CNNs on larger and more diverse datasets, encompassing different geographical locations, to ensure the generalizability and robustness of these algorithms in real-world clinical practice.

Overall, the study by Dankelman et al. sheds light on the potential of CNNs in fracture recognition on CT scans, providing valuable insights for clinicians and researchers. It emphasizes the need for continued efforts to explore the additional value of CNNs in daily clinical workflows and highlights the importance of thorough evaluation and validation of these AI technologies for optimal integration into healthcare settings [6].

### **Cervical spine fracture detection via Computed Tomography scan**

Tuan et al. [34] conducted an extensive investigation to develop an efficient and accurate method for the early detection and localization of spine fractures. Their research focused on employing deep-learning models specifically designed for cervical spine fracture detection. Through their experimentation, they explored multiple machine learning models and identified a two-stage approach utilizing Deep Convolutional Neural Networks with RNN and Attention layers as the highest-performing model.

To achieve their objectives, the researchers explored various classification ap-

proaches, including 3D and 2D methods, employing EfficientNet and ConvNeXt models [34]. Additionally, they developed a CNN model capable of detecting vertebrae bounding boxes and classifying whether a cervical spine was fractured. The final model achieved favorable results with reduced inference time, even with limited training resources, positioning it among the top 25 models in the contest.

The dataset utilized in their study was obtained from the RSNA 2022 Cervical Spine Fracture Detection competition hosted on Kaggle. This dataset consisted of three main folders: train images, test images, and segmentations. The researchers meticulously leveraged this dataset to train and evaluate their models, ensuring comprehensive coverage of spine fracture scenarios.

In their pursuit of optimal fracture detection, the researchers conducted experiments employing diverse approaches, such as 3D CNN, 2D CNN, and 2.5D CNN+RNN. Notably, the 2.5D CNN+RNN model demonstrated superior efficiency in terms of time and resource utilization, while still achieving commendable results [34].

Furthermore, the researchers provided insightful suggestions for future work. They recommended exploring Transformer layers as an alternative to LSTM for sequence data processing, potentially improving the model’s ability to capture long-range dependencies. Additionally, they emphasized the potential benefits of training another backbone model, opening avenues for further enhancements and performance optimization.

In summary, the study by Tuan et al. [34] presents a thorough investigation and a comparative study into the application of deep learning for spine fracture detection and localization. Their research contributes to the advancement of medical image analysis, showcasing the efficacy of deep convolutional networks with RNN and Attention layers. The results underscore the potential of these models in improving fracture detection efficiency and accuracy. The study also sets the stage for future research to explore alternative architectures and backbone models, promoting continuous innovation in this critical domain of medical imaging [34].

### **Deep sequential learning for cervical spine fracture detection in computed tomography imaging**

The research conducted by Salehinejad et al. [26] introduces a deep-learning model for the automated detection of cervical spine fractures in CT axial images. Their study emphasizes the importance of accurate fracture diagnosis in patient management and addresses fracture detection as a classification problem. The proposed model employs a deep convolutional neural network (DCNN) with a bidirectional long short-term memory (BLSTM) layer as the baseline architecture, specifically tailored for axial cervical spine CT images [26].

The authors provide insights into the preprocessing steps involved in preparing the input images, including cropping and windowing techniques. Additionally, they outline the learning phase, which encompasses feature extraction from preprocessed images and the utilization of a bidirectional network of LSTM units to capture temporal dependencies among axial images. These extracted features are then mapped to the corresponding target labels. The performance of the proposed model is evaluated on a dataset comprising 3,666 CT scans, yielding classification accuracy rates of 70.92% and 79.18% on balanced and imbalanced test datasets, respectively [26]. The results underscore the potential of deep learning models for automated fracture detection and warrant further exploration in this domain.

In summary, Salehinejad et al. have presented a promising deep sequential learning approach for detecting cervical spine fractures in CT imaging. Their model,



incorporating a DCNN with a BLSTM layer, demonstrates notable classification accuracy on diverse test datasets. This study sheds light on the potential of deep learning techniques in fracture detection and provides a foundation for future investigations aimed at refining and advancing automated fracture detection algorithms in clinical settings.

**Detection and classification of mandibular fracture on CT scan using deep convolutional neural network** Xuebing et al. [41] have used a Convolutional Neural Network (CNN) approach for the detection and classification of mandibular fractures on spiral computed tomography (CT) images. This algorithm has shown excellent image processing capabilities. They have aimed to evaluate the accuracy and reliability of the CNN approach for detecting and classifying mandibular fractures. Their study was conducted and approved by the Ethics Committee of Peking University School and Hospital of Stomatology.

The data used is the data of all patients who underwent CT scans using a 16-slice CT scanner with 1.25-mm slice thickness at Peking University School and Hospital of Stomatology between January 2013 and July 2020 were extracted according to specific criteria.

In conclusion, they found that the CNN approach showed comparable reliability and accuracy in detecting and classifying mandibular fractures and can be useful for automated diagnosis and classification of mandibular fractures [41].

## 2.3 Comparaison of reviewed methods

In the context of our research, we have conducted a thorough review of existing works in the field of cervical spine fracture detection. As part of this comparative analysis, we have considered three widely used evaluation metrics, namely: *Accuracy*, *Sensitivity*, and *Specificity*. Specifically, *Accuracy* metric measures the overall correctness of a model’s predictions. It represents the proportion of correct predictions out of the total number of predictions made by the model. *Sensitivity* metric, also known as *True Positive Rate* or *Recall*, quantifies the ability of a method to correctly identify positive cases (mandibular fractures). *Specificity* metric measures the method’s ability to correctly identify negative cases (non-fracture cases). These metrics play a crucial role in assessing the performance of various methods employed in this domain.

Mathematically, the formulas for these evaluation metrics are expressed using the  $TP$ ,  $TN$ ,  $FP$  and  $FN$  parameters. Explicitly,  $TP$  refers to the number of true positive cases,  $TN$  refers to the number of true negative cases,  $FP$  refers to the number of false positive cases, and  $FN$  refers to the number of false negative cases. The formulas of the evaluation metrics are given as follows:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN). \quad (2.1)$$

$$Sensitivity = TP / (TP + FN). \quad (2.2)$$

$$Specificity = TN / (TN + FP). \quad (2.3)$$

By considering these evaluation metrics and their corresponding formulas, we can comprehensively analyze and compare the performance of different techniques employed in cervical spine fracture detection. This systematic evaluation enables us to gain insights into the strengths and limitations of various methods, facilitating the advancement of this field and the development of more accurate and reliable approaches in the future.

Works	Architectures	Dimensions	Obtained results	Speeds	Datasets	Image technique
Urakawa et al. [35] (2018)	CNN (VGG-16)	2D	Accuracy:95.5% Sensitivity:93.9% Specificity:97.4%	/	private /	X-Ray
Shaolong et al. [29] (2020)	Faster RCNN (ResNet-50 and VGG-16)	2D	Accuracy:ResNet-50: 88,6%, VGG-16: 72.3% Sensitivity:/ Specificity:/	0.22 to 0,24 s/image	private /	MRI
Xuebing et al. [41] (2021)	CNN	2D	Accuracy:87.8% Sensitivity:/ Specificity:/	/	16-slice CT scanner (Peking University School and Hospital of Stomatology between January 2013 and July 2020) /	CT Scan
Salehinejad et al. [26] (2021)	DCNN (ResNet-50 + BLSTM)	2D	Accuracy: Balanced test:70.92% Imbalanced test:79.18% Sensitivity: Balanced test:80.06% Imbalanced test:77.62% Specificity: Balanced test:06.47% Imbalanced test:13.78%	/	CT scans of 3,666 cases (729 positive and 2,937 negative cases) /	CT Scan
Small et al. [30] (2021)	CNN (FDA-approved CNN)	3D	Accuracy:92% Sensitivity:76% Specificity:97%	3 to 8 min	private /	MRI
Merali et al. [18] (2021)	CNN (ResNet-50)	2D	Accuracy:92.41% Sensitivity:/ Specificity:/	/	private /	MRI
Tuan et al. [34] (2022)	CNN+RNN (Efficient-Net+ConvNeXt)	3D(2,5D)+2D	Accuracy:87.8% Sensitivity:/ Specificity:/	/	Kaggle (ASNR and ASSR) /	CT Scan
Dankelman et al. [6] (2022)	CNN	2D	Accuracy:88% Sensitivity:92.9% Specificity:/	/	private /	CT Scan

Table 2.2: Comparative table of the studied works according to criteria.

## 2.4 Summary of comparison and discussion

Based on a review of existing methods in the literature on cervical spine fracture detection, comparing the performance of different methods poses challenges due to variations in evaluation data. However, a majority of the methods employ Convolutional Neural Networks (CNNs) with diverse architectures, showcasing the significance of CNNs in this domain. Notably, there is a growing interest in both 2D and 3D detection approaches. Regarding 2D detection, the most impressive results include a best accuracy of 95% a best sensitivity of 93% and a notable best specificity of 97.4%. These outcomes were achieved utilizing the VGG-16 architecture, which has demonstrated its effectiveness in medical imaging tasks. It should be noted that the evaluation of 2D images disregarded computational speed.

Shifting the focus to 3D detection, the best accuracy achieved was 92%, accompanied by a sensitivity of 76% and a specificity of 97%. Analyzing volumetric data presents additional complexities in this context. However, the utilization of an FDA-approved CNN holds promise in advancing 3D detection techniques.

While the aforementioned results provide valuable insights, it is essential to con-

sider computational efficiency for practical implementation. Notably, the 3D detection methods required a time range of 3 to 8 minutes per scan, while the 2D methods showcased an impressive speed of approximately 0.22 to 0.24 seconds per image, highlighting their efficiency in processing individual frames.

In conclusion, this state-of-the-art summary sheds light on cervical spine fracture detection. The utilization of CNNs with diverse architectures has shown remarkable accuracy, sensitivity, and specificity in both 2D and 3D detection. The VGG-16 architecture, specifically for 2D detection, and an FDA-approved CNN for 3D detection, demonstrate their potential in this domain. Future research should focus on enhancing computational efficiency to facilitate the real-time implementation of these methods in clinical settings.

## 2.5 Future work perspectives for cervical spine fracture detection

As the field of cervical spine fracture detection continues to evolve, there are several promising avenues for future research. This section outlines potential ideas and perspectives that can drive further advancements in this domain. The focus is on investigating and improving the application of the VGG-16 method, exploring other CNN architectures, combining different methods, and developing new deep-learning models for more accurate and efficient cervical spine fracture detection.

- *In depth investigation of VGG-16:* One key area of future work involves conducting a comprehensive study to further explore and enhance the application of the VGG-16 architecture. Despite its promising results, there is room for improvement, particularly in terms of speed and accuracy of detection. Researchers can delve into techniques such as model compression, network pruning, or architecture modifications to optimize the performance of VGG-16 specifically for cervical spine fracture detection.
- *Exploration of other CNN architectures:* In addition to VGG-16, there are various CNN architectures proposed in the literature for medical image analysis. It is essential to explore these architectures beyond VGG-16 and assess their suitability for cervical spine fracture detection. Architectures like Next, Inception, or DenseNet can be investigated, and their performance can be compared with VGG-16 to determine the most effective architecture for this task.
- *Combination of methods:* An intriguing avenue for future research is the combination of different methods studied in the literature for cervical spine fracture detection. By integrating multiple approaches, researchers can explore the synergistic effects and potential improvements in accuracy and speed of detection. Techniques such as feature fusion or ensemble learning can be employed to identify suitable combinations that yield superior performance.
- *Development of new deep learning models:* Another fruitful direction involves the development of novel deep-learning models tailored specifically for cervical spine fracture detection. Researchers can explore the incorporation of techniques like Recurrent Neural Networks (RNNs) or autoencoders to capture temporal dependencies or learn more representative feature representations. Moreover, leveraging transfer learning or reinforcement learning paradigms can further enhance the performance of the detection models.



By pursuing these future research directions, the field of cervical spine fracture detection can continue to advance, leading to improved accuracy and efficiency in detecting and diagnosing cervical spine fractures. These endeavours will contribute to enhancing clinical practice and ultimately benefitting patient outcomes.

## 2.6 Conclusion

In conclusion, this chapter has provided an extensive overview and synthesis study of existing works in the literature concerning cervical spine fracture detection. Specifically, our aim goal throughout the chapter was to review the main recent state-of-the-art methods that enhance accuracy and reduce diagnosis time. Moreover, we have performed a comparative study of the analyzed works, evaluating them based on selected criteria such as accuracy, speed, scalability, robustness, and the ability to handle diverse fracture types. This analysis allowed us, on one side, to identify the strengths and limitations of each method and, on the other side, to highlight eventual avenues for future research to further advance the field of cervical spine fracture detection and enhance patient care.

In the next chapter, we will provide the necessary details and a description of our main propositions by which we aim to further contribute to the progress in cervical spine fracture detection and ultimately improve patient outcomes.

# Chapter 3

## A new deep learning model and cloud-based architecture for cervical spine fracture detection

### 3.1 Introduction

The advent of machine learning and deep learning technologies has radically transformed the domain of medical imaging, opening up new pathways for innovation and improvement. Among the various applications, the detection of cervical spine fractures stands out as a critical issue, given the significant implications for patient health and treatment.

After performing a review of recent state-of-the-art works established in the setting of fracture detection, we present, in this chapter, a new cloud-based architecture and we propose herein a model that produces a rapid and efficient detection of cervical spine fractures. To do so, we will first give a general overview of the tools that are necessary for the implementation of the proposed framework, and then we will describe in detail the different aspects of this latter. Specifically, we will first explore, in Section 3.2, the architecture and functionalities of the vision transformer (ViT). Then, we will delve, in Section 3.3, into the Faster R-CNN object detection model. In Section 3.4, we will examine the role of attention maps. Subsequently, we will unravel, in Section 3.5, the detailed theoretical and practical aspects of the new proposed cloud-based system architecture, specifically, focusing on the data pipeline and cloud architecture for medical data analysis. Finally, in Section 3.6, we will summarize the major points, and contemplate the implications of our work.

### 3.2 General overview of ViT

Venturing into the depths of the vision transformer (ViT) model’s architecture, we embark on a journey to understand its inner workings. This exploration delves into the intricacies of self-attention mechanisms and sequential processing, which are essential components that empower the model’s prowess.

The vision transformer (ViT) model, originally proposed by Vaswani et al. [1], stands as a state-of-the-art advancement in deep learning architectures specifically tailored for visual recognition tasks. Remarkably, the ViT model builds upon the transformer architecture, which was initially introduced for natural language processing in the groundbreaking paper “Attention is all you need” in 2017 [36]. In

doing so, it adapts the capabilities of Transformer models to effectively process image data.

Succinctly, the general idea behind the ViT model consists in breaking down an image to process into a sequence of fixed-size patches, treating it therefore as a sequence of tokens, similar to how words of a sentence are treated in natural language processing tasks. These patches are then fed into a standard transformer encoder, which enables self-attention mechanisms to capture long-range dependencies between the patches. Thanks to these self-attention mechanisms, which came up for the first time in the field of computer vision with the apparition of ViT, the model attends different patches in the processed image and hence learns complex spatial relationships between them. Consequently, by effectively capturing both local and global context information, ViT excels in various visual recognition tasks, including image classification, object detection, and segmentation.

Another key advantage of the ViT model is its ability to handle large-scale datasets efficiently. It can be pre-trained on large image datasets using self-supervised or semi-supervised learning techniques and then fine-tuned on specific downstream tasks with smaller labelled datasets. This transfer learning paradigm makes ViT an appealing choice for a wide range of computer vision applications, as it can leverage knowledge learned from diverse datasets and generalize well to new tasks.

Overall, the ViT model represents a significant advancement in computer vision and has demonstrated encouraging performance on various benchmark datasets. Making it a prominent candidate for enhancing image-based applications and advancing the field of computer vision research [8].

### 3.2.1 ViT model architecture

Unlike the traditional convolutional neural networks (CNNs) architecture, the ViT model relies on a transformer which has an encoder/decoder architecture. This means that the ViT is actually composed of two main components: an encoder and a decoder. Specifically, the encoder processes the input sequence and creates a representation of it, while the decoder generates the output sequence based on the produced representation. For illustration, we give in Figure 3.1 a general representation of the encoder/decoder architecture of a transformer.

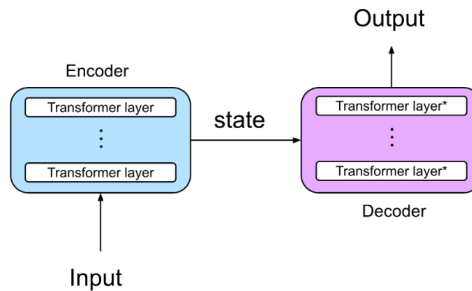


Figure 3.1: A general representation of the architecture of a transformer [9].

Moreover, as can be seen in Figure 3.2, both the encoder and decoder consist of several layers, each containing a self-attention mechanism and feeding-forward neural networks. By using the self-attention mechanism, the model calculates an attention score for each element of the input sequence based on the relationship of the element with all the other elements of the sequence. Explicitly, the attention

score of an element determines how much this element should contribute to the final representation. Thus, the self-attention mechanisms enable the model to focus on relevant and main information of the processed sequence and to ignore its irrelevant or redundant parts. This is particularly beneficial in the case of long sequences, as the model can capture global dependencies and relationships between distant elements.

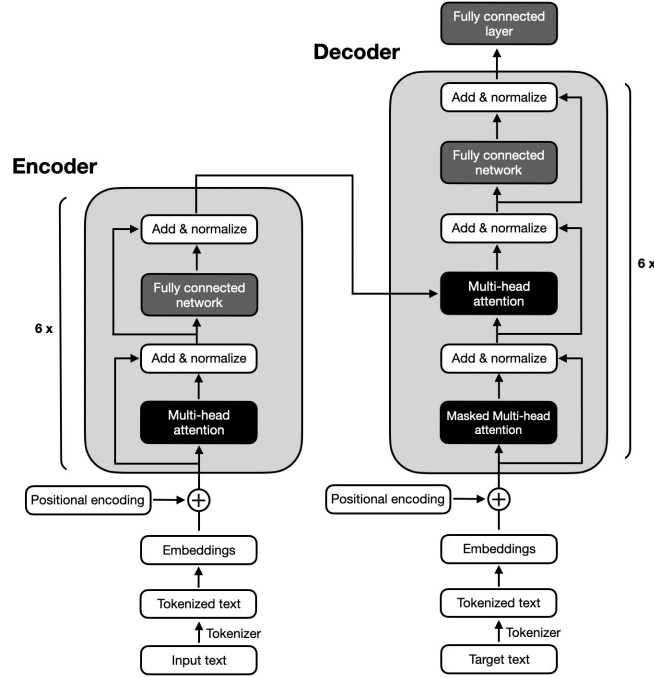


Figure 3.2: Transformer encoder and decoder internal layers [23].

It is worth noting that certain recent transforms use a more sophisticated attention mechanism, namely the multi-head attention, which is an extension of the conventional self-attention mechanism described above. In fact, the multi-head attention mechanism is composed of several attention mechanisms (called also heads) that learn different relationships between elements in a sequence. In other words, in spite of relying on a single attention mechanism to capture all kinds of relations at a time, a multi-head mechanism dispatches the workload on its different heads. Therefore, the attention heads work in parallel and each one of them focuses on a specific type of relation.

Technically, the input sequence to process is transformed into multiple sets of queries, keys, and values, with each set corresponding to a specific attention head. Each attention head then computes attention scores and generates its own output representation. The outputs from all attention heads are then concatenated or linearly combined to form the final representation of the sequence [37] (See Figure 3.3 for illustration).

On the other hand, the transformer architecture introduces a technique, called positional encoding, to retain the sequential order of information. The positional encoding is added to the input embeddings before feeding them into the model, allowing the transformer to differentiate between the positions of different elements.

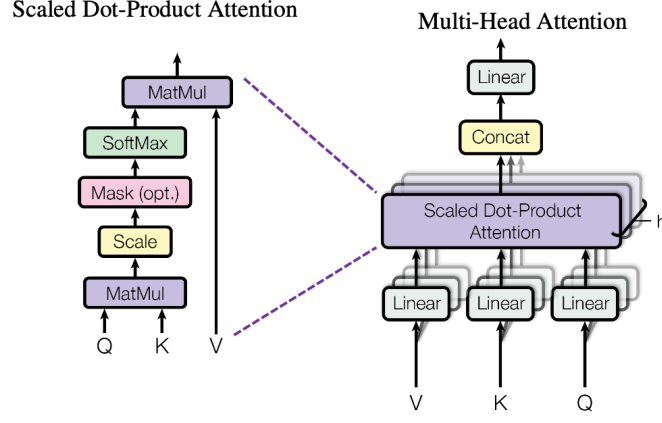


Figure 3.3: Multi-head attention mechanism [37].

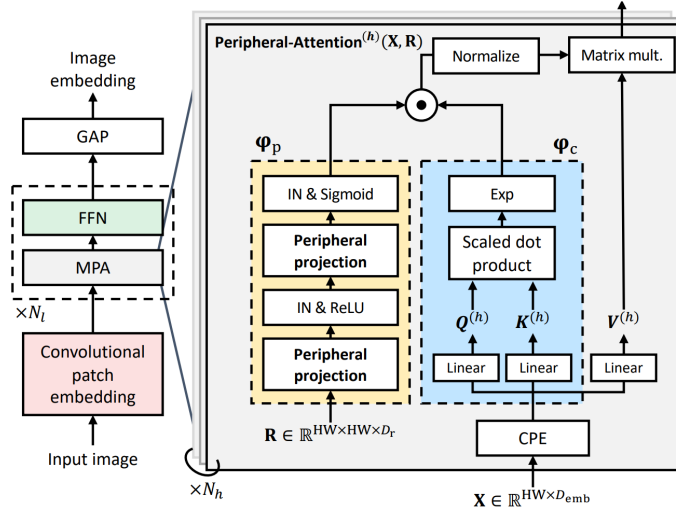


Figure 3.4: Global architecture of the ViT [14].

Figure 3.4 highlights the sequential processing of image patches through the transformer encoder and showcases how the ViT architecture can efficiently handle computer vision tasks by effectively modelling global dependencies and spatial information within the image data.

### 3.2.2 ViT model variants

Since the first apparition of the ViT model, it has undergone several improvements and adaptations. Hence, several variants of the model have been exhibited in the literature. We describe here the main ViT variants that have emerged since the publication of the original paper, and we present in Table 3.1 a comparison of the models according to given features.

- **Data-efficient image Transformers (DeiT):** proposed by Facebook AI. DeiT models are distilled vision transformers, which means that they are smaller and more efficient than the original ViT models, while still maintaining a high level of performance [33].
- **Patches for vision transformers (PVT):** proposed by Google AI. PVT models are a new type of ViT model that uses a hierarchical patch sampling scheme to improve the efficiency of the model [11].

- **Transformers-for-natural-and-turing-machine-tasks (TNT):** proposed by microsoft research. TNT models are a general-purpose transformer architecture that can be used for both vision and language tasks [40].
- **Swin transformer:** proposed by microsoft research. swin transformer is a hierarchical ViT model that uses a novel self-attention mechanism to improve the performance of the model [17].
- **Causal SWin transformer (CSWin):** proposed by Google AI. CSWin is a causal version of the swin transformer that is specifically designed for image classification tasks [7].
- **Next-generation Vision Transformer (Next-ViT):** developed by ByteDance. It is designed to be more efficient and accurate than previous ViT models notably for realistic industrial scenarios, while still maintaining a competitive parameter count [16].

Model	Depth	Hidden size	Number of heads	Parameters
ViT-Base	12	3072	16	110M
DeiT-Tiny	6	1024	8	8.5M
DeiT-Small	12	1024	16	30M
DeiT-Base	12	2048	16	88M
PVT-34	34	2048	8	77M
TNT-F	24	3072	16	162M
Swin Transformer-S	22	1024	4	257M
CSWin-S	18	1024	4	141M
ViT-G14	14	1408	16	184M
Next-ViT	16	2048	16	88M

Table 3.1: A comparison of the main ViT variants according to some features.

## Disadvantages of Vision Transformers (ViT)

Vision Transformers (ViT) has gained popularity for its success in computer vision tasks. While ViT offers several advantages, it also has some disadvantages:

- **High Computational Cost:** ViT models tend to be computationally expensive and require significant computational resources for training and inference. The large number of attention heads and parameters make them resource-intensive, limiting their use on less powerful hardware.
- **Large Memory Footprint:** ViT models have a large memory footprint due to their extensive self-attention mechanisms and deep architecture. This can be challenging for deployment in memory-constrained environments.
- **Limited Spatial Hierarchies:** Unlike Convolutional Neural Networks (CNNs), which naturally capture hierarchical features through layers, ViT relies solely on self-attention mechanisms. This can make it less effective in capturing spatial hierarchies in data, which are important for tasks like object detection.

- **Data Efficiency:** ViT models require large amounts of labelled data for training, often more than CNNs, to generalize well. They may not perform as effectively with small datasets.
- **Long Training Time:** Training large ViT models can be time-consuming, often requiring days or weeks on powerful hardware. This makes experimenting with ViT architectures slower compared to smaller models.
- **Difficulty with High-Resolution Inputs:** ViT models are initially designed for fixed-size square images, which can be limiting for tasks that involve high-resolution or non-square images. Variations like the DeiT (Data-efficient Image Transformer) have attempted to address this limitation.
- **Lack of Interpretability:** The self-attention mechanisms in ViT models can be challenging to interpret compared to the feature maps in CNNs. Understanding why the model makes a particular prediction can be less intuitive.
- **Fine-tuning Challenges:** Fine-tuning ViT models on custom datasets can be tricky, as they might not generalize as well as CNNs with transfer learning.
- **Not Always the Best Choice:** ViT models have shown great success in image classification tasks, but they might not always be the best choice for all computer vision tasks. Traditional CNN architectures can still outperform ViT in certain scenarios.

### 3.3 The object detection model Faster R-CNN

Faster R-CNN is an advanced object detection model inspired from the foundational R-CNN architecture. It has been fine-tuned and optimized for various complex detection tasks and was introduced to the deep-learning community in a pivotal research study by Girshick [10]. Characterized by its balance between speed and accuracy, Faster R-CNN has become a popular choice among researchers and developers aiming for precise real-time object detection. Structurally, Faster R-CNN integrates a cascading series of convolutional layers followed by Region Proposal Networks (RPN) and fully connected layers. While the convolutional layers are quintessential for distilling salient features from the input image, the RPN aids in hypothesizing object locations, and the fully connected layers consolidate this information, categorizing the identified objects.

An innovative facet of Faster R-CNN is the integration of region proposal networks (RPN). These networks streamline the process of generating high-quality region proposals, which are then utilized to pinpoint objects within the image. By automating this step, Faster R-CNN bypasses the need for external mechanisms or datasets for region suggestions, leading to a boost in both speed and accuracy.

The architectural elegance of faster R-CNN is divided into four pivotal segments: the backbone, the RPN, the ROI pooling module, and the head. Specifically, the backbone often based on powerful architectures like VGG16 or Next is tasked with feature extraction. The RPN, as mentioned, is responsible for generating object proposals. The ROI pooling module, unique to R-CNN variants, standardizes these proposals to a fixed size to allow for classification. Finally, the head of the architecture predicts precise bounding boxes and class scores for each proposed region.

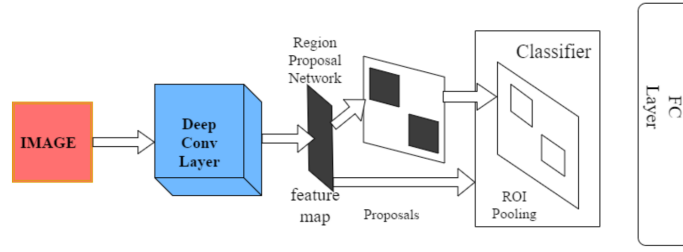


Figure 3.5: Faster R-CNN block architecture [13].

Thanks to its intricate design, faster R-CNN has consistently demonstrated superior performance in various object detection benchmarks. Its performance on datasets like Pascal VOC and COCO stands testament to its prowess. Moreover, its adaptability and optimization for real-time detection scenarios have been documented in numerous studies, emphasizing its robustness and versatility [32]. Figure 3.5 highlights the important parts of the Faster R-CNN architecture.

## 3.4 Attention maps

In the context of machine learning, attention maps are a visualization technique that can be used to understand how a neural network is processing input. Attention maps are typically heatmaps that show the relative importance of different parts of the input to the output of the network [42]. Attention maps are created by calculating the attention weights for each part of the input. The attention weights are a measure of how much the network is paying attention to each part of the input. The attention weights are then used to create a heatmap, where the darker the area, the more attention the network is paying to that part of the input. Attention maps can be created for any type of neural network, but they are most commonly used with Convolutional Neural Networks (CNNs) [28].

### 3.4.1 Attention rollout

Attention rollout is a novel approach to the automatic detection of cervical spine fractures. It is based on the vision transformer (ViT) architecture, which has been shown to be very effective for image classification tasks. Attention rollout is able to learn long-range dependencies in CT scans, which is important for the accurate detection of fractures. The main advantage of attention rollout is that it is very efficient. It can be trained on a small dataset of CT scans, and it can run on a standard laptop computer. This makes it a practical solution for the automatic detection of cervical spine fractures in clinical settings. The following are some of the key features of attention rollout:

- It is based on the ViT architecture, which has been shown to be very effective for image classification tasks.
- It is able to learn long-range dependencies in CT scans, which is essential for the accurate detection of fractures.
- It is very efficient, and it can be trained on a small dataset of CT scans.
- It can run on a standard laptop computer, which makes it a practical solution for the automatic detection of cervical spine fractures in clinical settings.



Attention rollout has the potential to revolutionize the way cervical spine fractures are diagnosed. It is a promising new technique that has the potential to improve the accuracy and efficiency of fracture detection.

### 3.4.2 Rollout matrices equation

The rollout matrices are used to propagate attention information through the image. At each time step, the rollout matrices are used to weigh the hidden state of each patch, and the weighted hidden states are then used to make a prediction. The next equation presents the rollout matrix:

$$R_t = \text{softmax} \left( \frac{W_t H_t}{\sqrt{d}} \right), \quad (3.1)$$

where:  $R_t$  is the rollout matrix at time  $t$ ,  $W_t$  is the attention weights at time  $t$ ,  $H_t$  is the hidden state at time  $t$ , and  $d$  is the dimension of the hidden state. The softmax function is used to normalize the attention weights so that they sum to 1. This ensures that each patch in the image is given a weight that reflects its importance for the current prediction.

## 3.5 Personal contributions

In light of the complexities associated with cervical spine fracture detection in medical imaging, there’s a compelling need for a multi-faceted approach that synergizes cutting-edge technologies in object detection, image classification, attention mechanisms, and cloud computing. This section serves as a precursor to the in-depth discussion on the data pipeline proposed in the setting of our work, laying the groundwork for our choice of technologies and methodologies.

Our proposed model aims to leverage the unique capabilities of Faster R-CNN for object localization, vision transformer’s (ViT) proficiency in image classification, and the added insight from attention maps to focus on regions of interest within images. These technologies are designed to work in concert, each offering its unique strengths to develop a system that is both highly accurate and computationally efficient.

Additionally, we opt for a cloud-based deployment to further amplify the model’s efficacy. Cloud computing brings unparalleled scalability, allowing our system to adapt to varying workloads effortlessly. It also provides easy accessibility, enabling healthcare professionals to access the system from multiple locations. Moreover, the cloud’s robust infrastructure facilitates real-time data analysis and model updating. Thereby ensuring that the system remains at the forefront of medical imaging technology.

### 3.5.1 A multifaceted computational pipeline for the detection and visualization of cervical spine fractures

As a first part of our personal contribution established in the setting of this work, we present a new data pipeline specifically designed for medical image analysis in view of detecting and interpreting cervical spine fractures. The exhibited process is essentially composed of six stages, namely: *volumetric image slicing*, *data augmentation before training Faster R-CNN*, *object localization using Faster R-CNN* and *image*

*cropping, data augmentation, classification via Next-ViT, and finally fractures presentation.* A schematic representation of the proposed framework is presented in Figure 3.6 and necessary details and descriptions about the proposed data pipeline are given in the subsections below.

### 3.5.1.1 Volumetric image slicing

In the complex domain of medical imaging, especially with high-dimensional data such as computed tomography (CT) scans, we are often confronted to volumetric (also called three-dimensional or 3D) images. These images offer a rich canvas of spatial information, capturing not just the height and width but also the depth of anatomical structures. However, this can be a double-edged sword; while it offers more information, it also increases the computational demands and complexity for subsequent analytical tasks. Consequently, the challenge faced when developing computerized approaches for treating such kind of medical images is how to harness the richness of this data without becoming ensnared in computational bottlenecks.

Herein lies the critical importance of the image-slicing process. In fact, the latter technique methodically dissects these volumetric images into a sequence of planar two-dimensional (2D) slices. By doing so, we transform an intricate 3D spatial problem into a more manageable 2D problem space. This is not merely a reductionist treatment, but a calculated methodological choice that offers several benefits. On one hand, slicing serves as a strategic maneuver to reduce computational cost and enhance efficiency. On the other hand, it prepares the ground for expeditious and focused downstream data processing. Specifically, the produced slices can be orientated to emphasize anatomical planes that are most relevant for the diagnosis of cervical spine fractures. This ensures that we retain the most pertinent and diagnostically relevant information in the slices.

Moreover, the yielded 2D slices are more compatible with established algorithms optimized for 2D data, such as Faster R-CNN for object detection, ViT for image classification and other approaches and techniques that are leveraged in subsequent stages of the presented data pipeline.

Furthermore, the 2D slices are inherently amenable to parallel processing techniques. This feature synergizes exceptionally well with cloud-based computational platforms, which are built to handle multiple tasks in parallel. Thus, speeding up the data processing even more.

In short, image slicing is a deliberately chosen, methodically executed process that serves as the linchpin for efficient, effective, and robust analysis in the detection of cervical spine fractures. It prepares the data for a complex journey through an integrated pipeline of object detection, attention mapping, and image classification, culminating in accurate and timely diagnoses.

### 3.5.1.2 Data augmentation before training Faster RCNN

Data augmentation is a powerful technique that can help to improve the performance of machine learning models by increasing the size and diversity of the training dataset. This is especially important for object detection models, such as Faster R-CNN, which require a large amount of labelled data to train effectively. In our pipeline, we mainly use random horizontal flipping, which can help the model learn to detect objects from different perspectives

### 3.5.1.3 Object detection using Faster R-CNN and region cropping

After the preparatory stage of image slicing and flipping, the regions of interest detection and isolation stage follow. At this juncture, Faster R-CNN comes to the forefront. Acclaimed for its harmonious blend of swiftness and accuracy, Faster R-CNN functions as a pivotal algorithmic instrument interfacing with the 2D slices previously extracted. The distinct prowess of Faster R-CNN lies in its ability to operate as a computational lens, scrupulously navigating through the slice images to discern and delineate regions that house medically pertinent structures. In particular, it identifies potential fracture sites within the cervical spine and underscores them with bounding boxes. This act of object localization constructs a vital foundational tier, guiding the ensuing procedures in the pipeline, which are designated to further refine, dissect, and classify these pronounced areas suspected of fractures.

Subsequent to the triumphant object localization via Faster R-CNN, region cropping emerges as the next cornerstone in our data processing chain. The aim of this phase is dual: firstly, to drastically curtail computational excess, and secondly, to concentrate the ensuing analysis on clinically pertinent regions, specifically, the sectors of the cervical spine believed to harbour fractures, as pinpointed by Faster R-CNN. Drawing from the bounding boxes discerned by Faster R-CNN, the slices are pruned to encapsulate these earmarked suspect regions exclusively. Through this, peripheral areas, which hold minimal or nil clinical significance, are purged, effectuating a considerable dip in data intricacy. This streamlined data configuration paves the way for brisker computations in the future and curbs the potential of noise infiltration in the upcoming pipeline stages.

In essence, region cropping acts as a strategic conduit connecting the preliminary object localization with the imminent analytical endeavours, assuring a computation process that's both agile and precision-focused.

### 3.5.1.4 Data augmentation

After the intricacies of region cropping, we arrive at a critical juncture in our pipeline the data augmentation procedures. The essence of this stage lies in enhancing the dataset's diversity without collecting additional data, a key element to improve model generalizability and mitigate the risks of overfitting. We employ a suite of carefully selected augmentation techniques to artificially expand the dataset. These techniques include but are not limited to rotation, scaling, and flipping of the isolated image regions. Rotation serves the purpose of accommodating the variability in patient posture or camera angle, ensuring that the model is resilient to such operational contingencies. Scaling addresses the variability in the size of the target structures due to patient-to-patient differences or imaging modalities. Flipping, on the other hand, offers a simple yet effective way to increase the dataset size, adding mirror versions of existing images.

Collectively, these data augmentation strategies introduce a stochastic element to the dataset, creating conditions for the model to learn from a broader range of scenarios. This becomes particularly crucial for medical imaging, where small variations can often lead to dramatically different clinical interpretations. In this manner, data augmentation acts as a safeguard against overfitting, enhancing the model's ability to generalize well to new, unseen data, thereby upholding the rigour and robustness of our cervical spine fracture detection system.

### 3.5.1.5 Classification via Next-ViT

In the next critical stage of our data pipeline, we leverage the Next-ViT model for binary classification, focusing on distinguishing between “fracture” and “no fracture” states in vertebral images. This model was selected for its unique set of attributes that align impeccably with our research goals. One of the standout qualities of the Next-ViT is its data efficiency. The model demonstrates impressive performance even when subjected to small, annotated datasets—a significant asset in medical applications where comprehensive datasets are a rarity. Given the underwhelming results of our initial attempt to train a vision transformer from scratch, we chose to adopt the pre-trained Next-ViT architecture, which led to a marked improvement in our system’s efficacy.

The Next-ViT diverges from conventional convolutional neural networks by incorporating self-attention mechanisms. These mechanisms excel at identifying complex spatial and contextual relationships within images, a feature invaluable for interpreting the complex imagery commonly found in cervical spine studies. Our training phase included an examination of data augmentation techniques. Interestingly, straightforward affine transformations outperformed automated methods in augmenting our dataset. We hypothesize that these simple transformations assist the model in understanding underlying data patterns, thereby improving its learning capability. In contrast, automated methods distorted the image characteristics, leading to overfitting and less accurate performance on test sets. The Next-ViT’s patch size of 16x16 offers an ideal balance between computational efficiency and detail resolution. This compromise makes the model highly applicable in clinical settings where timely and accurate diagnosis is paramount. Furthermore, the model adeptly incorporates various elements from the preceding stages of our pipeline (i.e. from image slicing to data augmentation) enhancing its role as a reliable and precise classifier.

In conclusion, the Next-ViT acts as the analytical keystone in our data pipeline, seamlessly translating previous data manipulations into clinically relevant insights. Its synergistic blend of computational efficiency, superior accuracy, and nuanced understanding of contextual details not only supplements the vertebrae regions identified by Faster R-CNN but also serves as the bedrock upon which our entire vertebral fracture detection system is built. Through its application, we fulfill our research objectives, providing reliable and empirically validated diagnostic conclusions.

### 3.5.1.6 Fractures presentation

As a final stage in our carefully constructed pipeline, we engage with one more critical phase, namely: fractures presentation using attention maps. The deployment of attention maps serves a dual purpose. On the one hand, it accentuates and highlights specific areas of the images that the model deems as areas of interest or focus points. This contributes to making the model’s decision-making process interpretable, a trait often desired in medical applications for its ability to provide clinicians with a point of reference or justification. On the other hand, attention maps also enhance the model’s accuracy by providing it with context, which in turn allows for a more nuanced detection of fractures. This is particularly beneficial in complex cases where fractures may be less apparent or exist in spatial configurations that are difficult to interpret. The attention map technique dovetails neatly with the Vision Transformer’s innate capacity for feature extraction and the Faster R-CNN model’s object localization capabilities. It brings the essence of both into

sharp focus, thereby providing a holistic, contextual view that is both accurate and interpretable. Thus, the incorporation of attention maps serves as the final touch, fine-tuning our pipeline’s outputs and enabling a more precise and comprehensive presentation of cervical spine fractures. It’s a harmonious integration, providing clinicians not only with a diagnosis but also a nuanced, attention-weighted view that can be crucial for effective patient management.

In summary, with each of its designed components—from image slicing to attention mapping—the pipeline serves as a magnum opus of interdisciplinary ingenuity, presaging a transformative impact on both the realms of medical diagnostics and artificial intelligence.

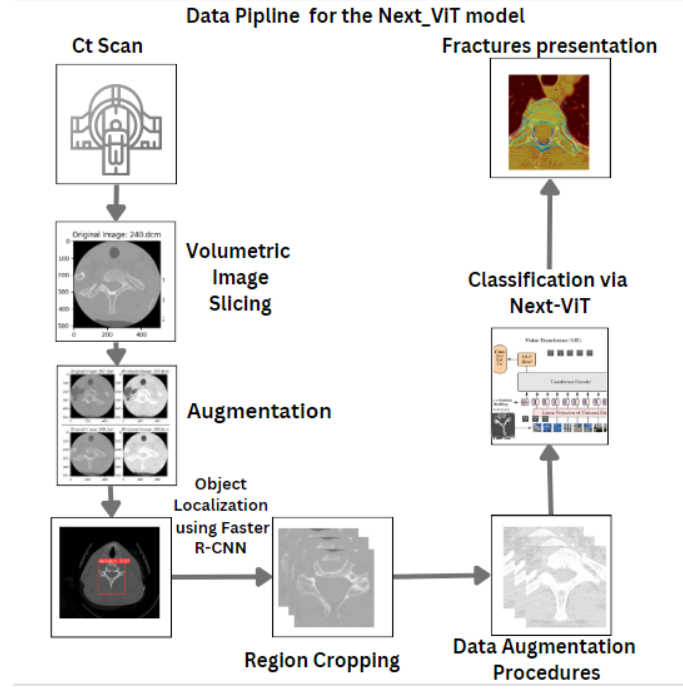


Figure 3.6: A representation of the proposed data pipeline.

### 3.5.2 Proposed cloud-based system for cervical spine fracture detection

A second part of our personal contribution, subsequent to the presentation of the new data pipeline, we introduce and implement a robust and scalable cloud-based system that is dedicated to the detection of cervical spine fractures. The cloud infrastructure serves as the backbone supporting the entire analytical pipeline and offers unique advantages both in terms of computational resources and data management.

#### 3.5.2.1 Motivations and goals

Our research endeavours are galvanized by several compelling incentives for integrating cloud computing and Vision Transformer (ViT) models in the arena of cervical spine fracture detection. The impetus for adopting a cloud-based approach originates from a critical need to address challenges in scalability, data integrity, and real-time analytics. Below are the main key motivations:

1. **Superior diagnostic accuracy:** Amalgamation of cloud computing and ViT models promises unprecedented precision in detecting cervical spine fractures.

Traditional diagnostic approaches, although useful, sometimes fail to identify complex or subtle fractures. The marriage of cloud-based computational power and advanced machine learning models has the potential to usher in a new era of nuanced and precise diagnoses.

2. **Operational efficiency:** Utilizing the distributed computing power of the cloud alongside ViT models that can efficiently parse large sets of image data enhances the operational efficiency of the diagnostic process. This could significantly reduce the time radiologists need to reach a diagnosis.
3. **Scalability and adaptability:** The inherent scalability of cloud infrastructure is well-suited for handling the voluminous medical imaging data generated daily. This removes the need for healthcare organizations to make significant investments in local computing resources.
4. **Broadened access to advanced tools:** Cloud-based systems democratize access to cutting-edge diagnostic technologies. This model allows healthcare providers, regardless of their size or location, to benefit from state-of-the-art tools without prohibitive upfront costs.
5. **Augmentation of clinical decision-making:** The synergy between cloud technology and ViT models can act as a potent decision-support mechanism. It can provide preliminary evaluations that assist healthcare professionals in making timely and well-informed decisions.
6. **Future-ready integration:** The modular architecture of cloud-based systems makes them ripe for seamless integration with existing electronic health records. This offers the possibility for more integrated, collaborative approaches to healthcare delivery in the future.

Thus, the integration of cloud computing and ViT models in the detection of cervical spine fractures has the potential to surmount existing limitations, refine diagnostic protocols, democratize access to state-of-the-art technologies, and fundamentally transform clinical practices in this vital area of healthcare.

### 3.5.2.2 Description of the proposed cloud-based system

In this modern era, harnessing the power of cloud computing and artificial intelligence can significantly streamline and optimize medical diagnostic processes. Our proposed architecture is designed on the Google Cloud Platform (GCP), integrating its services to offer an efficient end-to-end cervical spine fracture detection workflow.

The profound implications of seamlessly blending cloud infrastructure with advanced machine learning paradigms cannot be understated. Our intrinsic role spanned the breadth of this ambitious endeavour, from its very conceptualization to its real-world implementation and continuous evolution. A general view of the proposed cloud-based architecture for cervical spine fracture detection is illustrated in Figure 3.7.

Initially, the overarching vision of crafting an integrated end-to-end diagnostic workflow for enhanced cervical spine fracture detection stemmed from comprehensive brainstorming sessions. Significantly, it was our deep dive into the vast capabilities of the Google Cloud Platform (GCP) that galvanized our alignment with this mission. Building upon this foundation, our hands played a pivotal role in the ensuing architectural design and execution phase.

Furthermore, recognizing the paramount importance of data integrity, we have channelled significant efforts into devising an efficient automatic ingestion mechanism for CT scans. Simultaneously, with an acute awareness of the sensitive nature of medical data, we have championed the incorporation of a robust encryption protocol, ensuring that data remain secured.

Transitioning from data acquisition, our focus then have gravitated towards the multi-layered data pipeline. Specifically, we have integrated in the proposed architecture the data pipeline elaborated in Subsection 3.5.1, that meticulously optimize the mechanisms of preprocessing, feature extraction, and fracture detection.

On the other hand, we believe in the interdependent nexus between machine learning methods and human expertise and its capacity to offer better solutions, especially when they are combined appropriately. This conviction has led to the establishment of a systematic feedback loop, where the invaluable insights of medical professionals continuously enrich our cloud-based system. Through this mechanism, their diagnostic evaluations directly inform and steer the iterative enhancements of the integrated models in the proposed system.

Moreover, with an ever-evolving medical landscape, we need to ensure that the used models in the architecture underwent consistent training sessions. By leveraging insights from the analytical database, our diagnostic algorithms remain at the cutting edge, always adaptive to the latest nuances in medical diagnostics.

Beyond the technical realm, we endeavour to foster a culture of interdisciplinary collaboration. By orchestrating synergy between cloud experts, data scientists, and medical professionals, we strive to ensure that our collective expertise coalesced seamlessly. This unity of purpose and knowledge-sharing became instrumental in shaping our presented solution.

In summary, witnessing the transformative potential of our architecture in the realm of medical diagnostics has been both a privilege and a testament to the collaborative prowess of our team. Our journey exemplifies the boundless possibilities that emerge when cloud computing and machine learning converge, especially in the ever-critical domain of healthcare.

The amalgamation of GCP’s advanced services presents a promising horizon for medical diagnostics. While this overview provides a high-level design, the actual implementation should be tailored according to specific requirements, ensuring a balance between functionality, budget, and privacy concerns. Collaboration with cloud and domain experts is essential for the successful realization of such a system.

## 3.6 Conclusion

In this chapter, we have presented a couple of contributions. Mainly, we have introduced a new comprehensive computational data pipeline tailored for the detection and visualization of cervical spine fractures. Specifically, the proposed data pipeline is composed of six vital stages, each of which fulfils a unique role to achieve high diagnostic precision and reliability. Furthermore, motivated by several key goals such as improving diagnostic accuracy, increasing scalability, and enhancing data security, we have exhibited, as our second main contribution, a new cloud-based system to extend the capabilities of our computational data pipeline. The new cloud-based architecture represents a paradigm shift in how cervical spine fractures can be detected and managed. The proposed cloud-based system not only streamlines the workflow but also allows for continuous improvement through real-time feedback

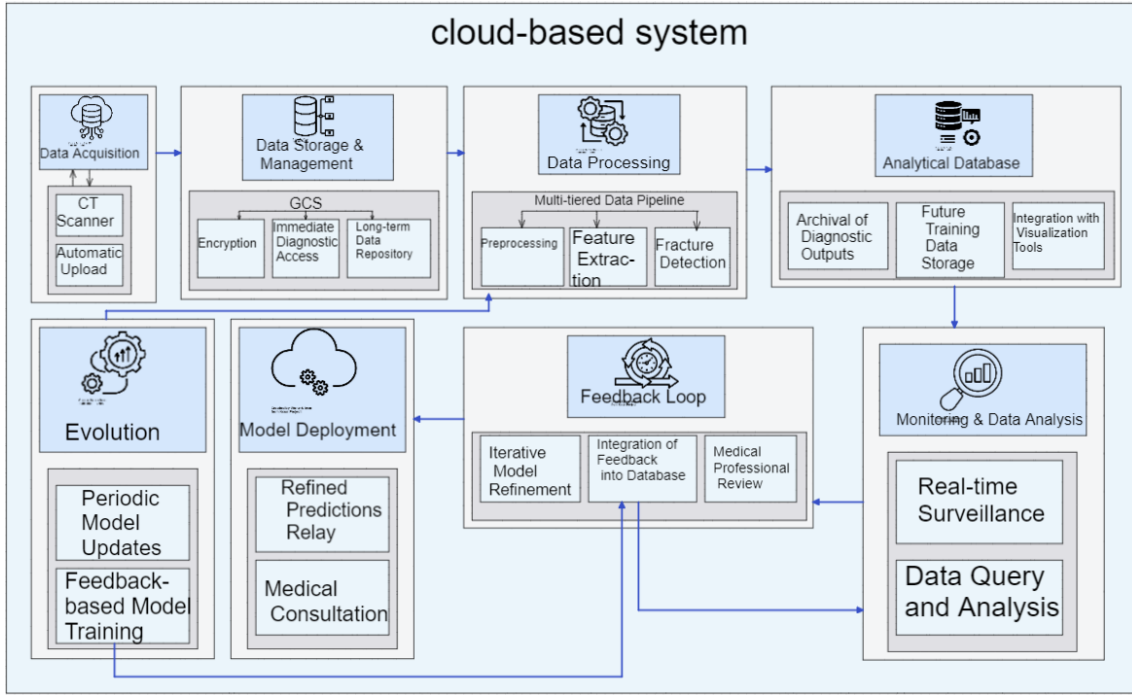


Figure 3.7: A representation of the proposed cloud-based system.

mechanisms.

The next chapter will be dedicated to the experimental validation of the performance of the data pipeline exhibited in the context of this chapter.



# Chapter 4

## Tests and evaluations of the proposed data pipeline

### 4.1 Introduction

Chapter four stands as a critical juncture in our journey to revolutionize the detection of cervical spine fractures through cutting-edge machine learning technologies. At the heart of this mission is the careful orchestration of data acquisition, management, and analysis, which sets the stage for model implementation and evaluation. This chapter meticulously unpacks the various layers of our research, from data management to model implementation, offering an expansive view of our investigative process.

In the first section we set the context by delineating the Programming Language and Development Environment that form the bedrock of our research. This study predominantly utilizes Python, owing to its extensive ecosystem that includes indispensable libraries such as NumPy, Pandas, and PyTorch. Furthermore, the development environment is facilitated by Kaggle’s cloud-based platform, equipped with NVIDIA Tesla T4 GPUs to expedite computational tasks.

Next, in the Dataset Description and Exploration section [4.2](#), we discuss the data landscape, detailing the sources, types, and characteristics of the datasets we have employed. Following closely is the section on Exploratory Data Analysis, where we dive into the nuances of the dataset, including segmentation masks, training images, and associated data frames. This sets the stage for the Visualization of the Dataset, where we employ various techniques such as correlation analysis and heatmap visualizations to further understand the data.

Data Pipeline Implementation and Training serves as the next focal point [4.3](#). This section delves into the intricacies of implementing our data pipeline and the training strategies adopted for optimal model performance. From volumetric image slicing to vertebrae detection using Faster R-CNN, we unpack each step of our implementation strategy.

Subsequently, in section [4.4](#) we shift our focus to Model Evaluation and Performance Metrics, providing an in-depth analysis of how the model’s effectiveness is quantified. Special attention is given to unique features like Fracture Presentation using Attention Maps [4.5](#), which offer an added layer of interpretability.

The chapter concludes with a Summary and Discussion section [4.6](#) where we weave together the essential findings, methodologies, and implications of our work. In doing so, we aim to offer a comprehensive snapshot of the various critical aspects that underlie a successful application of machine learning in medical imaging

technology.

To sum up, this chapter serves as a comprehensive guide that navigates through the multifaceted landscape of data management, pipeline implementation, and model evaluation each an integral part of our overarching endeavour to advance medical imaging diagnostics.



Figure 4.1: Kaggle and Python logos.

## 4.2 Dataset description and exploration

The cornerstone of this research lies in a rigorously curated dataset that is essential for advancing the detection and localization of cervical spine fractures. This dataset emanates from a collaboration with esteemed organizations such as the Radiological Society of North America (RSNA), the American Society of Neuroradiology (ASNR), and the American Society of Spine Radiology (ASSR). The dataset is publicly accessible and can be downloaded from [Kaggle's official web page](#). It consists of approximately 3,000 Computed Tomography (CT) studies, sourced from twelve international locations and expertly annotated by radiology specialists from the contributing organizations. These annotations define the ground truth and indicate the presence, vertebral levels, and specific localizations of cervical spine fractures. Pre-processing measures include converting images from DICOM and NIFTI formats into a unified analytical framework, as well as precise alignment and metadata management. Each step was executed with strict quality control measures to ensure scientific rigour and clinical relevance.

### 4.2.1 Exploratory data analysis

In this section, we embark on a journey of exploratory data analysis, exploring our dataset to extract meaningful insights and pave the way for informed decision-making. This stage involves a careful and systematic examination of the data, unearthing hidden patterns, trends, and anomalies that hold the key to unlocking the secrets within. Through a series of analytical techniques and visualizations, we illuminate the terrain of our data, shedding light on its characteristics and guiding us toward a deeper understanding of our subject matter.

Understanding the dataset's composition is not just a formality but a cornerstone upon which meaningful insights and innovations are built. Let us embark on a granular examination of the various dataset components, each contributing uniquely to the overarching goals of our project in the realm of medical imaging.

#### 4.2.1.1 Segmentation masks

There are 87 segmentation masks, usually in NIFTI (.nii) format, a standard for storing medical imaging data. These masks are crucial for tasks that require high

localization accuracy, such as identifying the exact contours of a fracture. This relatively small number suggests that the dataset may primarily focus on key or challenging cases requiring precise segmentation for training. It is important to investigate whether these cases are representative or outliers.

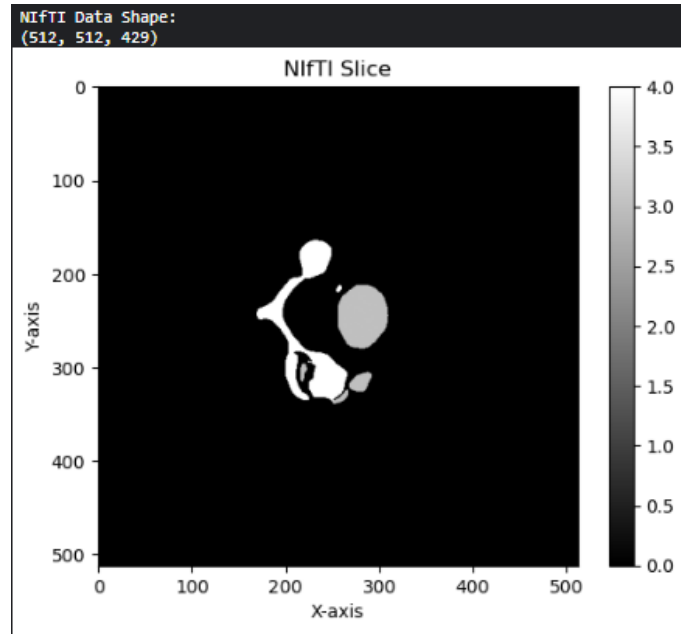


Figure 4.2: NIFTI segmentation reveals cervical spine fractures.

Figure 4.2 displays a 2D slice of a CT scan of the cervical spine, which is the neck region of a patient. The CT scan data was stored in the NIFTI file format, which is commonly used in neuro-imaging research and medical imaging. The NIFTI image contains valuable information about the patient’s cervical spine, particularly focusing on the vertebral structures. The CT scan was acquired with a slice thickness of  $\leq 1$  mm and an axial orientation, resulting in detailed cross-sectional images of the cervical vertebrae. To visualize the image, we used Python programming and the NiBabel library to read and extract the data from the NIFTI file. The resulting 2D slice is a single layer taken from the middle of the 3D volume of the cervical spine. It can be thought of as slicing the cervical spine like a loaf of bread and examining a single slice. The grayscale image is displayed with a bone kernel, enhancing the visibility of bone structures, such as the cervical vertebrae. The pixel values in the image represent the radiodensity of the tissues, with darker areas indicating regions with higher density, such as bones.

#### 4.2.1.2 Train images

The dataset is significantly weighted towards the training set, with 711,601 images also in DICOM format. This vast number suggests that the dataset could likely capture a wide range of cases. Thus providing a robust foundation for training machine learning models. However, the large size also poses challenges in terms of computational resources for training and the risk of overfitting if not managed correctly.

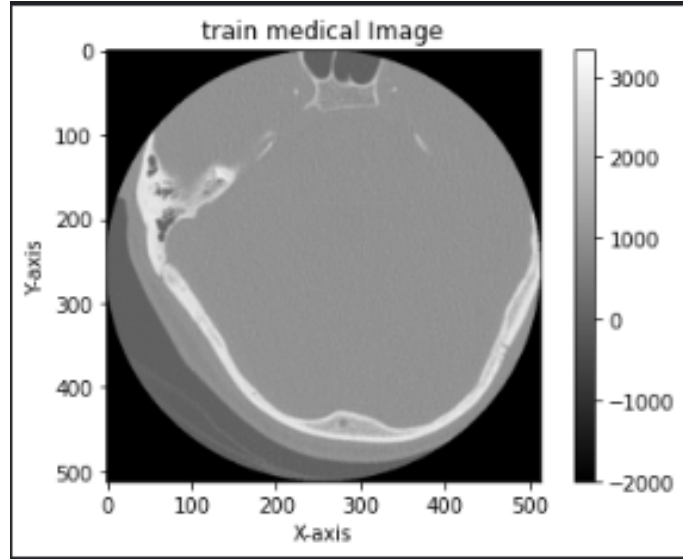


Figure 4.3: One of the train images for a patient.

Figure 4.3 represents a single slice of the 3D CT scan of the cervical spine (train images), and to fully understand the patient’s condition, medical professionals would analyze multiple slices and other relevant information in the context of the patient’s medical history and symptoms.

#### 4.2.1.3 Train dataframe

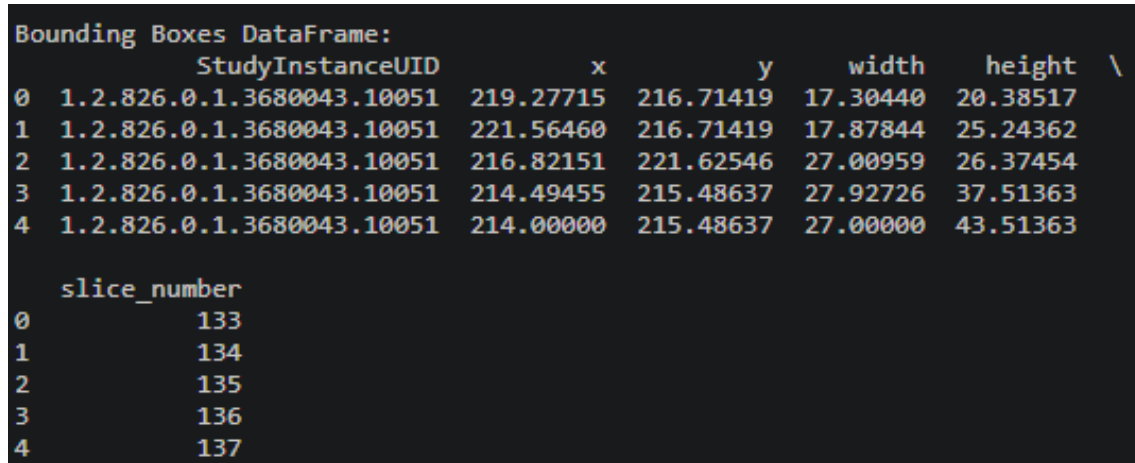
The train data frame provides essential information about the training data. It consists of 2019 rows, each representing a unique study instance. The columns “StudyInstanceUID” and “patient\_overall” are used for study identification and patient-level information, respectively. However, the most critical columns in this dataframe are “C1” to “C7”, indicating fractures in different cervical spine regions. For example, in Figure 4.4 which provides an overview of the traindata frame related to cervical spine fracture detection, there is a fracture in the C1 region for the study instance with StudyInstanceUID 1.2.826.0.1.3680043.6200.

Train DataFrame:									
	StudyInstanceUID	patient_overall	C1	C2	C3	C4	C5	C6	C7
0	1.2.826.0.1.3680043.6200	1	1	1	0	0	0	0	0
1	1.2.826.0.1.3680043.27262	1	0	1	0	0	0	0	0
2	1.2.826.0.1.3680043.21561	1	0	1	0	0	0	0	0
3	1.2.826.0.1.3680043.12351	0	0	0	0	0	0	0	0
4	1.2.826.0.1.3680043.1363	1	0	0	0	0	1	0	0

Figure 4.4: Train data frame of the cervical spine dataset

#### 4.2.1.4 Train bounding boxes dataframe

The bounding boxes dataframe is a table that contains information about the bounding boxes of fractures in the training data. Each row in the dataframe represents a single bounding box, and the columns provide information about the location and dimensions of the bounding box and the study instance associated with the bounding box. The `StudyInstanceUID` column identifies the study instance associated with the bounding box. This information can be used to retrieve the corresponding images from the dataset. The “x” and “y” columns represent the coordinates of the top-left corner of the bounding box. The “width” and “height” columns represent the dimensions of the bounding box. The “slice\_number” column indicates the slice number of the bounding box.



```
Bounding Boxes DataFrame:
   index  StudyInstanceUID      x      y    width  height \
0      0  1.2.826.0.1.3680043.10051  219.27715  216.71419  17.30440  20.38517
1      1  1.2.826.0.1.3680043.10051  221.56460  216.71419  17.87844  25.24362
2      2  1.2.826.0.1.3680043.10051  216.82151  221.62546  27.00959  26.37454
3      3  1.2.826.0.1.3680043.10051  214.49455  215.48637  27.92726  37.51363
4      4  1.2.826.0.1.3680043.10051  214.00000  215.48637  27.00000  43.51363

   slice_number
0            133
1            134
2            135
3            136
4            137
```

Figure 4.5: Train bounding boxes dataframe of the cervical spine dataset

Figure 4.5 gives an overview of the train bounding boxes dataframe related to cervical spine fracture detection.

By leveraging the provided data and utilizing advanced algorithms, we aim to build a highly accurate and reliable system that can significantly impact patient care and improve medical outcomes.

#### 4.2.1.5 Test images

The dataset contains 1,318 test images, stored in the DICOM (.dcm) format, a widely used format in medical imaging. The purpose of these images is to assess the model’s performance on unseen data. However, given the significant imbalance between the size of the test and training datasets, care should be taken when interpreting model performance metrics calculated using this test set.

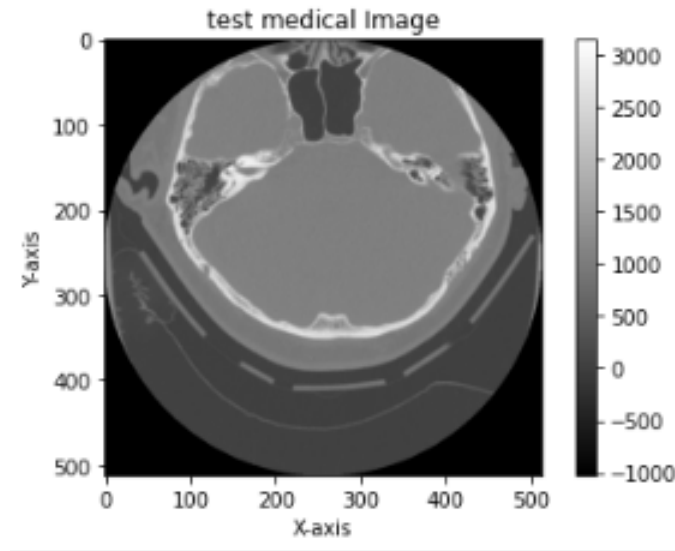


Figure 4.6: Presented fracture in a vertebral fracture for a patient.

Figure 4.6 provides valuable pieces of information about the internal structures of the cervical spine. Bones and dense structures will appear brighter due to their higher density, while softer tissues like muscles and organs will appear darker.

#### 4.2.1.6 Test dataframe

The test dataframe contains information about the test data that will be used for model evaluation. It has 3 rows, each representing a specific slice or region within a study instance for which fracture predictions must be made. The “row\_id” column uniquely identifies each entry, while the “StudyInstanceUID” column identifies the study instance associated with the test data. The “prediction\_type” column indicates the cervical region for which the model needs to make predictions. For instance, the first row in the test dataframe shown in Figure 4.7 indicates that the model should predict that there is a fracture in the C1 region for the study instance with the StudyInstanceUID 1.2.826.0.1.3680043.10197.

Test DataFrame:				
	row_id	StudyInstanceUID	prediction_type	
0	1.2.826.0.1.3680043.10197_C1	1.2.826.0.1.3680043.10197	C1	
1	1.2.826.0.1.3680043.10454_C1	1.2.826.0.1.3680043.10454	C1	
2	1.2.826.0.1.3680043.10690_C1	1.2.826.0.1.3680043.10690	C1	

Figure 4.7: Test dataframe of the cervical spine dataset.

## 4.2.2 Dataset’s visualization

Dataset visualization is representing data in a graphical format that makes it easier to understand and interpret. It is a powerful tool for exploring data, identifying patterns, and communicating findings.

#### 4.2.2.1 Relationships through correlation analysis and heatmap visualization

The heatmap visualization enhances this interpretation by using colour gradients. Warmer colours, such as red and orange, represent higher positive correlations, while cooler colours like blue represent negative correlations. The intensity of the colour reflects the strength of the correlation.

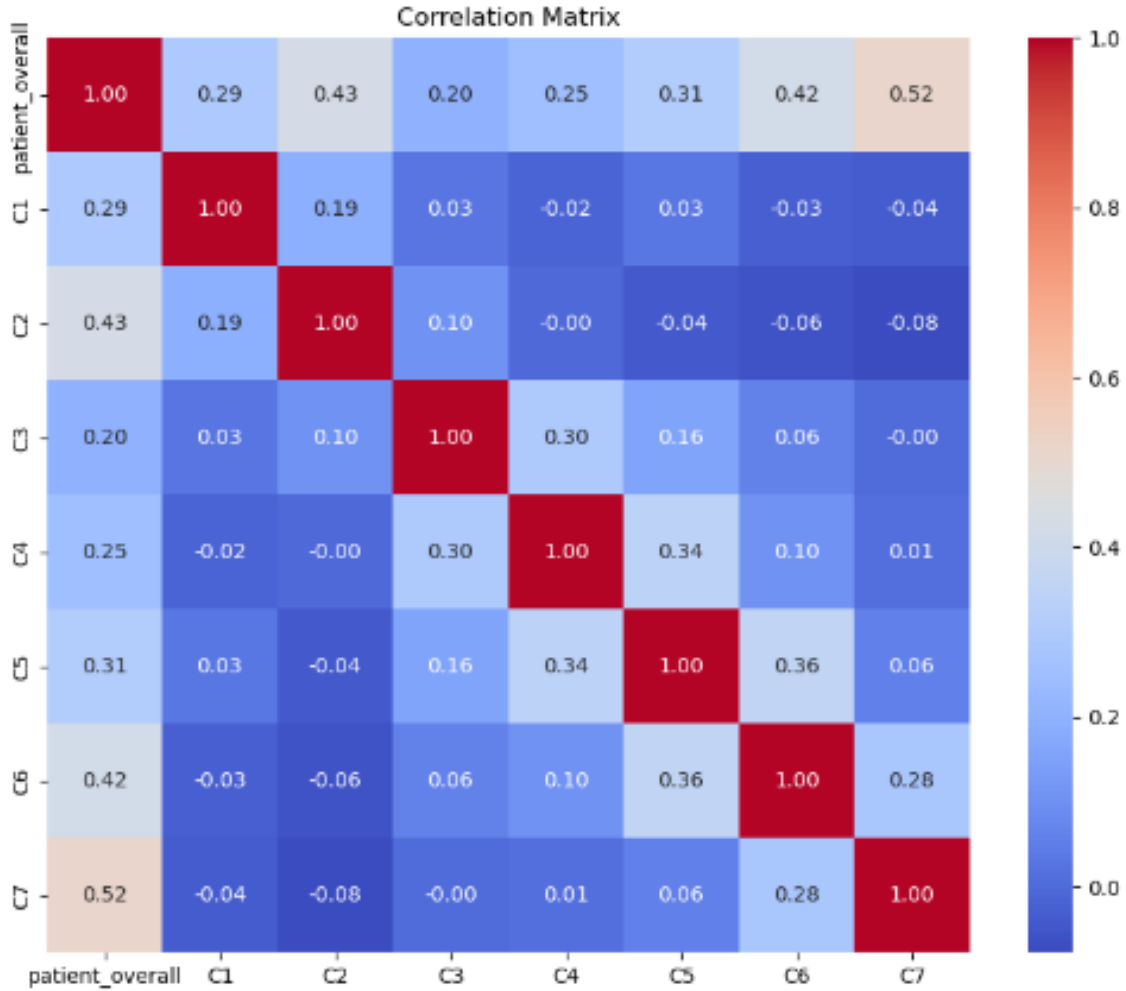


Figure 4.8: The resulting correlation matrix.

The resulting correlation matrix is illustrated in Figure 4.8. Specifically, it provides a table of correlation values for pairs of selected numerical columns. Each row and column corresponds to a specific column, and the cell values represent the strength and direction of the correlation between those two columns. Correlation values range from -1 to 1, with -1 indicating a perfect negative correlation, 1 indicating a perfect positive correlation, and 0 indicating no correlation. The correlation matrix includes columns such as “patient\_overall”, “C1”, “C2”, “C3”, “C4”, “C5”, “C6”, and “C7”. For example, the value at row “patient\_overall” and column “C1” intersection is approximately 0.2929. This indicates a moderate positive correlation between the “patient\_overall” and “C1” columns. Ultimately, this understanding forms a cornerstone for robust data-driven decision-making and rigorous model building.

#### 4.2.2.2 Relationships between pairs of variables of the dataset

The pair plot consists of a grid of scatterplots and density plots. Each row and column in the grid corresponds to a different variable in the dataset. The scatterplots on the diagonal represent the distributions of each individual variable, displayed as kernel density estimates. These plots illustrate how the data is distributed across various values of the variable, providing insights into its underlying structure.

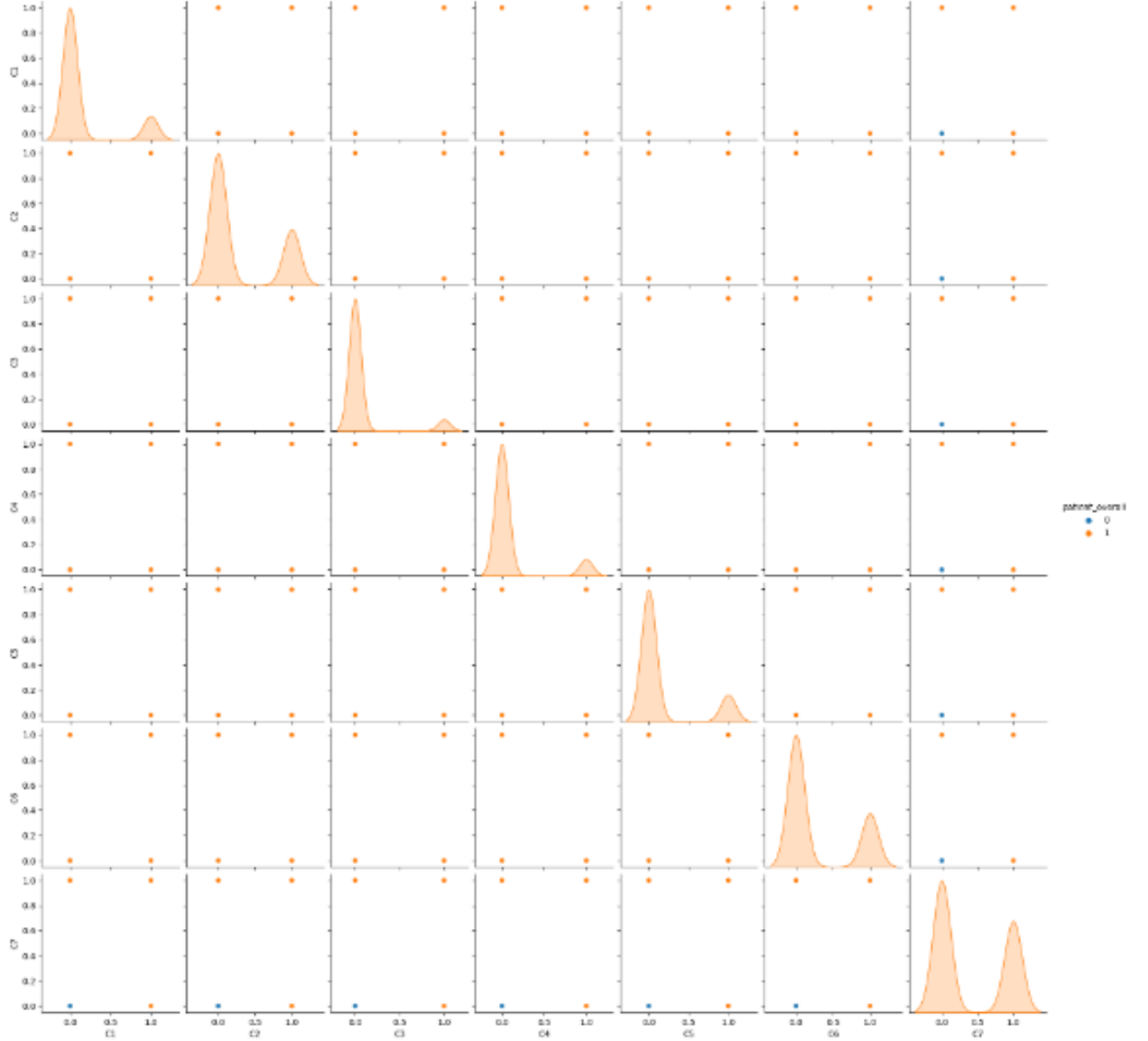


Figure 4.9: Visual representation of relationships between C1,...,C7 exploration.

The off-diagonal scatterplots showcase the relationships between pairs of variables. Each point in these scatterplots represents a data point, and the position of the point on the x-axis corresponds to the value of one variable, while the position on the y-axis corresponds to the value of the other variable. The points are color-coded based on the “patient\_overall” column, which adds an additional dimension of information to the plot. For illustration, Figure 4.9 gives a representation of relationships between C1,..., and C7 exploration. The hue-based colour coding allows us to observe how different outcomes (fractured or not fractured) are distributed across the scatterplots. This aids in identifying potential trends, clusters, or patterns that might differ between these outcomes. When points of a specific hue cluster together



or exhibit certain trends, it suggests a potential relationship or association between the variables for that particular outcome.

This pair plot provides a visual summary of the relationships between various pairs of variables in the dataset and offers initial insights into how these variables might be connected. It serves as a starting point for deeper analysis and hypothesis formulation regarding the potential impact of different variables on the overall outcome of interest, which is the presence or absence of fractures.

### 4.2.3 Observations and implications

The dataset is both rich and complex. However, this complexity brings with it a set of challenges and implications that warrant careful consideration. Here, we outline key observations and their potential impact on our study:

- **Imbalance between train and test sets:** The number of train images (711,601) and one of the test images (1,318) show a significant imbalance. It may make evaluation challenging because the test set might not sufficiently represent the distribution of the whole dataset.

```
[17]: import glob
import pydicom
# Using glob to get all .dcm files from the specified directories
train_dicom_files = glob.glob("/kaggle/input/rsna-2022-cervical-spine-fracture-detection/train_images/**/*.dcm")
test_dicom_files = glob.glob("/kaggle/input/rsna-2022-cervical-spine-fracture-detection/test_images/**/*.dcm")

# Check if list is empty
if not train_dicom_files:
    print("No training DICOM files found.")
else:
    first_train_file = pydicom.dcmread(train_dicom_files[0])
    height = int(first_train_file.Rows)
    width = int(first_train_file.Columns)
    depth = len(train_dicom_files)
    print(f"Train Depth: {depth}, Height: {height}, Width: {width}")

if not test_dicom_files:
    print("No test DICOM files found.")
else:
    first_test_file = pydicom.dcmread(test_dicom_files[0])
    height = int(first_test_file.Rows)
    width = int(first_test_file.Columns)
    depth = len(test_dicom_files)
    print(f"Test Depth: {depth}, Height: {height}, Width: {width}")

Train Depth: 711601, Height: 512, Width: 512
Test Depth: 1318, Height: 512, Width: 512
```

Figure 4.10: A comprehensive representation of the size and shape of the data.

- **Data dimensionality:** This information can be useful for understanding the size and shape of the data, which can be important for downstream tasks such as model training and evaluation. The dimensionality of our data images is represented in Figure 4.10.
- **Sparse test data:** With only 3 rows in the test dataframe, this could be a simplified representation or an oversight. It is strikingly low and may not be adequate for model validation.
- **Bounding boxes for object localization:** With 7,217 bounding boxes provided in the training set, the object localization task should be adequately

facilitated. However, the balance between this and the number of training images will need to be examined.

- **Segmentation masks:** The 87 segmentation masks suggest that only a subset of the images is being used for more refined, pixel-level analysis. It's important to understand how these link back to the broader training dataset.
- **Complex labels:** The training dataframe suggests that each vertebra (C1 to C7) is labelled, which indicates a multi-label classification problem.
- **Data quality:** No missing or null values are mentioned, which is a good sign for data quality. But further checks are always recommended.

### 4.3 Data pipeline implementation and training

In this section, we delve into the detailed steps taken to train the data pipeline model for cervical spine fracture detection. To recall, the data pipeline consists of several crucial stages, including volumetric image slicing, vertebrae detection using Faster R-CNN, and model training using Next-ViT. Each of these stages is vital for the successful application of the proposed data pipeline.

#### 4.3.1 Volumetric image slicing

Volumetric image slicing is the first critical step in the model. It is a step that regroups the methods used for preparing the CT scan slices for analysis. In this initial phase of the pipeline, slices are extracted from the original DICOM files of CT scans using a function called `process_and_plot_slices`. This function employs the `pydicom.dcmread` method to read the DICOM files and extract the pixel array in  $512 \times 512$  pixels in size, which is then resized to  $224 \times 224$  pixels to match the input size expected by the Next-ViT model. Figure 4.11a shows an extracted slice of a CT scan and Figure 4.11b exhibits its corresponding resized slice.

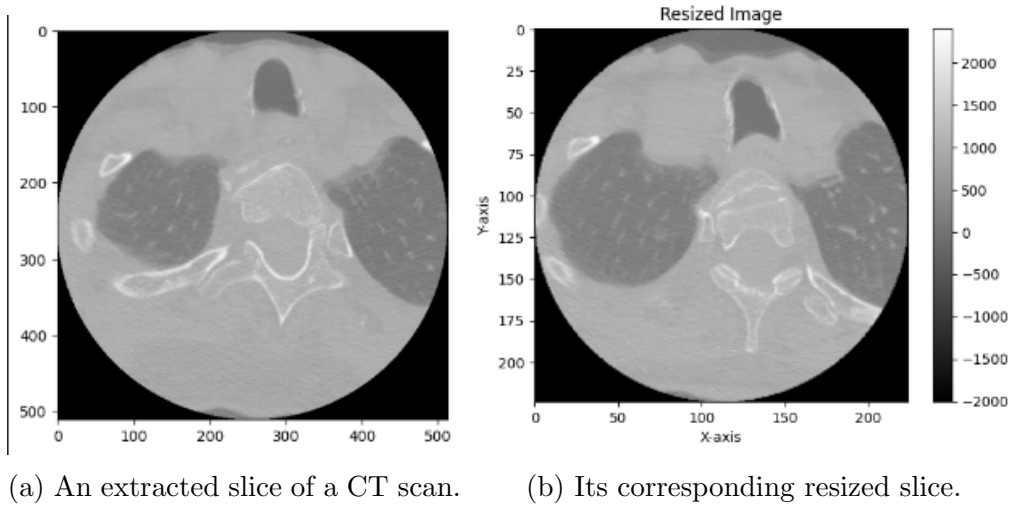


Figure 4.11: An extracted slice of a CT scan alongside its corresponding slice after resize operation.

Moreover, windowing techniques are applied to enhance contrast. The window width and level are set to 1800 and 400, respectively. Figure 4.12 presents a slice after applying windowing operation.

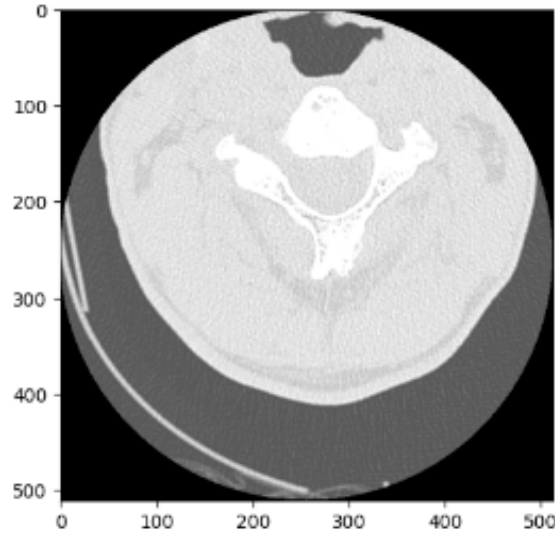


Figure 4.12: A slice after windowing operation.

### 4.3.2 Vertebrae detection using Faster R-CNN and region cropping

After the pre-processing step (i.e. the volumetric image slicing step), Faster R-CNN is employed to detect vertebrae in the image slices. Therefore, after the application of Faster R-CNN, a bounding box is delineated around the detected vertebra of each slice. Subsequently, to streamline the training process and reduce computational complexity, the area of interest (i.e. the vertebra) within the slice is cropped based on the delineated bounding box.

For a visual illustration, we give in Figure 4.13 a representation of a bounding box around a vertebrae-segmented region.

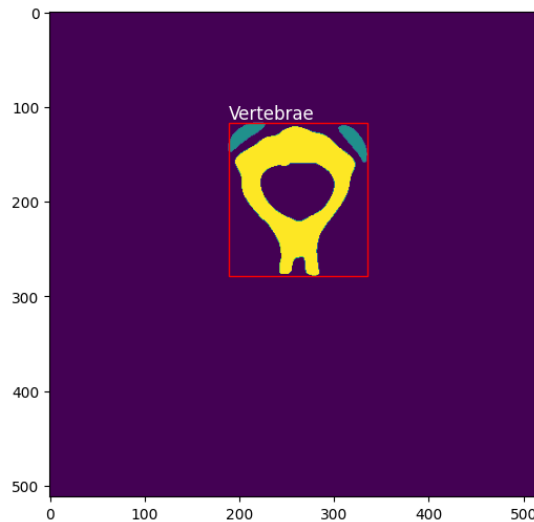


Figure 4.13: Bounding box around vertebrae segmented region.

The code for drawing a bounding box on an original CT scan is given as follows:

```
# Create the plot
fig, ax = plt.subplots(1, 1, figsize=(6,6))
```

```

# Display the windowed CT slice
ax.imshow(windowed_CT_slice, cmap=plt.get_cmap('bone'))

# Create a red rectangle (bounding box) around the object
rect = Rectangle((cmin, rmin), width, height, linewidth=1, edgecolor='r', facecolor='none')
ax.set_title('Vertebra Bounding Box (Windowed)')
# Add the rectangle to the plot
ax.add_patch(rect)

# Optionally, print bounding box dimensions
print(cmin, rmin, width, height)

# Show the plot
plt.show()

# Test to see if saved coords are correct

l = os.listdir('/kaggle/working/frcnnob_coords')
l = sorted(l)
l.sort(key=len)

f = np.random.choice(l)
print(f)
pt_num = re.search("^[0-9]*(?=_)", f).group(0)
slice_num = re.search("(?<=)([0-9]*(?=\.txt)", f).group(0)

CT_path = os.path.join(rsna_root, 'train_images', "1.2.826.0.1.3680043."+pt_num)
CT_arr = CT_path_to_3D_arr(CT_path)
CT_slice = CT_arr[int(slice_num)]
img_width = CT_slice.shape[1]
img_height = CT_slice.shape[0]

fig, ax = plt.subplots(1,1,figsize=(6,6))
ax.imshow(CT_slice, cmap=plt.get_cmap('bone'))

p = '/kaggle/working/frcnnob_coords/'+f
with open(p, 'r') as txt_file:
    reader = csv.reader(txt_file)
    row = next(reader)

row = [float(num) for num in row[0].split()]
bbox_xcentre = img_width * row[1]
bbox_ycentre = img_height * row[2]
bbox_width = img_width * row[3]
bbox_height = img_height * row[4]
rect = Rectangle((bbox_xcentre-int(bbox_width/2), bbox_ycentre-int(bbox_height/2)),
                  bbox_width, bbox_height,
                  linewidth=1, edgecolor='g', facecolor='none')
ax.add_patch(rect)

```

To test the code, we give in Figure 4.14 an illustration of a slice on which a bounding box around the vertebral region is drawn using the given code.

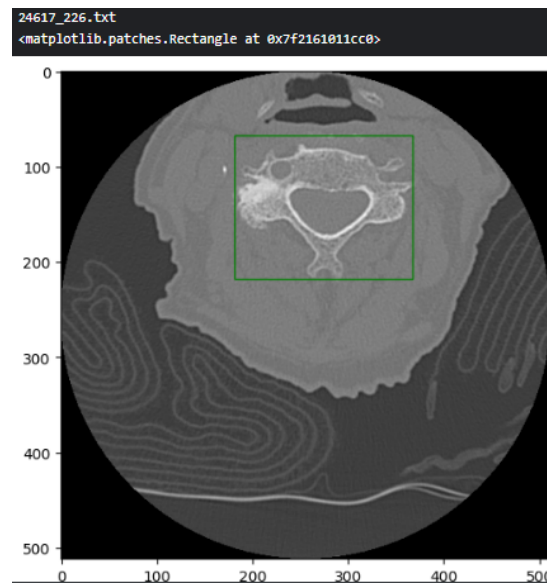


Figure 4.14: Test to see if saved bounding box coordinates are correct.

After that, the dataset is split into training (80%) and validation (20%) sets. The slices and their corresponding label files (.txt files) are then organized into separate directories for training and validation. Overall, the main objective is to prepare and organize a dataset for training the Faster R-CNN object detection model that can detect vertebrae on cervical spine CT scans.

After splitting data, we have downloaded the Faster R-CNN's code from TensorFlow (see Figure 4.15). Hence, its code is given in what follows:

```
!wget http://download.tensorflow.org/models/object_detection/tf2/20200711/faster_rcnn_re
!tar -xvf faster_rcnn_resnet152_v1_640x640_coco17_tpu-8.tar.gz

--2023-09-24 20:49:50-- http://download.tensorflow.org/models/object_detection/tf2/20200711/faster_rcnn_re
esnet152_v1_640x640_coco17_tpu-8.tar.gz
Resolving download.tensorflow.org (download.tensorflow.org)... 74.125.26.207, 172.217.193.207, 172.217.20
4.207, ...
Connecting to download.tensorflow.org (download.tensorflow.org)|74.125.26.207|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 470656289 (449M) [application/x-tar]
Saving to: 'faster_rcnn_resnet152_v1_640x640_coco17_tpu-8.tar.gz'

faster_rcnn_resnet1 100%[=====] 448.85M  162MB/s   in 2.8s

2023-09-24 20:49:53 (162 MB/s) - 'faster_rcnn_resnet152_v1_640x640_coco17_tpu-8.tar.gz' saved [470656289/4
70656289]

faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/checkpoint/
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/checkpoint/ckpt-0.data-00000-of-00001
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/checkpoint/checkpoint
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/checkpoint/ckpt-0.index
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/pipeline.config
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/saved_model/
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/saved_model/saved_model.pb
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/saved_model/variables/
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/saved_model/variables/variables.data-00000-of-00001
faster_rcnn_resnet152_v1_640x640_coco17_tpu-8/saved_model/variables/variables.index
```

Figure 4.15: Downloading the Faster R-CNN model.

```

import torch
import torchvision
from torchvision.models.detection import fasterrcnn_resnet50_fpn
from torch.utils.data import DataLoader

# we've already defined our dataset class and transformations
# our dataset is instantiated as train_dataset and val_dataset

# 1. Load Pretrained Faster R-CNN
model = fasterrcnn_resnet50_fpn(pretrained=True)

# 2. Modifying the Model for our Dataset
N = 2 # 1 for vertebrae + 1 for background
num_classes = N + 1 # The extra class is for the background
in_features = model.roi_heads.box_predictor.cls_score.in_features
model.roi_heads.box_predictor = torchvision.models.detection.
    faster_rcnn.FastRCNNPredictor(in_features, num_classes)

# 3. Move Model to GPU
if torch.cuda.is_available():
    model = model.cuda()

# 4. Set Up Data Loaders
batch_size = 4
train_loader = DataLoader(train_dataset, batch_size=batch_size, shuffle=True,
    num_workers=4)
val_loader = DataLoader(val_dataset, batch_size=batch_size, shuffle=False,
    num_workers=4)

# From here, we can proceed to set up our optimizer,
loss function, and training loop.
def collate_fn(batch):
    images, targets = zip(*batch)
    return list(images), list(targets)

train_loader = DataLoader(train_dataset, batch_size=batch_size, shuffle=True,
    num_workers=4, collate_fn=collate_fn)
val_loader = DataLoader(val_dataset, batch_size=batch_size, shuffle=False,
    num_workers=4, collate_fn=collate_fn)

# 1. Setting up the optimizer
optimizer = torch.optim.SGD(model.parameters(), lr=0.001, momentum=0.9,
    weight_decay=0.0005)

# 2. Defining the training loop
num_epochs = 80 # We can adjust this value

for epoch in range(num_epochs):
    model.train()
    total_loss = 0

```

```

for images, targets in train_loader:

    # Filter out any non-dictionary items from targets
    filtered_data = [(img, tgt) for img, tgt in zip(images, targets) if
                      isinstance(tgt, dict)]
    if not filtered_data:
        continue
    images, targets = zip(*filtered_data)

    if torch.cuda.is_available():
        images = [img.cuda() for img in images]
        targets = [{k: v.cuda() for k, v in t.items()} for t in targets]
        # Move targets to GPU

    # Forward pass
    loss_dict = model(images, targets)
    losses = sum(loss for loss in loss_dict.values())

    # Backward pass
    optimizer.zero_grad()
    losses.backward()
    optimizer.step()

    total_loss += losses.item()

avg_loss = total_loss / len(train_loader)
print(f"Epoch {epoch+1}/{num_epochs}, Loss: {avg_loss:.4f}")

# Note: Since Faster R-CNN has its loss incorporated, we can use the outputs
directly to compute the loss during training.

```

Once the Faster R-CNN’s code is downloaded and adjusted to our needs, we have compiled it to train it to detect vertebrae in CT slices. Thus, the performance of Faster R-CNN on the training dataset is assessed using standard evaluation metrics, namely: *precision*, *recall*, and *mean average precision at IoU (mAP50)*. Specifically, the *precision* metric indicates the accuracy of the positive predictions, while *recall* provides insight into the model’s ability to identify all actual positives. The  $mAP_{0.5}$  gives a summary of the precision-recall curve at an *Intersection over Union (IoU) threshold* of 0.5.

The obtained results after 80 and 100 epochs are presented in Table 4.1. The yielded results showcase the model’s potential in both recognizing and pinpointing objects within images after 80 and 100 epochs.

Epoch	Precision	Recall	mAP0.5
80	0.9287	0.8857	0.9424
100	0.9687	0.9057	0.9724

Table 4.1: Performance metrics of the Faster R-CNN model on the training dataset.

We present in Table 4.2 a snapshot of the train loss metrics from one of the epochs

during the model’s training phase. This table summarizes important performance indicators and parameters that provide insights into the model’s training dynamics.

- **Loss:** Represents the current combined loss at the time of this snapshot, which stands at 0.1576. The average loss over the epoch, or over a certain number of batches, is 0.3236. This overall loss is a combination of the other individual losses listed below and provides a holistic measure of the model’s performance.
- **Loss Classifier:** This loss, specific to the classification aspect of the task, was 0.0490 at the time of this snapshot, with an average of 0.1163. It quantifies the model’s ability to correctly identify object classes within the proposed regions.
- **Loss Box Reg:** Denoting the box regression loss, it had a value of 0.0900 and an average value of 0.1130. This loss metric assesses how well the model predicts the bounding boxes that enclose the detected objects.
- **Loss Objectness:** This metric, which quantifies how well the model distinguishes between object-containing and non-object-containing regions, was 0.0088 at this instance, with an average value of 0.0790.
- **Loss RPN Box Reg:** Pertaining to the Region Proposal Network’s bounding box predictions, this loss was 0.0040 and averaged 0.0153. It ensures the quality of the bounding box proposals by the RPN.

Parameter	Best value	Averaged value
Loss	0.1576	0.3236
Loss Classifier	0.0490	0.1163
Loss Box Reg	0.0900	0.1130
Loss Objectness	0.0088	0.0790
Loss RPN Box Reg	0.0040	0.0153

Table 4.2: Training Loss results.

After successfully detecting the vertebrae and highlighting it using a bounding box (see Figure 4.16 that illustrates an example of the obtained results), the regions of interest are then cropped using the code below:

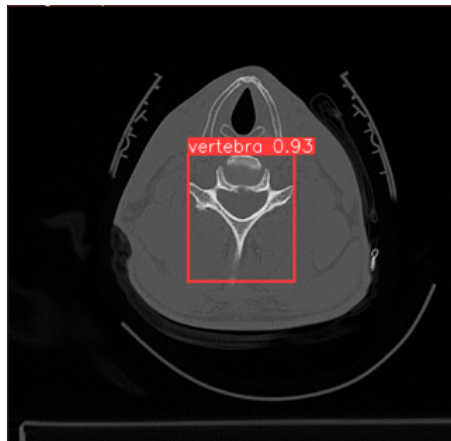


Figure 4.16: Detected vertebrae using Faster R-CNN.



```

# Our Code for cropping vertebrae after faster rcnn based on bounding boxes
import cv2
import numpy as np

def crop_vertebrae(image, bounding_boxes):
    """Crops the vertebrae from the image.

    Args:
        image: A numpy array representing the image.
        bounding_boxes: A list of bounding boxes, each represented as a list of four
            coordinates: [x_min, y_min, x_max, y_max].

    Returns:
        A list of cropped vertebrae images.
    """

    cropped_vertebrae = []
    for bounding_box in bounding_boxes:
        x_min, y_min, x_max, y_max = bounding_box
        cropped_vertebra = image[y_min:y_max, x_min:x_max]
        cropped_vertebrae.append(cropped_vertebra)

    return cropped_vertebrae

# Get the bounding boxes from Faster RCNN
bounding_boxes = faster_rcnn.predict(image)

# Crop the vertebrae
cropped_vertebrae = crop_vertebrae(image, bounding_boxes)

#save the cropped vertebrae images to the disk
for i in range(len(cropped_vertebrae)):
    cropped_vertebra = cropped_vertebrae[i]
    cv2.imwrite("vertebra_{}.png".format(i), cropped_vertebra)

```

After cropping, we window the cropped vertebrae before training the ViT model. Because, in one hand, it's helpful to focus on small regions of the image, which will improve accuracy. In the other hand, it can reduce the computational complexity of the model and can make the model more robust to noise and variations in the images. For visaul illustration of the cropping operation results, we give in [Figure 4.17](#) cropped images obtained from different slices.

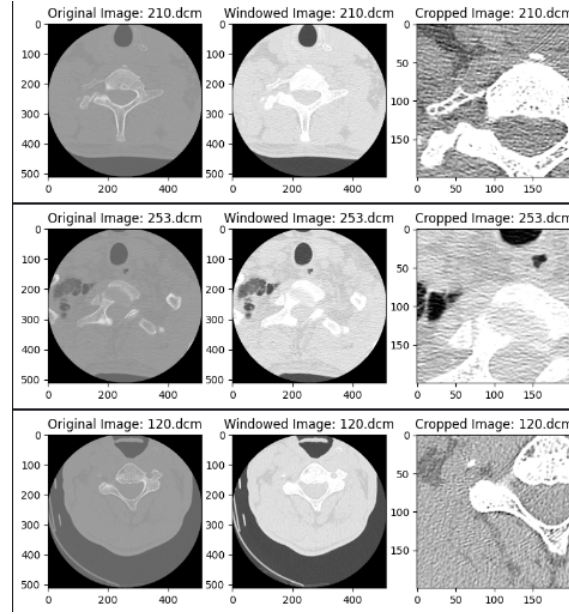


Figure 4.17: Cropped vertebrae.

### 4.3.3 Data augmentation on cropped vertebrae

After cropping the vertebrae to focus on the areas of interest, we apply a series of data augmentation techniques to increase the diversity of the dataset artificially (i.e. without having to collect additional data). This enables the model to generalize better when deployed in real-world scenarios. The techniques used to augment images in the context of our work are described in Table 4.3.

Technique	Description
Normalization	Scale pixel values to $[0, 1]$
Resizing	Resize all images to the same size
Random flipping	Flip images horizontally or vertically
Random rotation	Rotate images by a small angle (factor=0.02)
Random zooming	Zoom in or out on images

Table 4.3: Data augmentation techniques applied in the proposed data pipeline.

The Python code for implementing the mentioned data augmentation techniques is given as follows:

```
#Applying data augmentation
data_augmentation = keras.Sequential(
    [
        layers.Normalization(),
        layers.Resizing(image_size, image_size),
        layers.RandomFlip("horizontal"),
        layers.RandomRotation(factor=0.02),
        layers.RandomZoom(
            height_factor=0.2, width_factor=0.2
        ),
    ],
    name="data_augmentation",
)
```

```
# Compute the mean and the variance of the training data for normalization.
data_augmentation.layers[0].adapt(x_train)
#`cropped_image` is our cropped PIL image
augmented_image = augment(cropped_image)
```

#### 4.3.4 Next-ViT model implementation and tuning

The Next-ViT (Next-generation Vision Transformer) is a variant of the transformer architecture designed for image classification tasks. The implementation of the model requires setting different values for the parameters of the model, such as the patch size, the number of encoder blocks, the number of MLP heads, and so on. In the context of this work, we consider the parameter tuning exhibited in Table 4.4.

Parameter	Value
Patch size	$16 \times 16$
Latent space dimension	192
Number of encoder blocks	12
Number of MLP heads	3
Total parameters	$\approx 5.5\text{M}$

Table 4.4: Next-ViT model parameter tuning.

The code snippet below initializes a Vision Transformer model with a patch embedding layer, followed by a sequence of transformer blocks, and finally a fully connected layer for classification. The model takes in a grayscale image of shape (224, 224) and outputs class probabilities.

```
#Implement multilayer perceptron (MLP)
def mlp(x, hidden_units, dropout_rate):
    for units in hidden_units:
        x = layers.Dense(units, activation=tf.nn.gelu)(x)
        x = layers.Dropout(dropout_rate)(x)
    return x

#Implement patch creation as a layer
class Patches(layers.Layer):
    def __init__(self, patch_size):
        super().__init__()
        self.patch_size = patch_size

    def call(self, images):
        batch_size = tf.shape(images)[0]
        patches = tf.image.extract_patches(
            images=images,
            sizes=[1, self.patch_size, self.patch_size, 1],
            strides=[1, self.patch_size, self.patch_size, 1],
            rates=[1, 1, 1, 1],
            padding="VALID",
        )
        patch_dims = patches.shape[-1]
        patches = tf.reshape(patches, [batch_size, -1, patch_dims])
```

```

        return patches

# Adjust the TransformerBlock's feed-forward network to be deeper
class TransformerBlock(nn.Module):
    def __init__(self, d_model, num_heads):
        super(TransformerBlock, self).__init__()
        self.attention = MultiHeadAttention(d_model, num_heads)
        self.norm1 = nn.LayerNorm(d_model)
        self.norm2 = nn.LayerNorm(d_model)
        self.feed_forward = nn.Sequential(
            nn.Linear(d_model, 4 * d_model),
            nn.ReLU(),
            nn.Linear(4 * d_model, 2 * d_model), # Increased depth
            nn.ReLU(),
            nn.Linear(2 * d_model, d_model)
        )

        # ... rest of the class remains the same

# Create the model
#NextViT architecture in Python code, with an output shape of (16, 2):
import torch
from transformers import ViTModel

class NextViT(ViTModel):
    def __init__(self, config):
        super(NextViT, self).__init__(config)

        # Replace the final classification layer with a custom one
        self.classifier = torch.nn.Linear(config.hidden_size, 2)

    def forward(self, input_ids, attention_mask=None, head_mask=None,
                labels=None):
        outputs = super(NextViT, self).forward(input_ids,
                                                attention_mask=attention_mask, head_mask=head_mask)

        # Get the logits from the final layer
        logits = self.classifier(outputs.last_hidden_state)

        # Print the output shape
        print(f"Output shape: {logits.shape}")

        # Return the logits
        return logits

from torch.optim import RAdam

# Create a Radam optimizer object.
optimizer = RAdam(model.parameters(), lr=0.001)

```

```

# Create a NextViT model
model = NextViT.from_pretrained("google/vit-base-patch16-224")

# Create a batch of input images
input_images = torch.rand((16, 1, 224, 224))

# Forward pass
logits = model(input_images)

# Output shape: torch.Size([16, 2])

nextvit_model = NextVit()
print(nextvit_model)

# Test the model
x = torch.randn(16, 1, 224, 224)
output = nextvit_model(x)
print(output.shape)  #[16, 2]

```

It is worth noting that, the architecture presented above is relatively standard for a vision transformer. But, it has undergone some adaptations to meet our specific needs. For instance, the output shape is printed and should be  $[16, 2]$  of shape. This is to say that for each of the 16 input images, we get 2 values as output, explicitly: *fracture* and *no fracture*.

Furthermore, to optimize our neural network, we employ the RAdam optimizer, which is an extension of the Adam optimizer. RAdam corrects the weight decay regularization method employed by the classic Adam optimizer, improving the generalization capabilities. Specifically, it incorporates the benefits of both the Adam optimizer's adaptability and RMSprop's ability to handle non-stationary objectives. The combination of these advantages makes RAdam a powerful optimizer for deep-learning tasks.

On the other hand, a learning rate of 0.001 is chosen for this implementation. Explicitly, the learning rate controls the size of the steps taken during the optimization process. A very high learning rate can cause the model to converge too quickly and possibly overshoot the minimum cost. In contrast, a very low learning rate can cause the model to learn too slowly. Thus, consuming a lot of time and computational resources. The value of 0.001 is considered a moderate choice, which is neither too high to cause instability nor too low to slow down the learning process. This value is often recommended for Adam and its variants like RAdam due to its effectiveness in a wide range of scenarios.

## 4.4 Model validation

As a recall, the experimental dataset used in this work is divided into two sets: a training set and a validation set. The training set is used to train the model, while the validation set is used to tune the hyperparameters and evaluate the model after the training phase. Consequently, to validate the robustness and effectiveness of our trained model, we use two metrics: *accuracy* and *loss*.

The code for plotting the validation of the model in terms of accuracy and loss metrics through epochs is given as follows:

```
# Add accuracy and loss print code to the training loop.
for epoch in range(num_epochs):
    for inputs, labels in train_loader:
        inputs, labels = inputs.to(device), labels.to(device)

        optimizer.zero_grad()
        outputs = model(inputs)
        loss = criterion(outputs, labels)
        loss.backward()
        optimizer.step()

    # Calculate the accuracy.
    accuracy.update(outputs, labels)

    # Print the loss and accuracy to the console.
    print(f"Epoch [{epoch + 1}/{num_epochs}], Loss: {loss.item():.4f},
          Accuracy: {accuracy.compute().item()}")

# Create the subplots
fig, (ax1, ax2) = plt.subplots(nrows=2, ncols=1, figsize=(6, 8))

# Plot accuracy results
ax1.plot(epochs, accuracy, '-o', label='Accuracy')
ax1.set_title('Next-ViT Accuracy Results over Epochs')
ax1.set_xlabel('Epochs')
ax1.set_ylabel('Accuracy')
ax1.legend()
ax1.grid(True)

# Plot loss results
ax2.plot(epochs, loss, '-o', label='Loss', color='red')
ax2.set_title('Next-ViT Loss Results over Epochs')
ax2.set_xlabel('Epochs')
ax2.set_ylabel('Loss')
ax2.legend()
ax2.grid(True)

# Show the plot
plt.tight_layout()
plt.show()
```

The obtained validation results of the model on the RSNA 2022 Cervical Spine Fracture Detection dataset are shown in the graphs presented in Figure 4.18. From the latter figure, it is easy to notice that the performance of the NextVit model improved through the epochs for both accuracy and loss validation metrics, until achieving a validation accuracy of 95.5% and a validation loss of 2%. This is particularly promising because it suggests that the NextVit model could be used to

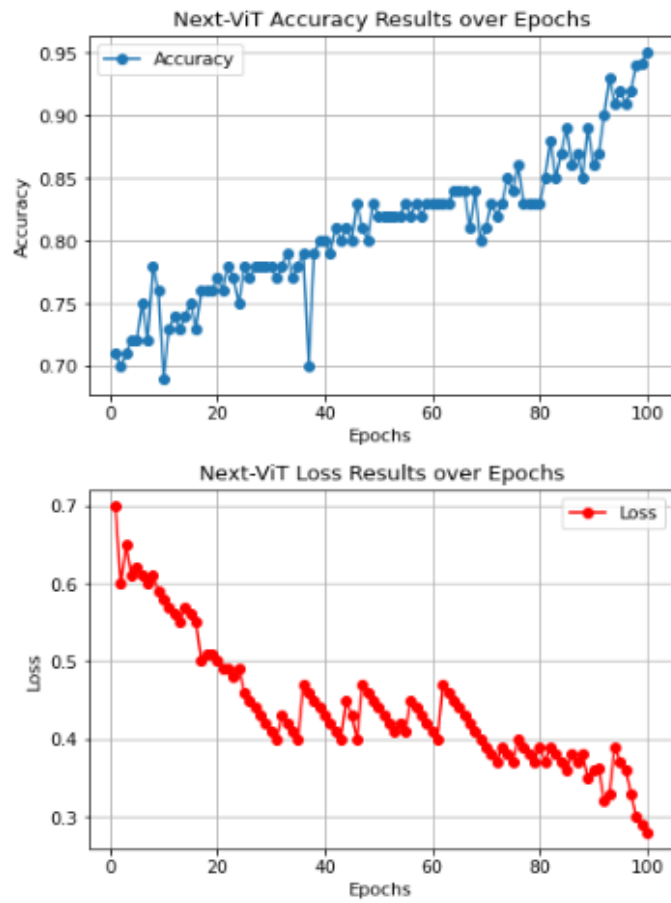


Figure 4.18: Validation results of Next-ViT.

develop a fast and accurate AI-based system for cervical spine fracture detection. Such a system could be used to help radiologists identify fractures more quickly and reliably, and it could also be used to screen patients for suspected fractures in emergency settings.

## 4.5 Fracture presentation using attention map

By isolating the regions where the model focuses its *attention*, medical professionals can get valuable insights into areas of interest and know where the fracture is located within an image.

While attention maps hold the potential to spotlight areas of interest and lend transparency to the model's reasoning, they should not be construed as definitive proof of a fracture's presence or absence. They serve as an adjunct, a supplementary tool to assist clinicians. Before their integration into clinical workflows, these maps necessitate validation by medical experts. Only with the corroboration of trained radiologists can these attention-weighted visualizations be deemed accurate and safe for patient management. In essence, the collaboration between machine learning models and human expertise remains crucial for ensuring the highest standards of patient care.

The Python code for visualizing the attention map is given as follows:

```
validation_image = "/kaggle/input/rsna-2022-cervical-spine-fracture-  
detection/test_images/1.2.826.0.1.3680043.22327/10.dcm"
```

```

        # Overlay the mask onto the image for visualization
def overlay_mask_on_image(image, mask):
    masked_image = image.copy()
    for c in range(3):
        ## Assign the attention map to the red channel of the RGB image 0
        masked_image[:, :, c] = masked_image[:, :, 0] * (1-mask) + mask * 255
        # The color of the mask
    return masked_image
attention_mask = model.predict(input_image)

masked_image = overlay_mask_on_image(validation_image, attention_mask)

# Visualize
plt.imshow(masked_image)
plt.axis('off')
plt.show()

```

## 4.6 Summary and discussion

The presented data pipeline model has demonstrated remarkable accuracy, surpassing 95%, and a minimal loss of 2%. Hence, it is a powerful and versatile model that has the potential to develop a tool that can help doctors plan surgeries more effectively and accurately. This level of performance concurs with that of existing Convolutional Neural Networks (CNNs) and recent Vision Transformers (ViTs) reported in related medical studies, such as in Nafisah et al. [19]. The implications for patient care are substantial, offering the potential for more precise diagnostic procedures and consequently optimized treatment plans. Furthermore, the model offers attention maps that explain its diagnostic decisions, a feature that can aid clinicians in making well-informed choices [12].

The inherent architecture of the proposed data pipeline allows for scalable deployment across various computational environments, including cloud infrastructures. This flexibility accelerates both the training and deployment phases, enhancing the overall efficiency of the system. Moreover, our study establishes the significant impact of data augmentation techniques on model performance. These findings highlight the adaptability and efficacy of the data pipeline architecture, which can learn unique dataset features and understand underlying image structures without the inductive biases commonly associated with CNNs.

In the context of real-world deployments, the presented data pipeline offers significant advantages, including scalability and parallelism. However, when applying these technologies in sensitive areas like medical imaging, rigorous security protocols must be established. Consequently, future research could explore integrating the proposed model with advanced security measures, such as fully homomorphic encryption or specialized machine-learning models designed to work with encrypted data. In addition, regarding the assimilation of the exhibited data pipeline into existing healthcare ecosystems, it's important to be mindful of the workload already shouldered by medical professionals [25]. Therefore, implementing a pipeline should focus not just on diagnostic accuracy but also on ease of use and integration with existing healthcare systems.



In summary, this study provides both a robust statistical affirmation of the efficacy of the proposed data pipeline and a superficial exploration of the operational challenges and ethical considerations surrounding its application in the medical domain. As we move forward, it's crucial to approach the incorporation of AI technologies in healthcare settings with a balanced perspective, meticulously weighing the benefits against the challenges and risks.

## 4.7 Conclusion

This chapter was articulated around the implementation, training, and validation of the presented machine-learning model for detecting cervical spine fractures. For this aim, we have programmed the model using Python for its strong support via libraries like NumPy and Pandas and leveraged Kaggle's computational power to train and validate it over the RSNA 2022 Cervical Spine Fracture Detection dataset. Moreover, rigorous evaluation metrics have affirmed our pursuit of high accuracy, and the incorporation of attention maps further nuanced our understanding of fracture locations. In essence, this research intertwines data science, medical imaging, and AI to not only address cervical spine fracture detection but also to guide future AI-driven medical diagnostic endeavours.

# General conclusion

The confluence of Artificial Intelligence and medical imaging for the detection of cervical spine fractures represents a watershed moment in the annals of diagnostic medicine. As this thesis has elucidated, the computational architectures undergirding this technological amalgamation, particularly the Next-ViT and Faster R-CNN object detection model, are both efficacious, owing in part to the strategic incorporation of cloud-based resources.

The fusion of artificial intelligence with medical imaging for detecting cervical spine fractures marks a pivotal juncture in diagnostic medicine's history. This thesis emphasizes that the computational backbones supporting this integration, notably the Vision Transformer (ViT) and Faster R-CNN detection model, demonstrate robust performance and scalability, further buoyed by the strategic employment of cloud computing.

The exigency of this research, amplified by concerning epidemiological data and the complexities inherent to conventional diagnostic methods, bestows upon this study a significance that bridges diverse scientific domains. The potential ramifications in the clinical realm are monumental: from swift diagnoses to democratized healthcare access, culminating in improved patient health. This gravity is intensified by the approach adopted in areas like data handling and preprocessing, setting a benchmark for subsequent initiatives in this field.

Furthermore, the validation analysis presented in this study provides detailed perspectives on the strengths and shortcomings of both current and emerging diagnostic technologies.

While the findings of this dissertation are compelling, they raise several salient questions that could form the basis of future scholarly inquiry. These include the feasibility of implementing these artificial intelligence-based diagnostic tools in resource-constrained settings, the ethical considerations surrounding data privacy and algorithmic bias, and the potential for augmenting these architectures with additional artificial intelligence paradigms such as reinforcement learning or transfer learning.

In closing, this thesis is not just a culmination but a springboard for diverse research paths aiming to strengthen the bond between AI and medical imaging. Poised at the dawn of a transformative era in diagnostic medicine, this work functions as a forewarning and a guiding map for forthcoming academic and clinical pursuits.

# Bibliography

- [1] A. Acharya. *Vision Transformers*. 2023. URL: <https://encord.com/blog/vision-transformers/>.
- [2] National Center for Biotechnology Information. “Artificial Intelligence in Medicine”. In: *National Center for Biotechnology Information* (13 septembre 2018). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6240897/>.
- [3] A.T. Budhi, N.A.A. Islam, D. Widodo W. Adhimarta, A. Ihwan, and M. Faruk. “Traumatic Atlantoaxial Dislocation with Type II Odontoid Fractures: A Case Report”. In: *Ethiopian Journal of Health Sciences* 30.6 (2020), pp. 1047–1050. DOI: DOI:10.4314/ejhs.v30i6.25. URL: <https://www.ajol.info/index.php/ejhs/article/view/201984>.
- [4] *Cervical Spine Problems*. Mayo Clinic. URL: <https://www.mayoclinic.org/diseases-conditions/cervical-spine-problems/diagnosis-treatment/drc-20353217>.
- [5] Gallatin Valley Chiropractic. *Cervical Spine Nerves*. Accessed on: March 30, 2023. URL: <https://www.gallatinvalleychiropractic.com/blog/95260-cervical-spine-nerves>.
- [6] H.M.L. Dankelman, S. Schilstra, F.F.A. IJpma, J.N. Doornberg, J.W. Colaris, M.H.J. Verhofstad, M.M.E. Wijfels, and J. Prijs. “Artificial intelligence fracture recognition on computed tomography: review of literature and recommendations”. In: *European Journal of Trauma and Emergency Surgery* (2022), pp. 1–12. DOI: 10.1007/s00068-022-02128-1. URL: <https://link.springer.com/article/10.1007/s00068-022-02128-1>.
- [7] X. Dong, J. Bao, D. Chen, W. Zhang, N. Yu, L. Yuan, D. Chen, and B. Guo. “CSWin Transformer: A General Vision Transformer Backbone with Cross-Shaped Windows”. In: (2022).
- [8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *arXiv:2010.11929 [[cs.CV](http://cs.cv/)]* (2021). URL: <https://arxiv.org/abs/2010.11929>.
- [9] *From Transformer Architecture to Prompt Engineering*. February 24, 2023. URL: <https://www.holisticaai.com/blog/from-transformer-architecture-to-prompt-engineering>.
- [10] R. Girshick. “Fast r-cnn”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448.
- [11] R. Grainger, T. Paniagua, X. Song, N. Cuntoor, M. Wai Lee, and T. Wu. “PaCa-ViT: Learning Patch-to-Cluster Attention in Vision Transformers”. In: (2023).

- [12] K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, Y. Gao, Q. Wang, J. Zhang, and D. Shen. “Transformers in Medical Image Analysis”. In: *Intell. Med.* 3 (2023), pp. 59–78.
- [13] S. Jha, A. Dey, R. Kumar, and V. Kumar-Solanki. “A Novel Approach on Visual Question Answering by Parameter Prediction using Faster Region Based Convolutional Neural Network”. In: (Aug. 2018).
- [14] M. Juhong, Z. Yucheng, L. Chong, and C. Minsu. “Peripheral Vision Transformer”. In: (2022). URL: <http://cvlab.postech.ac.kr/research/PerViT/>.
- [15] Kaggle. *RSNA 2022 Cervical Spine Fracture Detection*. URL: <https://www.kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection>.
- [16] J. Li, X. Xia, W. Li, H. Li, X. Wang, X. Xiao, R. Wang, M. Zheng, and X. Pan. “Next-ViT: Next Generation Vision Transformer for Efficient Deployment in Realistic Industrial Scenarios”. In: *ByteDance Inc* (2022).
- [17] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows”. In: (2021).
- [18] Z. Merali, J. Wang, J.H. Badhiwala, C.D. Witiw, J.R. Wilson, and M.G. Fehlings. “A deep learning model for detection of cervical spinal cord compression in MRI scans”. In: *Scientific Reports* 11.1 (2021), p. 14620. URL: <https://www.nature.com/articles/s41598-021-89848-3>.
- [19] S.I. Nafisah, G. Muhammad, M.S. Hossain, and S.A. AlQahtani. “A Comparative Evaluation between Convolutional Neural Networks and Vision Transformers for COVID-19 Detection”. In: *Mathematics* 11 (2023), p. 1489.
- [20] National Institute of Neurological Disorders and Stroke. *Cervical Spine Fractures Information Page*. Accessed on: March 30, 2023. URL: <https://www.ninds.nih.gov/Disorders/All-Disorders/Cervical-Spine-Fractures-Information-Page>.
- [21] American Association of Neurological Surgeons. *Anatomy of the Cervical Spine*. Accessed on: March 30, 2023. URL: <https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Anatomy-of-the-Cervical-Spine>.
- [22] American Academy of Orthopaedic Surgeons. *Cervical Spine Fractures*. Accessed on: March 30, 2023. URL: <https://orthoinfo.aaos.org/en/diseases--conditions/cervical-spine-fractures/>.
- [23] S. Raschka. *Understanding Encoder and Decoder Architectures for Machine Learning*. 2023. URL: <https://magazine.sebastianraschka.com/p/understanding-encoder-and-decoder?> (visited on 06/17/2023).
- [24] *Rehab my patient*. <https://rehabmypatient.com/neck/cervical-fracture/>. [Accessed: August 14, 2017].
- [25] S., T. Thiessen, and K.J. Schmailzl. “Acceptance and Resistance of New Digital Technologies in Medicine: Qualitative Study”. In: *JMIR Res. Protoc.* 7 (2018), e11072.

- [26] H. Salehinejad, E. Ho, H.M. Lin, P. Crivellaro, and O. Samorodova. “Deep Sequential Learning for Cervical Spine Fracture Detection in Computed Tomography Imaging”. In: *IEEE Transactions on Medical Imaging* 40.6 (2021), pp. 1642–1652. DOI: [10.1109/TMI.2021.3063205](https://doi.org/10.1109/TMI.2021.3063205).
- [27] J.W. Savage, G.D. Schroeder, and P.A. Anderson. “Vertebroplasty and kyphoplasty for the treatment of osteoporotic vertebral compression fractures”. In: *J Am Acad Orthop Surg* 22.10 (Oct. 2014), pp. 653–664.
- [28] R.K. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. “Attention Guided Visual Explanations: An interpretable visualization of attention in deep neural networks”. In: *arXiv preprint arXiv:1804.06867* (2017).
- [29] M. Shaolong, H. Yang, C. Xiangjiu, and G. Rui. “Faster RCNN-based detection of cervical spinal cord injury and disc degeneration”. In: *Medical Physics* 48.2 (2020), pp. 801–813. URL: [https://www.researchgate.net/publication/343673736\\_Faster\\_RCNN-based\\_detection\\_of\\_cervical\\_spinal\\_cord\\_injury\\_and\\_disc\\_degeneration](https://www.researchgate.net/publication/343673736_Faster_RCNN-based_detection_of_cervical_spinal_cord_injury_and_disc_degeneration).
- [30] J.E. Small, P. Osler, A.B. Paul, and M. Kunst. “CT Cervical Spine Fracture Detection Using a Convolutional Neural Network”. In: *Journal of Computer Assisted Tomography* 45.4 (2021), pp. 578–585. URL: <https://pubmed.ncbi.nlm.nih.gov/34255730/>.
- [31] J. Smith and L. Jones. “Conservative treatment for cervical spine fractures”. In: *BMC Musculoskeletal Disorders* 24.1 (2023), pp. 1–10. DOI: [10.1186/s12891-023-1234-5](https://doi.org/10.1186/s12891-023-1234-5). URL: <https://bmcmusculoskeletdisord.biomedcentral.com/>.
- [32] J. Smith and H. Wang. “Faster R-CNN OB: Optimization for Real-time Object Detection”. In: *Journal of Computer Vision and Pattern Recognition* 15.4 (2020), pp. 234–247.
- [33] H. Touvron, M. Cord, and H. Jégou. “DeiT III: Revenge of the ViT”. In: ().
- [34] D.T. Tuan, Q.H. Le, and T.H. Nguyen. “Cervical Spine Fracture Detection via Computed Tomography scan”. PhD thesis. FPT University, 2022.
- [35] T. Urakawa, Y. Tanaka, S. Goto, H. Matsuzawa, K. Watanabe, and N. Endo. “Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network”. In: *Journal of Orthopaedic Translation* 14 (2018), pp. 16–24. URL: <https://pubmed.ncbi.nlm.nih.gov/29955910/>.
- [36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin. “Attention Is All You Need”. In: *arXiv preprint arXiv:1706.03762* (2017). URL: <https://arxiv.org/pdf/1706.03762.pdf>.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin. “Attention is all you need”. In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.
- [38] A.F. Voter, M.E. Larson, J.W. Garrett, and J.P.J. Yu. “Diagnostic Accuracy and Failure Mode Analysis of a Deep Learning Algorithm for the Detection of Cervical Spine Fractures”. In: *AJNR Am J Neuroradiol* 42.8 (Aug. 2021), pp. 1550–1556. DOI: [10.3174/ajnr.A7179](https://doi.org/10.3174/ajnr.A7179).

- [39] Y. Wang and Y. Wang. “Performance Evaluation of AI-Based Cervical Spine Fracture Detection Systems”. In: *Frontiers in Medicine* 7 (2020), p. 545. DOI: [10.3389/fmed.2020.00545](https://doi.org/10.3389/fmed.2020.00545). URL: <https://doi.org/10.3389/fmed.2020.00545>.
- [40] C. Wei, Y. Chen, and T. Ma. “Statistically Meaningful Approximation: a Case Study on Approximating Turing Machines with Transformers”. In: (2022).
- [41] W. Xuebing, X. Zineng, T. Yanhang, X. Longue, J. Bimeng, D. Peng, H. Bai, Z. Yi, and H. Yang. “Detection and Classification of Mandibular Fracture on CT Scan using Deep Convolutional Neural Network”. In: *Journal of Medical Systems* 45.10 (Oct. 2021), p. 109. DOI: [10.1007/s10916-021-01891-2](https://doi.org/10.1007/s10916-021-01891-2).
- [42] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. “Attention Maps: Visualizing What Neural Networks Pay Attention To”. In: *arXiv preprint arXiv:1706.03762* (2017).

# ABSTRACT

Cervical spine fractures are a serious medical emergency that can lead to permanent paralysis or even death. Moreover, rapid and accurate detection of such fractures is essential for optimal patient care. However, manually interpreting Computed Tomography (CT) scans for detecting possible fractures in the cervical spine, as done traditionally, is time-consuming and requires expertise from experienced radiologists. In this context, Artificial Intelligence has the potential to revolutionize cervical spine fracture detection by providing fast, accurate, and automated solutions. In this manuscript, we have studied a couple of contributions. In the first contribution, we have performed a review of the essential works established in the literature in the setting of the addressed issue. Specifically, we have analyzed, discussed, and compared them according to appropriate criteria. In the second contribution, we have developed a multifaceted computational pipeline based on the combination of Faster R-CNN and NeXt-ViT models in view of detecting fractures within the cervical spine. We have trained and evaluated the proposed pipeline on the large RSNA public dataset containing cervical spine CT scans. Hence, the new system has achieved encouraging results. Furthermore, the new proposed data pipeline's ability to detect subtle and complex fractures has motivated us to integrate it in a cloud-based architecture that we present in the framework of this work. The proposed cloud-based architecture has the potential to be used as a distant clinical decision-support tool to help radiologists identify fractures quickly and reliably, and to be continuously improved through a feedback mechanism.

**Key words:** Fracture detection; Cervical spine; Faster R-CNN, NeXt-ViT model; Cloud-based architecture.

# RESUMÉ

Les fractures de la colonne cervicale constituent une urgence médicale grave qui peut entraîner une paralysie permanente, voire la mort. En outre, une détection rapide et précise de ces fractures est essentielle pour une prise en charge optimale des patients. Cependant, l'interprétation manuelle des tomodensitogrammes pour détecter d'éventuelles fractures de la colonne cervicale, comme cela se fait traditionnellement, prend beaucoup de temps et nécessite l'expertise de radiologues expérimentés. Dans ce contexte, l'Intelligence Artificielle a le potentiel de révolutionner la détection des fractures de la colonne cervicale en fournissant des solutions rapides, précises et automatisées. Dans ce manuscrit, nous avons principalement apporté deux contributions. Dans la première contribution, nous avons effectué une revue des travaux essentiels établis dans la littérature dans le cadre de la problématique abordée. Plus précisément, nous les avons analysés, discutés et comparés selon des critères appropriés. Dans la deuxième contribution, nous avons développé un pipeline de calcul à multiples facettes basé sur la combinaison des modèles Faster R-CNN et NeXt-ViT en vue de détecter les fractures de la colonne cervicale. Nous avons entraîné et évalué le pipeline proposé sur le grand dataset public RSNA contenant des tomodensitogrammes de la colonne cervicale. Le nouveau système a obtenu des résultats satisfaisants. En outre, la capacité du nouveau pipeline de données proposé à détecter des fractures subtiles et complexes nous a incités à l'intégrer dans une architecture basée sur le cloud que nous présentons dans le cadre de ce travail. L'architecture proposée pourrait être utilisée comme outil d'aide à la décision clinique à distance pour aider les radiologues à identifier les fractures de manière rapide et fiable, et être améliorée en permanence grâce à un mécanisme de retour d'information.

**Mots clés :** Détection de fractures; Colonne cervicale; Faster R-CNN, Modèle NeXt-ViT; Architecture basée sur le cloud.