T.C.
ISTANBUL GEDİK UNIVERSITY
INSTITUTE OF GRADUATE STUDIES

AUTOMATED DETECTION AND CLASSIFICATION OF SPINAL DISC HERNIATION USING DEEP LEARNING ON MRI IMAGES

MASTER THESIS

Mustafa Isam AL-AJAJ

Department of Statistics and Data Science

Statistics and Data Science English Program

MAY 2024
ISTANBUL

T.C.
ISTANBUL GEDİK UNIVERSITY
INSTITUTE OF GRADUATE STUDIES

AUTOMATED DETECTION AND CLASSIFICATION OF SPINAL DISC HERNIATION USING DEEP LEARNING ON MRI IMAGES

MASTER THESIS

Mustafa Isam AL-AJAJ
(211225001)

Department of Statistics and Data Science

Statistics and Data Science English Program

Thesis Advisor: Assist. Prof. Dr. Rıza DİLEK

Istanbul 2024

T.C.

İSTANBUL GEDİK ÜNİVERSİTESİ

Lisansüstü Eğitim Enstitüsü Müdürlüğü

Jüri Tez Onay Formu

21.05.2024

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜ

Bu çalışma 21./05./2024 tarihinde aşağıdaki jüri tarafından İstatistik ve Veri Bilimi Anabilim Dalı, İstatistik ve Veri Bilimi İngilizce (Tezli Yüksek Lisans) Programı Yüksek Lisans Tezi olarak kabul edilmiştir.

TEZ JÜRİSİ

Dr. Öğr. Üyesi Rıza DİLEK
Danışman
İstanbul Gedik Üniversitesi

Doç. Dr. İzzet Paruğ DURU
Üye (İmza)
İstanbul Gedik Üniversitesi

Doç. Dr. Caner DEĞER
Üye (İmza)
Marmara Üniversitesi

DECLARATION

I'm Mustafa Isam AL-AJAJ, declare that this thesis titled “Automated Detection and Classification of Spinal Disc Herniation Using Deep Learning on MRI Images” is original work I completed this to receive my master's in Statistics and Data Science. I further declare that neither this thesis nor any part of it has ever been submitted to or presented for a research paper or other degree at any other university or institution. (21 /05/2024)

Mustafa Isam AL-AJAJ

TABLE OF CONTENTS

Page No:

ABBREVIATIONS

3D	: Three Dimensional
AAP	: Area between Anterior and Posterior
AI	: Artificial Intelligence
ANN	: Artificial Neural Network
AUC-ROC	: Area Under the Curve-Receiver Operating Characteristic
CLBP	: Chronic Lower Back Pain
CNNs	: Convolutional Neural Networks
CSF	: Cerebrospinal Fluid
CT	: Computerized Tomography
DCE-MRI	: Dynamic Contrast Enhanced-Magnetic Resonance Imaging
DenseNet	: Densely Connected Convolutional Networks
DSC	: Dice Similarity Coefficient
DTI	: Diffusion Tensor Imaging
DWI	: Diffusion-Weighted Imaging
FCN	: Fully Convolutional Network
FMRI	: Functional MRI
GAN	: Generative Adversarial Network
IoU	: Intersection over Union
IVDs	: Intervertebral Discs
LF	: Ligamentum Flavum
LSS	: Lumbar Spinal Stenosis
MRI	: Magnetic Resonance Imaging
PE	: Posterior Element
ReLU	: Rectified Linear Unit
ResNet	: Residual Network
RNNs	: Recurrent Neural Networks
ROC	: Receiver Operating Characteristic
RoI	: Region of Interest

SVM : Support Vector Machines

TS : Thecal Sac

VGG : Visual Geometry Group

LIST OF FIGURES

Page No:

AUTOMATED DETECTION AND CLASSIFICATION OF SPINAL DISC HERNIATION USING DEEP LEARNING ON MRI IMAGES

ABSTRACT

A common pathological condition that affects the intervertebral discs is spinal disc herniation. Nerve compression and pain result when the gel-like substance in a disc protrudes through the stiff outer layer. For ideal treatment arranging and patient consideration, an exact and provoke finding of spinal disc herniation should be made. Spinal disc herniation necessitates prompt and precise diagnosis before the appropriate treatment can be planned. Magnetic Resonance Imaging (MRI) has developed as an efficient imaging technique in the diagnosis and evaluation of spinal disc herniation, offering precise anatomical information and allowing for non-invasive examination. While MRI is an efficient imaging method for identifying spinal disc herniation, radiologists' manual review procedure is time-consuming and error-prone.

Development of an automated solution include the spine's complex anatomical structure, imaging technique variability, and the lack of annotated datasets raise significant challenges. In the last few years, deep learning algorithms have transformed the interpretation of medical imaging, demonstrating remarkable ability in automated tasks like as detection and classification. So, a computer system that uses deep learning techniques to identify and classify spinal disc herniation on MRI images is required. These strategies are superior to traditional computer-aided diagnostic approaches because they can instantly acquire specific patterns and features from large data sets.

However, there are many difficulties in creating a deep learning system that is both reliable and effective for distinguishing intervertebral disc herniation. The complex anatomy of the spine, the unpredictable nature of imaging strategies, and powerful algorithms that can handle the occurrence of abnormalities on MRI images. Also, one more obstruction to effectively training deep learning models is the absence of large, annotated datasets.

Hence, it is necessary to discover how to develop a deep learning system that overcomes those challenges and uses MRI scans to accurately and automatically identify and classify spinal disc herniations.

This research developed a deep learning system specifically designed for the automated detection and categorization of spinal disc herniation using MRI data, for identifying and categorizing spinal disc herniation from MRI data by utilizing deep learning capabilities. The research compared the efficiency of the conventional approaches to computer-aided diagnosis to the deep learning-based method. In addition to surveying the robotized framework's clinical helpfulness by assessing its ability to help radiologists in deciding the right course of treatment and the board of spinal disc herniation.

By applying such deep learning system into a therapeutically useful tool, radiologists can use it to correctly diagnose and classify spinal disc herniation from

MRI data. The viable reception of such a system can possibly work on indicative interaction, treatment arranging, and patient results. By demonstrating the efficacy of deep learning techniques in improving the diagnosis and treatment of spinal herniation of the disc, the research contributed to the advancement of computer-aided medical image analysis.

This study can possibly change the area of spinal disc herniation discovery and add to the development of PC supported examination of clinical pictures by utilizing the force of deep learning.

Keywords: S*pinal disc herniation, medical imaging, non-invasive examination, spinal disc herniation on MRI images, abnormalities in MRI images, intricate anatomical structure of spine, classify spinal disc herniation, automated detection and categorization of spinal disc herniation.*

MR GÖRÜNTÜLERİNDE DERİN ÖĞRENME KULLANILARAK OMURGA DİSK HERNİSİNİN OTOMATİK TESPİTİ VE SINIFLANDIRILMASI

ÖZET

Omurlar arası diskleri etkileyen yaygın bir patolojik durum spinal disk hernisidir. Diskteki jel benzeri madde sert dış tabakadan dışarı çıktığında sinir sıkışması ve ağrı meydana gelir. İdeal tedavi düzenlemesi ve hasta değerlendirmesi için spinal disk hernisinin kesin ve kışkırtıcı bir bulgusu yapılmalıdır. Spinal disk hernisi, uygun tedavi planlanmadan önce hızlı ve kesin bir tanı gerektirir. Manyetik Rezonans Görüntüleme (MRG), spinal disk hernisinin teşhisinde ve değerlendirilmesinde etkili bir görüntüleme tekniği olarak gelişmiştir, hassas anatomik bilgi sunar ve invaziv olmayan incelemeye olanak tanır. MRI, spinal disk hernisini tanımlamak için etkili bir görüntüleme yöntemi olsa da, radyologların manuel inceleme prosedürü zaman alıcıdır ve hataya açıktır.

Otomatik bir çözümün geliştirilmesi, omurganın karmaşık anatomik yapısını, görüntüleme tekniği değişkenliğini ve açıklamalı veri kümelerinin eksikliğini içerir ve önemli zorluklar ortaya çıkarır. Son birkaç yılda, derin öğrenme algoritmaları tıbbi görüntülemenin yorumlanmasını dönüştürdü ve tespit ve sınıflandırma gibi otomatik görevlerde dikkate değer bir yetenek gösterdi. Bu nedenle, MRI görüntülerinde spinal disk herniasyonunu tanımlamak ve sınıflandırmak için derin öğrenme tekniklerini kullanan bir bilgisayar sistemine ihtiyaç vardır. Bu stratejiler, büyük veri kümelerinden anında belirli desenleri ve özellikleri edinebildikleri için geleneksel bilgisayar destekli tanı yaklaşımlarından üstündür.

Ancak, omurlar arası disk herniasyonunu ayırt etmek için hem güvenilir hem de etkili bir derin öğrenme sistemi oluşturmada birçok zorluk vardır. Omurganın karmaşık anatomisi, görüntüleme stratejilerinin öngörülemez doğası ve MRI görüntülerinde anormalliklerin oluşumunu ele alabilen güçlü algoritmalar. Ayrıca, derin öğrenme modellerini etkili bir şekilde eğitmenin önündeki bir diğer engel de büyük, açıklamalı veri kümelerinin olmamasıdır.

Bu nedenle, bu zorlukların üstesinden gelen ve spinal disk herniasyonunu doğru ve otomatik olarak tanımlamak ve sınıflandırmak için MRI taramalarını kullanan bir derin öğrenme sisteminin nasıl geliştirileceğini keşfetmek gerekir.

Bu araştırma, derin öğrenme yeteneklerini kullanarak MRI verilerinden spinal disk herniasyonunu tanımlamak ve kategorize etmek için MRI verilerini kullanarak spinal disk herniasyonunun otomatik tespiti ve kategorizasyonu için özel olarak tasarlanmış bir derin öğrenme sistemi geliştirdi. Araştırma, bilgisayar destekli tanıya yönelik geleneksel yaklaşımların etkinliğini derin öğrenme tabanlı yöntemle karşılaştırdı. Robotik çerçevenin klinik yararlılığını, radyologların doğru tedavi yöntemine ve spinal disk hernisi kuruluna karar vermelerine yardımcı olma yeteneğini değerlendirerek araştırmanın yanı sıra.

Bu tür derin öğrenme sistemini terapötik olarak yararlı bir araca uygulayarak, radyologlar bunu MRI verilerinden spinal disk hernisini doğru bir şekilde teşhis etmek ve sınıflandırmak için kullanabilirler. Böyle bir sistemin uygulanabilir alımı,

muhtemelen gösterge etkileşimi, tedavi düzenlemesi ve hasta sonuçları üzerinde çalışabilir. Derin öğrenme tekniklerinin spinal disk hernisinin teşhisini ve tedavisini iyileştirmedeki etkinliğini göstererek, araştırma bilgisayar destekli tıbbi görüntü analizinin ilerlemesine katkıda bulunmuştur.

Bu çalışma, spinal disk hernisi keşfinin alanını değiştirebilir ve derin öğrenmenin gücünden yararlanarak klinik resimlerin PC destekli incelemesinin gelişimine katkıda bulunabilir.

Anahtar kelimeler: *Spinal disk hernisi, tıbbi görüntüleme, non-invaziv inceleme, MRI görüntülerinde spinal disk hernisi, MRI görüntülerindeki anormallikler, omurganın karmaşık anatomik yapısı, spinal disk hernisinin sınıflandırılması, spinal disk hernisinin otomatik tespiti ve kategorizasyonu.*

1. INTRODUCTION

1.1 Introduction

Spinal disc herniation is a frequent pathological disorder that affects the intervertebral discs. Nerve compression and pain result when the jelly-like material in the disc protrudes through the tough outer layer [1]. Accurate and prompt diagnosis of intervertebral disc herniation is essential for optimal treatment planning and patient care. The non-invasive examination and precise anatomical information provided by magnetic resonance imaging (MRI) have made it an effective imaging method for the diagnosis and evaluation of intervertebral disc herniation [2].

Deep learning algorithms have the potential to be invested deeply in all fields of medicine, from drug discovery to clinical decision making, significantly altering the way medicine is practiced. The success of Deep learning algorithms at computer vision tasks in recent years comes at an opportune time when medical records are increasingly digitalized [3]. This study aims to develop a computerized method for detecting and classifying intervertebral disc herniation from MRI data by taking advantage of deep learning capabilities. Deep learning algorithms, in particular convolutional networks, have rapidly become a methodology of choice for analyzing medical images. Deep learning procedures for detecting and staging intervertebral disc herniation have various expected benefits. First, by automating the labor-intensive process of manual review, it may extraordinarily reduce the burden on radiologists [4]. In addition, by reducing human error and potentially subjective interpretations, it has a great opportunity to increase heuristic consistency and accuracy. human error plays a significant role in shaping the outcomes and effectiveness of these crucial learning processes. [5].

Finally, CNN models are able to diagnose and grade spinal disc herniation using MRI scans with high accuracy, reliability, and post-hoc interpretation ability comparable or superior to expert radiologists and may function as a supporting diagnostic tool [6].

However, fostering a robust and reliable deep learning model and powerful algorithms equipped to manage artifact event in MRI image of intervertebral disc herniation presents different challenges. The complexity of the anatomical structure of the spine, and the unpredictability of imaging strategies, Deep learning is easier to train than traditional machine learning methods, but requires more data and much more care in analyzing results. It will automatically find the features of importance, but understanding what those features are can be a challenge. [7].

Furthermore, the lack of large annotated datasets represents an additional obstacle to successfully training deep learning models, deep learning requires the availability of large labeled datasets. Annotating imaging datasets is an extremely time-consuming and costly process that is normally undertaken by medical doctors [8].

This study proposes to overcome these issues by creating a deep learning system specifically for the automated identification and classification of intervertebral disc herniation using MRI scans. This approach will involve collecting a comprehensive data set of labeled MRI scans from patients with disc herniation, along with healthy control subjects.

We will use state-of-the-art deep learning models, such as convolution

Neural networks (CNNs), which will be investigated and selected to understand the complexity of the task and the currently existing computational resources. As a spot to begin, pre-trained models like Visual Geometry Group (VGG), Residual Network (ResNet), and Inception can be investigated.

To gain discriminative highlights and examples from MRI images, the chosen model will be trained on the labeled dataset, advancing its parameters by the minimization of a loss function [9]. Additionally, we will study methodologies, for example, augmenting data, transfer learning, and assembling to improve the generalizability and resilience of the proposed system, Data augmentation is a common solution to overcome this issue and various augmentation techniques have been applied to different types of images in the literature [10]. The creation of a therapeutically useful tool that radiologists can use to correctly diagnose and classify spinal disc herniation from MRI data is the ultimate goal of this research. The viable reception of such a system has the possibility to improve diagnostic process, treatment planning, and patient outcomes. This study can possibly change the area of spinal disc herniation discovery and add to the development of computer-aided analysis of medical images by using the power of deep learning.

1.2 Research Problem

Spinal disc herniation is a common illness that need an accurate and fast determination to get a fitting treatment. While MRI is a compelling imaging technique for detecting spinal disc herniation, the manual review technique performed by radiologists is tedious and prone to mistake. Therefore, a computerized system that uses deep learning techniques to identify and classify spinal disc herniation on MRI images is required. However, issues like the spine's complicated anatomical structure, imaging technique variability, and the shortage of annotated datasets present significant challenges to the creation of a robust and reliable automated solution. As a result, the research question is how to develop a deep learning system that overcomes these challenges and uses MRI scans to accurately and automatically identify and classify spinal disc herniations.

1.3 Research Questions

This thesis addresses the following questions:

1. Can a deep learning system discover the presence of a herniated spinal disc in MRI images?
2. What deep learning architectures, such as CNNs or Recurrent Neural Networks (RNNs), are most suitable for automated detection and classification of spinal disc herniation?
3. How can the deep learning system handle the variability in imaging protocols and artifacts commonly present in MRI images of spinal disc herniation?
4. To what extent does the performance of the automated system depend on the size and quality of the annotated dataset used for training?
5. Is it possible for data augmentation techniques to improve the robustness and generalizability of the deep learning system used to identify and classify spinal disc herniation?

6. How does the proposed deep learning-based approach compare to traditional computer-aided diagnosis methods in terms of accuracy, efficiency, and consistency?
7. Can the automated system assist radiologists in the diagnosis and treatment planning of spinal disc herniation, leading to improved patient outcomes?
8. What are the potential limitations and challenges in implementing the deep learning system in clinical practice, and how can they be addressed?

1.4 Thesis Objectives

The thesis objectives provide a roadmap for the research study, outlining specific goals and milestones to be achieved in the development and evaluation of an automated system for spinal disc herniation detection and classification using deep learning on MRI images. The following objectives will be covered in this thesis:

1. Create a deep learning system for detecting spinal disc herniation in MRI images.
2. Research and optimize deep learning architectures for the automated diagnosis and classification of spinal disc herniation, such as CNNs and U-net.
3. Examine the implications of data augmentation approaches on the efficiency and generalizability of the deep learning algorithm for detecting and classifying spinal disc herniation.
4. Comparing the effectiveness of the deep learning-based technique to computer-aided diagnosis ones
5. Evaluate the clinical utility of the automated system through an evaluating

Ability to assist radiologists in determining the correct course of treatment and location of intervertebral disc herniation.

6. Improving computer-aided medical image analysis by demonstrating the ability of deep learning approaches to improve spinal herniation disc analysis and treatment.

1.5 Research Significance and Contributions

This thesis makes several contributions over the earlier studies Contrary to the prior researched, this thesis may make its contributions as following:

1. Increase in diagnosis accuracy: Levered deep learning to develop an automated system for detection and classification of spinal disc herniation, which can significantly enhance diagnosis accuracy. It uses deep learning algorithms to analyze vast amounts of data to recognize subtle and complex patterns and features while reducing the chances of human error. Increased accuracy may steer patient detection and therapeutic management to a more accurate treatment plan.
2. Higher efficiency and workflow improvement: Deep learning assisted automation in identification and classification of spinal disc herniations will help the radiologists in reducing the time as well as expertise required for manual evaluation, The system can quickly analyze MRI images and give immediate results, which allows radiologists to concentrate more on difficult cases available to improve the overall efficiency of work process.
3. Enhancement in Medical Image Analysis: Other applications may benefit from these deep learning approaches to automated analysis of MRI images, specifically with respect to spinal disc herniation. The insights of knowledge and methodologies from deep learning models, advancement approaches, and data increase strategies could assist in the development of different medical imaging occupations outside spinal disc herniation.
4. Clinical Significance: An efficient automated system for detection and classification of spinal disc herniation in MRI images may be used clinically and, in practice, immediately. The created deep learning system might be integrated into current radiology processes and help radiologists in typical practice by offering a supportive device for proficient and precise diagnosis.

1.6 Research Challenges

The research challenges below underneath the complexity and factors that should be addressed to build a successful and solid deep learning system for the automated diagnosis and classification of spinal disc herniation utilizing MRI data:

1. Complex anatomical structure: The complexities of the anatomical structure of the spine make the recognition and classification of spinal disc herniation troublesome. Powerful algorithms capable of accurately segmenting and locating ruptured discs in MRI images are required due to the different shapes and sizes of the discs and surrounding structures.
2. Imaging protocol variation: Imaging protocol variation, such as changes in scan sequences, slice thickness, and contrast parameters, may be evident on MRI scans of spinal disc herniation. It is an important issue to fine-tune a deep learning system to handle these differences while ensuring its stability across different imaging methodologies.
3. Limited availability of classified datasets: The size and availability of datasets with annotations of MRI images including classified spinal disc herniations are usually limited. Building a large dataset with sufficient diversity and accurately labeled models is crucial to effectively training deep learning models.
4. Resolving class imbalance: Since spinal disc herniation is an uncommon disorder unlike healthy conditions, the data set has class imbalance. Tending to this class imbalance is troublesome since deep learning models might become biased for the larger part class. To accomplish fair representation of the two classes during training, appropriate sampling strategies, data augmentation, or loss function adjustments must be used. should be utilized.
5. Handling Artifacts and Noise: MRI images can be influenced by a variety of artifacts and noise, including motion artifacts, scanner-related artifacts, and physiological noise. These abnormalities can make proper identification and categorization of spinal disc herniation difficult. It is a huge problem to develop robust algorithms that can properly handle and limit the effects of such artifacts and noise on the performance of the deep learning system.

6. Clinical Interpretability: Deep learning models are often considered black-box models due to their complex architectures and internal representations. Interpreting the decisions and reasoning of the automated system in a clinically meaningful way is a challenge. Developing techniques to provide interpretable and explainable results to radiologists can enhance the acceptance and adoption of the system in clinical practice.

1.7 Related Work

E. Salehi, H. Yousefi [11] developed an automated system for the detection of spinal disc herniation on MRI images using machine learning techniques. Their work focused on accurately identifying the presence of herniated discs. The strength of their approach lies in achieving promising results in herniation detection. However, a limitation is the lack of exploration of deep learning techniques specifically.

Šušteršič et al. [12] proposed addresses the challenging task of localizing lumbar discs in MRI. It utilizes a deep learning-based approach for automatic segmentation and classification of disc herniation. The methodology comprises of some steps, including segmentation of the disc area as the region of Interest (ROI), bounding box cropping, RoI enhancement, and classification based on a CNN.

Researchers in [13] developed another methodology for autonomously recognizing and categorizing lumbar disc herniation in MRI scans. The methodology uses information from the MRI scans to classifies the discs into four groups: typical, protruding, prolapsed, and expelled. The CNN was trained utilizing a collection of MRI images that had been manually classified by radiologists. On a test dataset of MRI images, CNN accomplished 94% accuracy. CNN accomplished a precision of 94% on a test dataset of MRI images. This methodology beats laid out techniques for recognizing and categorizing lumbar disc herniation, like manual segmentation and grading. It is also more effective because it can quickly and easily classify huge datasets of MRI images. The researches argue that this technology could help with the identification and treatment of lumbar disc herniation. They too accept that it very well may be used to build new imaging methods for identifying and categorizing various types of clinical problems.

Hou et al. [14] focused on finding a way around the lack of labeled data in the herniation detection field. To prepare a deep learning model with both labeled and unlabeled information, they fostered a semi-supervised learning framework. The strength of their research lies in the use of unlabeled information, which can be all the more promptly available. Nonetheless, an impediment might be the likely effect of the quality and representativeness of the unlabeled information on the model's performance.

Cheng et al. [15] presented a multi-scale deep learning framework for the detection and segmentation of spinal disc herniation. Their work meant to catch progressive elements and work on the localization of herniated discs. The contribution of their work lies in the consolidation of multi-scale analysis. In any case, a potential shortcoming might be the increased complexity, what's more, computational above of the multi-scale approach.

Mbarki et al. [16] proposed a Generative Adversarial Network (GAN) for the synthesis of labeled training data for herniation detection. Their work resolved the issue of restricted named information by creating synthetic samples. The expansion of the training dataset lies in augmentation of the training dataset. However, a potential drawback could be a disparity between the synthetic and real data distributions.

Al-Kubaisi et al. [17] focused in on the use of transfer learning in herniation recognition utilizing deep learning models. Their work explored the transferability of pre-trained models from related clinical imaging tasks. The contribution of their work lies in the investigation of transfer learning for herniation identification. Notwithstanding, a potential shortcoming might be the requirement for domain adaptation to guarantee effective transfer. Shi et al. [18] investigated into how X-ray image herniation detection made use of attention mechanisms. Their work focused on salient features also, working on the interpretability of the deep learning models. The strength of their work lies in the integration of attention mechanisms. However, the potential increase in model complexity and computational resources may be a limitation.

Liu et al. [19] proposed a federated learning approach for herniation discovery, meaning to train models collaboratively across various clinical centers without sharing patient information. The contribution of their work lies in the security safeguarding nature of federated learning. Nonetheless, a potential shortcoming might be the expanded complexity and communication requirements of the federated framework.

1.8 Thesis Organization

This thesis is partitioned into five chapters. Every chapter starts with a short outline that offers an overall impact on the part. They are classified as follows**:**

Chapter Two: entitled "Literature Review" which involves:

- Overview of spinal disc herniation and its clinical significance.
- Traditional methods for spinal disc herniation detection and classification.
- Introduction to deep learning and its applications in medical imaging.
- Related work on automated detection and classification of spinal discrimination.
- Summary of strengths and weaknesses of existing research.

Chapter Three: entitled "The Proposed System". It covers the proposed system and its algorithms.

Chapter Four: entitled "Results and Discussion". This Chapter demonstrates the results of the proposed system and the research experiments. It also discusses the evaluation of the system's performance.

Chapter Five: entitled "Conclusions and Future Works". This Chapter presents the research conclusion and the possible future research directions to improve this work.

2. LITERATURE REVIEW

2.1 Introduction

In recent years, clinical imaging strategies play had a urgent impact in the early detection and exact diagnosis of different diseases and conditions. Among these imaging modalities, MRI stands as a powerful non-invasive tool widely used in the field of musculoskeletal and neurological disorders. Specifically, MRI has proven to be instrumental in the evaluation of spinal disc herniation—a common pathology affecting the intervertebral discs of the spinal column. Spinal disc herniation, commonly referred to as a "slipped disc" or "ruptured disc," occurs when the soft inner core of an intervertebral disc protrudes through its fibrous outer ring, resulting in compression or irritation of nearby spinal nerves. This condition can result in debilitating pain, may progress to numbness, and, in severe cases, neurological deficits [20].

Early and accurate detection of spinal disc herniation is very important for timely intervention and giving the effective treatment for patients. To detect and classify spinal disc herniation by MRI doctors or physicians, previously they relied primarily on manual examinations and standard image processing techniques. Although these techniques have yielded useful insights, they are often presented subjectively and with variance by observers and the processes are time-consuming. In recent years, the development of deep learning has shown tremendous potential in various fields, including medical image analysis. In this chapter we will conduct a comprehensive literature review of the latest methods, procedures, and research related to automatic detection and classification of spinal disc herniations using deep learning from MRI data. Through critical analysis of existing research, our goal will be to identify the advantages and disadvantages of current technologies, focus on research needs and create the foundations for a proposed automatic identification system. The main purpose of this literature review is to evaluate the feasibility and utility of deep learning models to improve the diagnosis of intervertebral disc herniation.

Also, we will investigate potential problems of implementing these types of models in a real clinical context, focusing on reliable and accurate diagnosis, while taking into account ethical and patient safety issues. The following sections provide an overview of the use of MRI in detecting spinal disc herniation. Next, classical detection and classification methods are explored, followed by a comprehensive study of deep learning in medical imaging.

Finally, this chapter concludes with a discussion of the novelty of the proposed approach and its contribution to the broader field of medical imaging.

By conducting a comprehensive literature review, this study aims to highlight the potential of deep learning to advance diagnosis of disc herniation and computer-aided diagnosis ultimately benefiting patients and healthcare providers in the same way.

2.2 Traditional Approaches for Spinal Disc Herniation Detection

Before the Advent of Advanced Imaging techniques such as MRI and CT, detection and diagnosis of spinal disc herniation relied heavily on traditional radiographic and medical examination approaches. However, these approaches lack the Precision and sensitivity of modern Imaging techniques, but they played a crucial role in the early understanding and diagnosis of spinal disc anomalies.

<u>X-ray radiography</u> [21],[22] was one of the earliest methods used to examine the spine for signs of disc herniation. Plain X-rays provide a two-dimensional view of the vertebral column, allowing for the evaluation of spinal alignment and any bony changes. While X-rays are beneficial for ruling out some illnesses, their sensitivity in detecting soft tissue abnormalities, such as herniated discs, is limited since they predominantly examine bony features.

<u>Myelography</u> [23],[24] a contrast agent is injected into the subarachnoid space around the spinal cord myelography. The contrast material improves vision of the spinal cord and nerve roots, allowing for a better understanding of any nerve compression produced by herniated discs. on the other hand, myelography is an invasive treatment with risks that has mostly been replaced by non-invasive MRI.

myelography still preferred in the diagnosis of spinal stenosis, particularly if positional components are involved in the patient's symptoms.

Discogram [25] (also called discography) is a type of imaging test that helps diagnose chronic (long-term) back pain. It may be able to tell your healthcare provider whether your symptoms result from a damaged spinal disc. Surgeons may also use discograms before spinal fusion surgery to confirm which discs need to be removed. healthcare providers refer to cervical or lumbar discograms. These terms refer to the same procedure but specify which section of your spine needs imaging. If you have neck or upper back pain, you'll need a cervical discogram. If you have lower back pain, you'll need a lumbar discogram. It works by injection a contrast material (dye) into each disk to highlight potential areas of concern. If the contrast stays contained inside of your disk, it's healthy. If the contrast spreads beyond the outer border of your disk, the disk might be worn or torn.

In addition, medical evaluation remains an important part of the diagnostic process, complemented by new imaging tools that allow complete evaluation of the patient and appropriate management of spinal disc herniation.

2.3 MRI in Spinal Disc Herniation and Lumbar Spinal Stenosis

MRI has become the backbone of modern Medical Imaging. MRI provides important information for the diagnosis and characterization of musculoskeletal and neurological problems such as spinal disc herniation. MRI has non-invasive features and high soft tissue contrast and multi-plane Imaging capabilities. MRI is there for an invaluable Aid in the evaluation of spinal diseases, especially intervertebral disorders.

MRI plays a critical role in the diagnosis and classification of spinal disk herniation by providing a wealth of anatomical information and accurate Imaging of spinal components. It identifies spinal disc herniations based on their location, size, shape and degree of nerve compression [26]. Advances in MRI technology have gradually increased the field strength of superconducting magnets in MRI devices, ranging from 0.15 T to 9.4 T in approximately situations [27]. Typical spine MRI techniques include axial and sagittal images generated by T1 and T2 weighted sequences [28]. The sagittal image T2 weighted pictures, such as short tau inversion recovery or contrast-enhance T1 images, are frequently included into sagittal Imaging according on the unique features and phase of the illness depending on the unique characteristics and stage of the disease [28]. T1 weighted MRI images provide excellent contrast between various soft tissues to help visualize the anatomy of the spine, vertebral bodies, and intervertebral discs. On the other hand, T2 weighted images are particularly sensitive to water content, which allows better visualization of the nucleus pulposus and annulus fibrosis of intervertebral discs.

In addition, T2 weighted images emphasize inflammation and edema around the herniated disc, which provides important information about the extent of nerve root or spinal cord compression [29].

Besides, refined MRI approaches like Diffusion Weighted Imaging (DWI) and Diffusion Tensor Imaging (DTI) give data on the microstructure of spinal structures such the intervertebral discs. these approaches might be used to examine water diffusion patterns and anisotropy inside the disc, giving essential knowledge into this degeneration and disease [30]. The T1 weighted and T2 weighted MRI pictures for the lumber spine District are shown in Figure 2.1.

(A) (B)

Figure 2.1: (A)T1 Weighted Image, (B) T2 Weighted Image[30]

Ongoing advances in MRI technology have likewise brought about the presentation of Functional MRI (FMRI) and Dynamic Contrast Enhanced-MRI (DCE-MRI) to investigate the dynamic behavior of spine and measure vascular perfusion in the herniated disc and encompassing tissues [30],[31]. These forefront imaging innovations can possibly work on how we might interpret the system of spinal disc herniation and its relationship to clinical side effects. It is key to underline that while MRI a valuable resource for envisioning spinal disc herniation, clinical judgment stays essential in the specific comprehension of MRI discoveries. Not all herniated discs lead to symptomatic radiculopathy or require immediate surgical intervention. Therefore, the integration of clinical history, physical examination, and MRI findings is crucial for arriving at an accurate diagnosis and formulating an effective treatment plan.

In conclusion, MRI stands as a fundamental imaging modality for the non-invasive assessment of spinal disc herniation and Lumbar Spinal Stenosis. Its ability to provide detailed and multi-dimensional imaging, coupled with advances in functional and dynamic imaging techniques, has solidified its place as the gold standard for diagnosing and characterizing spinal disc herniation, ultimately contributing to improved patient outcomes.

2.4 Lumbar Spine and Lumbar Spinal Stenosis

The lumbar spine, which is the lower back of the vertebral section, comprises of five interlocking spinal portions associating the upper thoracic spine and the lower sacral spine. The lumbar spine bears more burden than different locales of the spine, making it more inclined to degeneration and injury. Corruption or injury to the lumbar spine frequently prompts Ongoing Lower Back Torment (CLBP), an incapacitating condition portrayed by side effects, for example, emanating pain, abnormal leg pain, and neurogenic claudication [32].

When trying to figure out what's causing your back pain, an MRI is the best option. MRI considers envisioning the lumbar spine in three view-planes: sagittal (side), axial (top-down), and coronal (front facing). Ordinarily, just the sagittal and pivotal perspectives are used in lumbar spine MRI. The mid-sagittal MRI view, as shown in **Figure 2.2**, displays the five lumbar vertebrae and the Intervertebral Discs (IVDs) that separate them. The last IVD, labeled D5, separates the fifth lumbar vertebra (L5) and the sacrum, a large train gular-shaped bone at the base of the spine [33].

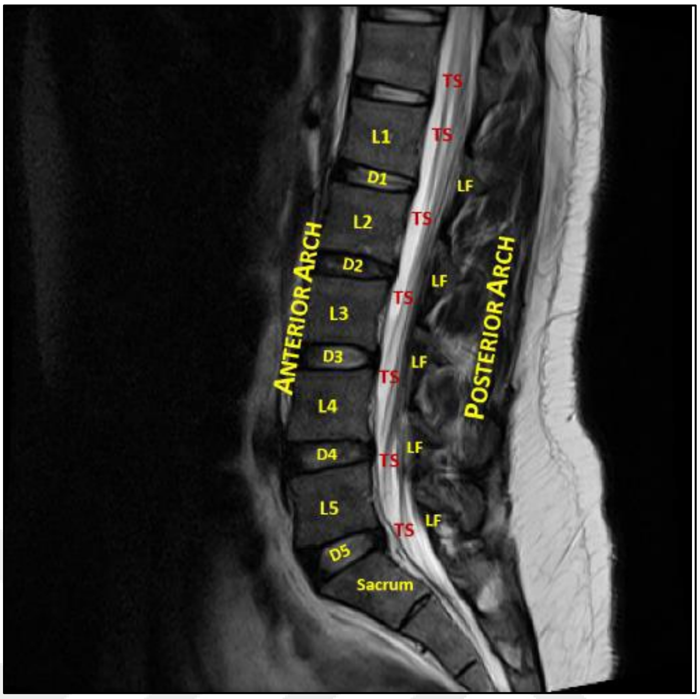


Figure 2.2: Mid-Sagittal MRI Scan on a Lumbar Spine [33]

Figure 2.2 likewise uncovers a long white opening addressing the front curve and the back curve. The apparent piece of this opening in the mid-sagittal cut is the Thecal Sac (TS), containing Cerebrospinal Fluid (CSF), which is like the liquid tracked down inside the brain. The rear of the opening, lining the front of the back curve, is covered with Ligamentum Flavum (LF).

In contrast, various slices of MRI images are displayed across each vertebra or IVD in the axial view of the lumbar spine, which is depicted in Figure 2.3. This view gives more itemized data about the tissues encompassing the vertebrae, IVDs, and the Posterior Element (PE) of the vertebral body. Also, the pivotal view obviously portrays the region between the front and back components, which contains the Thecal Sac and nerve roots in sidelong breaks. This region, which is referred to as the Area between Anterior and Posterior element (AAP), for lack of a better term, in [33].

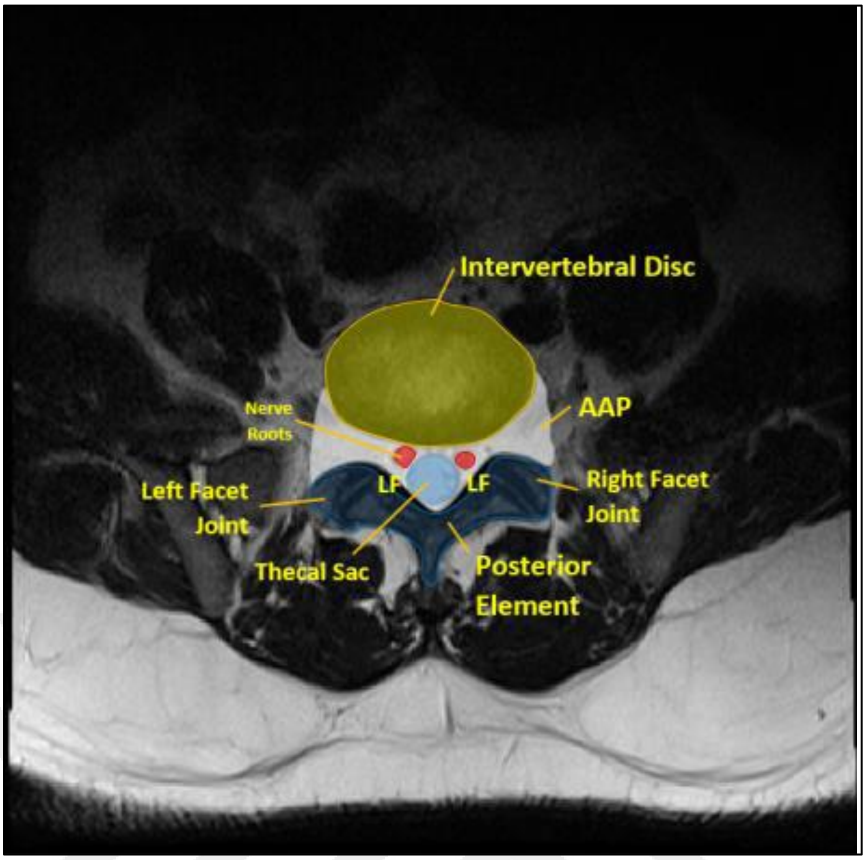


Figure 2.3: Axial View MRI of D4 [33]

Lumbar spinal stenosis, portrayed by the limiting of the AAP, can apply strain on the spinal nerve trench or roots, bringing about torment. Stenosis in various regions of the AAP can be caused by a number of defects, including posterior/posterolateral disc herniation, osteoarthritic thickening of the posterolateral vertebral body, or hypertrophy of LF [23].

In the AAP, clinicians typically measure the anteroposterior diameter of the spinal canal as well as the foramen's left and right widths. These measurements start with manual depiction of the limits between the AAP and the IVD, the left and right aspect joints, and the AAP and LF [32]. **Figure 2.4** show these measurements.

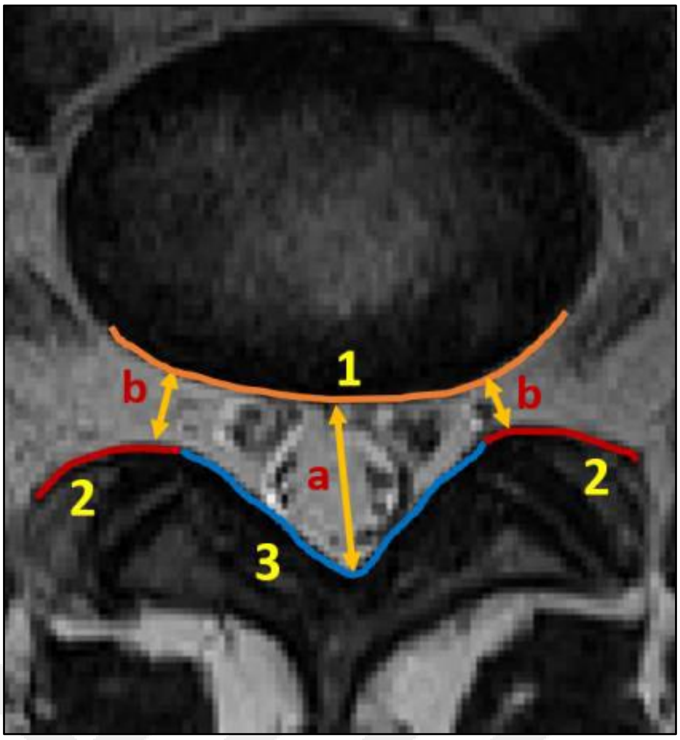


Figure 2.4: Manual Depiction of the Limits Between the AAP and the IVD [32].

2.5 Lumbar Spinal Stenosis Detection and Segmentation

Lumbar Spinal Stenosis (LSS) is a typical degenerative spinal problem characterized by a narrowing of the spinal channel. This results in compression of the spinal cord and nerve roots. [34]. The ability to diagnose LSS early and accurately is critical for prompt intervention and effective patient treatment. Furthermore, segmentation of afflicted areas in medical pictures is critical for evaluating stenosis severity and directing treatment options [35].

2.5.1 Detection approaches

<u>Conventional Image Processing:</u> for example, thresholding, edge detection, and region -based algorithms have been utilized to distinguish stenosis areas in lumbar MRI information [35], [36]. These approaches rely on handmade characteristics and established criteria. However, they may have difficulty dealing with complicated anatomical differences and picture artifacts. The Canny edge detection algorithm is shown in **Figure 2.5**.

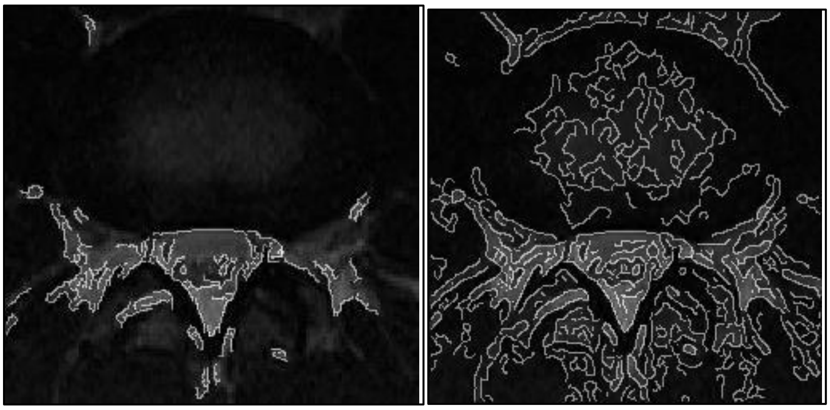


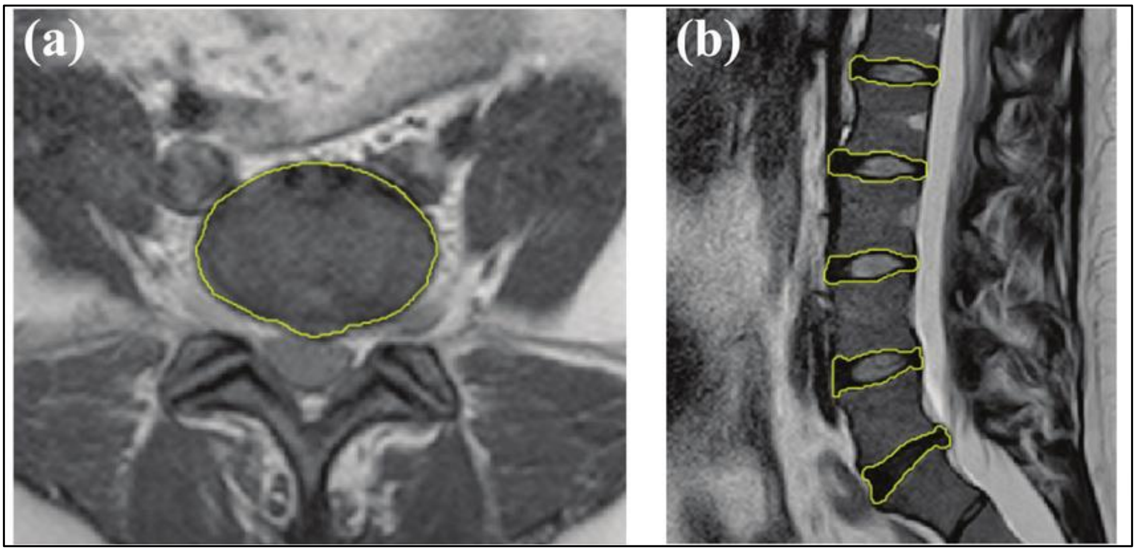
Figure 2.5: The Results of Canny_Edge Detection Algorithm [36].

Machine Learning: Supervised machine learning techniques, such as Support Vector Machines (SVM) and random forests, have been utilized for LSS detection [37], [38]. These approaches rely on manually engineered features extracted from MRI images and require a labeled dataset for training. More recently, deep learning models, such as CNNs, have shown promising results in automated LSS detection by automatically learning discriminative features from large datasets, eliminating the need for handcrafted features [39].

2.5.2 Segmentation approaches

Segmentation algorithms in medical imaging aim to divide an image into meaningful regions or segments, often corresponding to specific structures or regions of interest. Accurate segmentation is critical for quantitative analysis and clinical decision-making.

Manual Segmentation: Manual segmentation by expert radiologists is considered the gold standard for LSS region delineation. However, manual segmentation is time-consuming, subject to inter- and intra-observer variability, and impractical for large-scale studies [40]. Figure 2.6 shows the manual segmentation of two slices of an MRI image [41].



(A) (B)

Figure 2.6: Manual Segmentation, (A) Shows the Segmentation of the Intervertebral Disc L5-S1 in the Axial View, (B) Segmentation of All Lumbar Intervertebral Discs in the Sagittal View [41].

Semi-Automatic Segmentation: Semi-automatic segmentation approaches combine human interaction and computational methods. These methods typically involve the initialization of the segmentation algorithm by the user followed by automated refinement. Interactive segmentation tools and active contour models (snake algorithms) fall under this category [41],[42].

Fully Automatic Segmentation: Fully automatic segmentation techniques utilize machine learning and deep learning methods to delineate stenosis regions without any user intervention. CNNs and other deep learning architectures have been applied to accurately segment LSS regions from MRI images, demonstrating high accuracy and efficiency [43],[44]. While these approaches have shown promise in LSS detection and segmentation, challenges remain, such as dealing with the presence of image artifacts, variations in patient anatomy, and the need for large, annotated datasets for training deep learning models [35].

Additionally, the interpretability of deep learning models in the medical domain is an ongoing concern. Deep learning algorithms, particularly CNNs, have shown remarkable success in medical image segmentation. These models learn complicated properties from big datasets automatically. As a result, they are well-suited for segmenting various structures in medical pictures. The fully automatic segmentation using CNN shown in **Figure 2.7**.

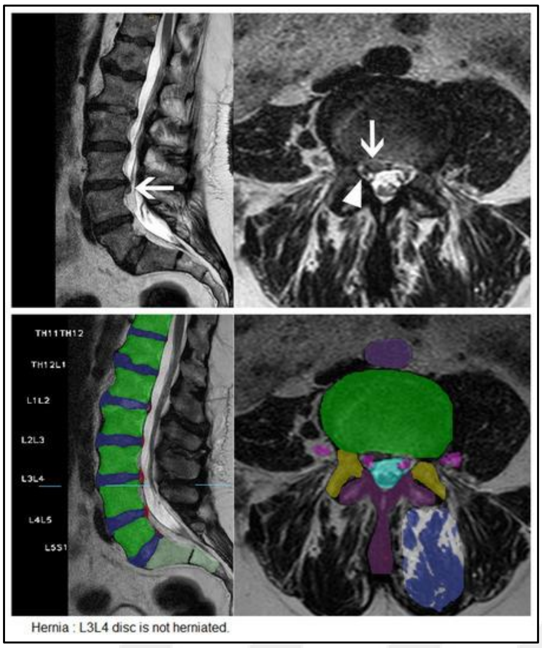


Figure 2.7: Fully Automatic Segmentation by CNN [37]

2.6 Deep Learning in Medical Imaging

Deep learning is a subset of machine learning and Artificial Intelligence (AI). It has become a game-changing technology in a variety of industries, including medical imaging [45]. Because of its capacity to learn and extract complicated patterns from vast datasets automatically. As a result, deep learning has demonstrated enormous potential for changing medical picture processing and interpretation. It contributes to more accurate and efficient illness and condition diagnosis.

Deep learning methods in medical imaging make use of artificial neural networks. It uses CNNs in particular to accomplish tasks including picture classification, object identification, segmentation, and reconstruction. Multiple layers of linked nodes in these neural networks learn and extract hierarchical information from input photos [46]. Neural networks learn from a vast collection of annotated medical pictures through a process known as training. Through training process altering internal parameters of neural networks to maximize performance.

Deep learning has demonstrated encouraging outcomes in medical imaging such as the identification and categorization of spinal disc herniation. Deep learning models outperform standard image processing and pattern recognition approaches in reliably identifying abnormalities in MRI and Computerized Tomography (CT) scans

[47], [48]. They are especially useful when dealing with complicated and diverse medical data. Since they provide resilience and generalization capabilities.

The following are some of the benefits of deep learning in medical imaging:

- Automated Detection: Deep learning models can detect and pinpoint problems in medical photos automatically. This reduces the need for radiologists to do time-consuming manual analysis.
- High Accuracy: Deep learning algorithms have shown remarkable accuracy in image classification tasks, achieving results comparable to, and sometimes even surpassing, human experts.
- Time-Efficient: Deep learning enables rapid analysis of large volumes of medical images, leading to faster and more efficient diagnosis, which is crucial in critical medical situations.
- Transfer Learning: Pre-trained deep learning models can be fine-tuned and adapted to specific medical imaging tasks with relatively smaller datasets, making them applicable in scenarios with limited labeled data.

2.6.1 CNNs and RNNs in medical image analysis

Deep learning represents a subset of machine learning that focuses on learning complex patterns and representations from data using artificial neural networks [49]. CNNs and RNNs are specialized architectures within deep learning that have revolutionized image and sequence data analysis, respectively.

In image analysis tasks, CNNs have demonstrated exceptional proficiency. The intricate anatomical structures found in MRI images of spinal discs correspond to their capacity to automatically learn hierarchical features from images [39]. CNNs excel in capturing spatial relationships, which is crucial in detecting and classifying spinal disc herniation based on MRI images.

RNNs on the other hand, are adept at sequence data analysis. While not the primary focus of this research, it's worth noting that RNNs could find applicability in scenarios where sequences of MRI images or patient data need to be analyzed to track the progression of spinal disc herniation or evaluate treatment effectiveness [49]-[51].

2.7 Segmentation Based on Deep Learning in Lumbar Spinal Stenosis

A promising strategy for automating the process of segmenting medical images in lumbar spinal stenosis is segmentation-based deep learning. Leveraging the power of artificial neural networks, deep learning models can learn to identify and delineate relevant structures in MRI or CT scans with remarkable accuracy. These models are trained on large datasets of annotated medical images, where the ground truth segmentations are provided by expert radiologists [36]. During the inference phase, the trained deep learning model can efficiently and consistently segment the spinal canal, nerve roots, and areas of stenosis in a given image [43]. This automation eliminates the need for manual delineation, reducing inter- and intra-observer variability and significantly expediting the segmentation process.

By providing precise and reliable segmentations, deep learning-based approaches hold immense potential to assist clinicians in diagnosing and assessing lumbar spinal stenosis more efficiently and accurately, ultimately leading to improved patient care and treatment outcomes. However, it is essential to validate these automated segmentations against expert annotations and ensure their reliability in clinical settings before widespread adoption. Segmentation in light of deep learning in lumbar spinal stenosis has acquired critical consideration because of the outcome of deep learning models in different clinical image examination undertakings. Deep learning algorithms, particularly CNNs, have shown remarkable performance in segmenting medical images, including lumbar spinal stenosis. Here are some common deep learning algorithms used for segmentation in this context:

- Fully Convolutional Network (FCN)
- U-Net
- DeepLab
- 3D U-Net

2.7.1 Fully convolutional network (FCN) architecture

FCN is a kind of deep neural network architecture explicitly intended for semantic division errands. Not at all like conventional CNNs that are frequently utilized for picture characterization, FCNs are custom fitted for pixel-wise marking, where every pixel in an info picture is doled out a class name or a class [52]. The essential thought behind FCNs is to supplant completely associated layers in an ordinary CNN with convolutional layers. This layer empowers the network to produce yield with similar spatial dimensions as the input picture. This is basic for undertakings like semantic division. Where it is essential to keep spatial information. Here's how an FCN works [53]:

a) **Input Image:** The FCN accepts an input image of any size.

- Convolutional Feature Extraction: The FCN's initial layers are like typical CNNs. Usually using convolutional processes to extract features from the input picture.

- Convolutional layer [54]**:** A convolutional layer is a crucial building component in CNNs. Which are frequently used for a variety of computer vision applications such as image analysis, object identification, and picture segmentation. Convolutional layers are intended to learn characteristics automatically and adaptively from input photos. As a result, they are particularly well-suited to jobs requiring visual data like as photographs and movies. A convolutional layer's fundamental function is to conduct the mathematical process known as convolution. Convolution is a technique for extracting local patterns or features from an input picture by applying a tiny filter (also known as a kernel) over it and performing element-wise multiplications followed by summing. This method is repeated at each position in the input picture, yielding a feature map representing the presence of certain patterns within the input data.

  Following is how a convolutional layer works, step by step:

1. Filter (Kernel): The filter is a tiny weighted matrix that is smaller than the input picture but greater than a single pixel. It is intended to capture certain characteristics such as edges, textures, or forms.
2. Convolution Operation: The filter is slid over the input picture in a methodical manner. At each point, element-wise multiplication is done between the filter's values and the corresponding values in the input picture. **Figure 2.8** shows convolutional operation.

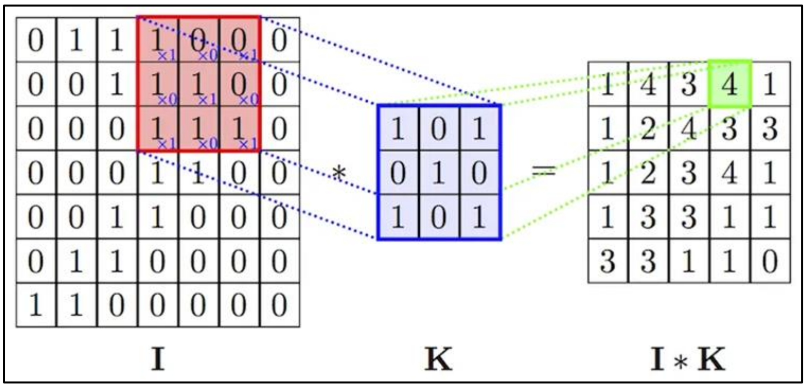


Figure 2.8: Convolutional Operation, Where I Input Image, K Kernel [54]

3. Summation: The results of element-wise multiplications are summed up to create a single value in the output feature map. This value represents the presence of the filter's pattern at that particular location in the input.

4. Strides: The filter is moved by a certain number of pixels at each step, called the stride. The strides opration are shown in **Figure 2.9**.

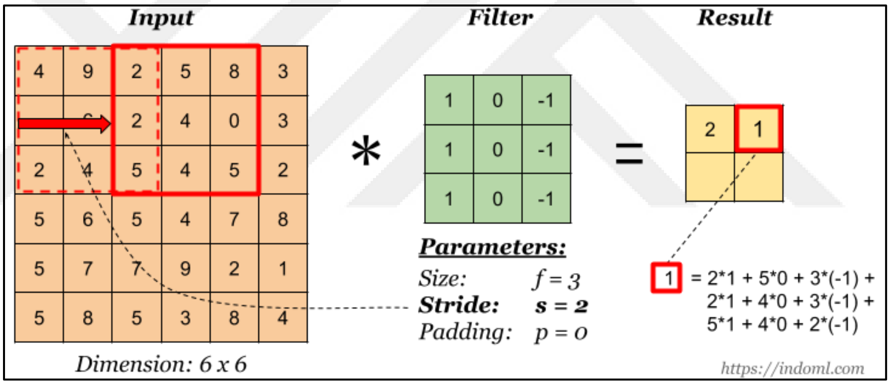


Figure 2.9: Filter With Stride (s) = 2 [53].

5. Padding: Sometimes, the input image is padded with additional pixels around its edges to control the spatial dimensions of the output feature map. Padding can help preserve spatial information and prevent the output from becoming too small. **Figure 2.10** shows padding operation.

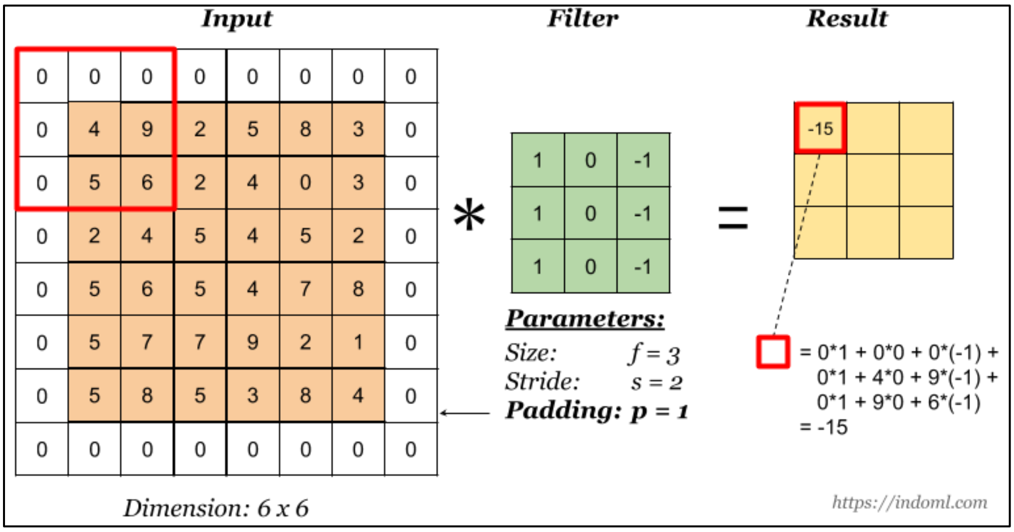


Figure 2.10: Padding Operation [54]

6. Activation Function [55]: An activation function like Rectified Linear Unit (ReLU) is typically used to introduce non-linearity following convolution and summation. ReLU sets negative qualities to nothing and leaves positive qualities unaltered. ReLU is a nonlinear activation function and is many times utilized in deep neural networks. This function returns zero if it received a negative value, but for any positive value returns the same received value. In another concept, ReLU makes to compare the input value with zero and take the maximum one as a winner. ReLU activation function is clarified in **(2.1)** and plotted in **Figure 2.11**.

$$\text{ReLU}(x) = \max(0, x) \tag{2.1}$$

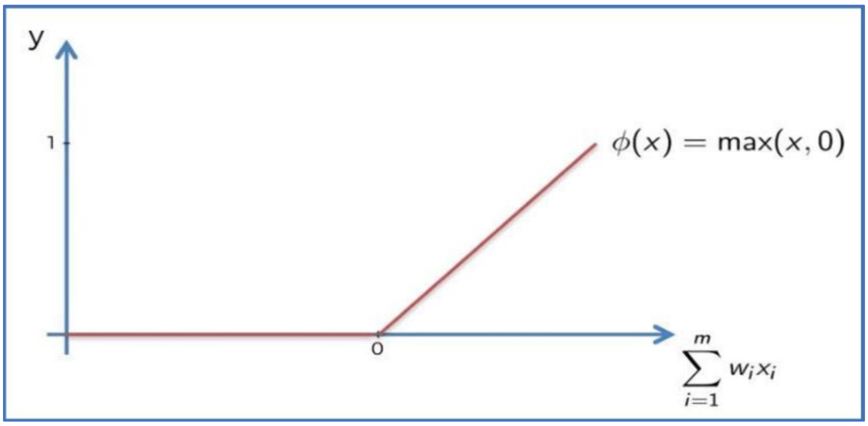


Figure 2.11: ReLU Function [55]

b) **Encoder:** The FCN typically has an encoder component that consists of multiple convolutional and pooling layers. The encoder progressively reduces the spatial dimensions of the feature maps while increasing the number of channels (features). This process captures both low-level and high-level features. Pooling is a form of spatial aggregation that groups neighboring pixels in a feature map and applies an aggregation function to summarize the values within each group. The most common pooling operation is max pooling, where the maximum value within each group is selected. Average pooling, where the average value is calculated, is another option. **Figure 2.12** shows max pooling process.

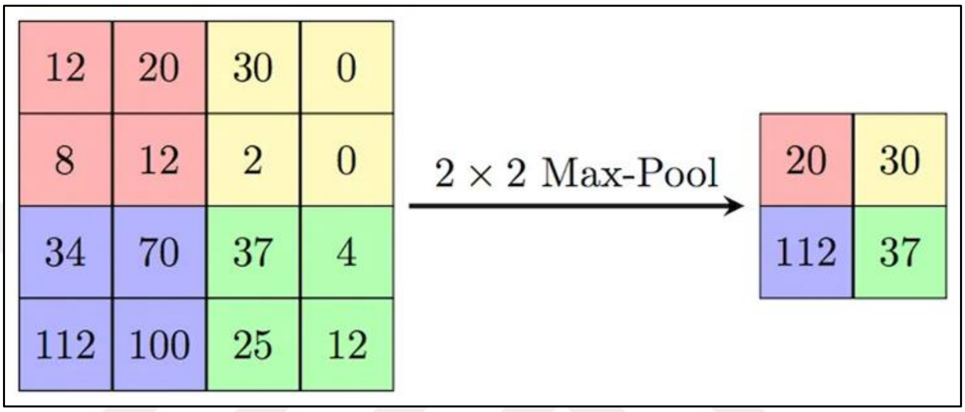


Figure 2.12: Max Pooling Process [55]

c) **Bottleneck Layer:** In some FCN variants, there might be a bottleneck layer that acts as a transition point between the encoder and decoder. This layer further abstracts the features.

d) **Decoder:** The decoder is where the FCN performs the upsampling and produces the final segmented output. It consists of transpose convolutional layers (also known as deconvolutional layers or up-convolutional layers). These layers increment the spatial elements of the component maps back to the first size. The objective is to create a intensive pixel-wise classification map. CNNs are made up of up-convolutional layers, which are also referred to as transpose convolutional layers or deconvolutional layers. They are used for image generation and segmentation. These layers play out the inverse activity of pooling layers, expanding the spatial dimensions of the feature maps. They are especially significant in structures like FCNs and U-Net, where keeping up with spatial resolution is urgent for pixel-wise division undertakings. The operation of upconvolutional layers is described as follows:

1-Input Feature Map: An up-convolutional layer's input is often a low-resolution feature map recovered from a past layer in the network.

2-Up-Convolution Operation: The up-convolutional layer upsamples the input feature map utilizing a filter (otherwise called a kernel). The up-convolutional operation, instead of ordinary convolution, increases the spatial dimensions of the feature map.

3. Strides and Padding: Up-convolutional layers have properties similar to regular convolutional layers. The stride shows how much the filter moves during activity, which influences the spatial dimensions of the result. The output size can be controlled through padding.

4-Upsampling Element: The filter's size and the stride together decide the upsampling factor. A 2x2 upconvolutional filter, for instance, doubles the spatial dimensions with a stride of 2.

5- Aggregation: Similar to standard convolution, up-convolution involves element-wise multiplication between the filter and the input values, followed by summation. The results are placed in the output feature map.

6- Output Feature Map: The output of the up-convolutional layer has larger spatial dimensions compared to the input, effectively producing a higher-resolution feature map. **Figure 2.13** shows Up-Convolution operation.

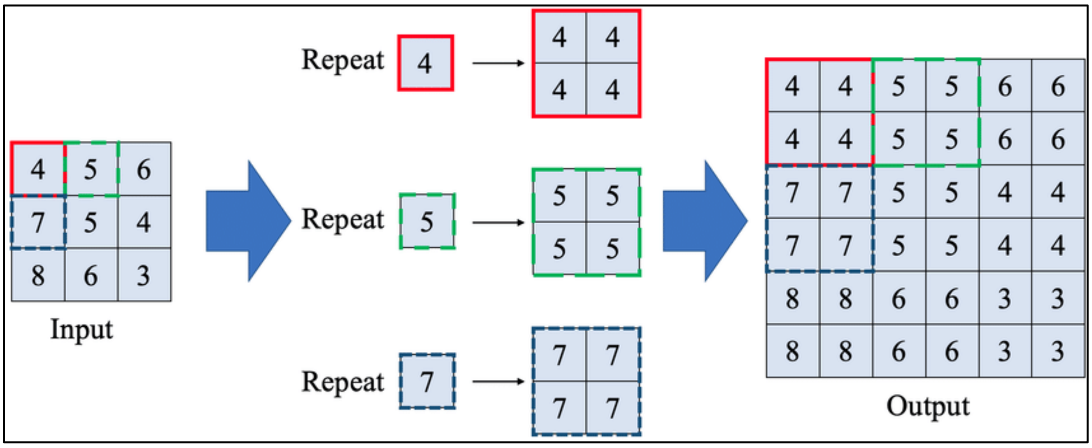


Figure 2.13: Up-Convolution Operation [56]

e) **e) Skip Connections:** In order to capture fine-grained details, many FCN architectures use skip connections. Layers from the encoder are connected to the decoder's corresponding layers by means of these connections. By combining high- resolution features from the encoder with features in the decoder, the FCN can all the more likely restrict object boundaries and details.

f) **f) Final Layer:** The last layer of the decoder outputs a pixel-wise classification map where each pixel's value represents the predicted class for that location.

g) **g) Softmax or Sigmoid Activation:** Depending on the specific task, the final layer might have a softmax or sigmoid activation function. For binary segmentation tasks, like segmenting an object from the background, a sigmoid activation is commonly used to produce per-pixel probabilities. Sigmoid activation is commonly used in scenarios where the task involves binary decisions, such as classifying images into two categories [57]. The sigmoid activation function maps input values to a range between 0 and 1. It's defined as shown in **(2.2)**.

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{2.2}$$

The SoftMax activation is often used in multi-class classification tasks, where the network needs to predict which of multiple classes an input belongs to. The class with the highest probability from the softmax output is typically selected as the network's prediction. The softmax activation function maps a vector of input values to a probability distribution over multiple class [57]. It's defined as shown in **(2.3)**.

$$softmax(x)i = \frac{e^{x_i}}{\sum_{j-1}^{N} e^{x_j}} \quad \text{for I =1,2,….,N} \tag{2.3}$$

The FCN architecture provides end-to-end learning for semantic segmentation problems without the requirement for post-processing processes. It has been effectively used to a variety of applications, including medical picture segmentation, object recognition, and scene parsing.

2.7.2 U-Net architecture

U-Net is a deep learning engineering presented in 2015 for biomedical picture division errands by Olaf Ronneberger, Philipp Fischer, and Thomas Brox [58]. was first introduced in the “U-Net: Convolutional Networks for Biomedical Image Segmentation” paper. The primary purpose of this architecture was to address the challenge of limited annotated data in the medical field. This network was designed to effectively leverage a smaller amount of data while maintaining speed and accuracy. The moniker "U-Net" is gotten from the engineering's U-shape. The U-Net plan is partitioned into two sections: the contracting way (encoder) and the extending way (decoder), which are connected together through skip connections. This method empowers the model to rapidly gather both local features and worldwide context, making it appropriate for segmentation errands requiring precise structure settlement. The U-net architecture is shown in **Figure 2.14**.

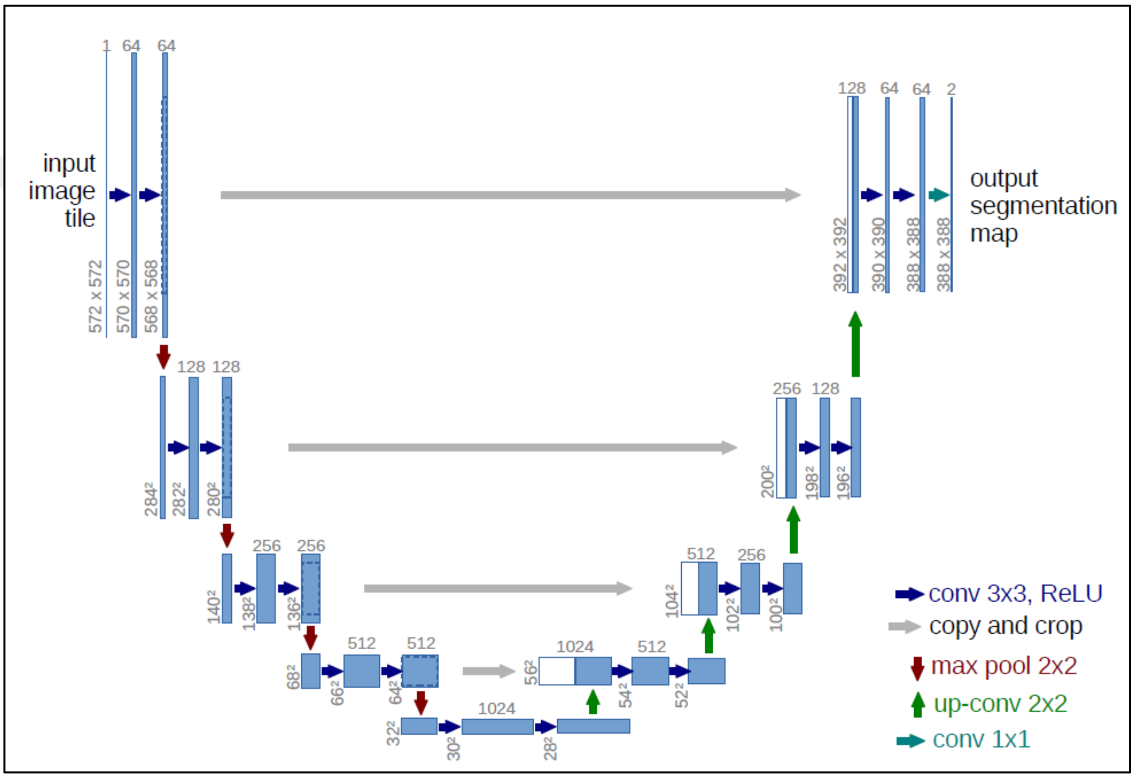


Figure 2.14: U-Net Architecture [58]

- <u>Contracting Path (Encoder):</u> The contracting route is made up of convolutional and pooling layers. This layer gradually reduces the spatial dimensions of the input picture while increasing the number of feature channels. These layers work as a feature extractor, learning to represent low-level and high-level characteristics in the input picture. Convolutions are applied with small filter sizes to capture local patterns, and max-pooling layers are used to down sample the spatial resolution. This down sampling helps increase the receptive field of the network, allowing it to recognize larger patterns in the input [59].

The contracting path resembles the architecture of a traditional CNN and is responsible for extracting abstract feature representations from the input image.

- Expansive Path (Decoder): The expansive path is a series of convolutional and upsampling layers that reduce the number of feature channels while gradually restoring the original size of the spatial dimensions. This path intends to deliver a segmentation map that has a similar resolution as the input picture. Each layer in the expansive path is connected with the relating layer in the contracting path through skip connections. These skip connections connect feature maps from the contracting path with the feature maps from the ongoing layer in the broad path. The skip connections permit the decoder to get to high- resolution features from the contracting path, which helps protect spatial data and helps in precise segmentation [60], [61]. This is especially useful for tasks involving the segmentation of biomedical images, where precise localization of structures is essential.
- Final Layers: The final layer of the U-Net typically consists of a 1x1 convolution followed by a suitable activation function. For binary segmentation tasks like lumbar spinal stenosis segmentation, a sigmoid activation function is commonly used to produce a probability map, where each pixel value indicates the likelihood of belonging to the target structure (e.g., stenosis or non-stenosis).
- Structures in U-Net: The application and the training data used determine the structures that the U-Net can segment. With regards to lumbar spinal stenosis segmentation, the U-Net is trained to fragment the areas of interest connected with stenosis in clinical pictures, for example, MRI or CT sweeps of the lumbar spine [37].

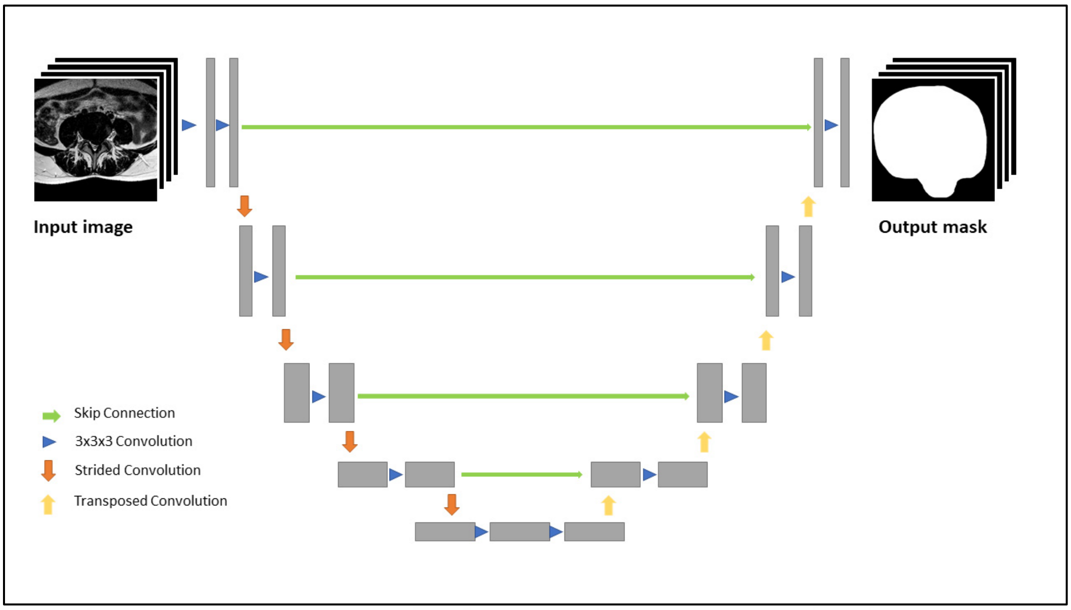


Figure 2.15: The U-Net- Segmentation Architecture for Spinal Canal Stenosis [37].

These regions may include the narrowed spinal canal, areas of nerve compression, or other stenotic regions [37]. The U-Net architecture is known for its simplicity, effectiveness, and ability to handle limited training data. Its design, with the contracting and expansive paths connected by skip connections, enables it to achieve accurate segmentation results, even with a relatively small number of training samples. **Figure 2.15** shows the U-Net- architecture for process of segmentation of disc herniation and spinal canal stenosis.

- Loss Function [62]–[65]

In early models of neural networks, the error is measured by calculating the difference between the real output and the expected output. At the moment, several formulas have appeared for calculating error in neural networks, these formulas are called Loss Functions [64]. When using various loss functions can lead to a different error value for the same prediction, therefore the type of loss function has a major impact on the output of the network.

The choice of loss function is critical to the successful training of the network in U-Net and similar architectures designed for image segmentation tasks. The most regularly involved shortfall capability for U-Net and related models is the Pixel-wise Cross-Entropy Deficit (otherwise called Binary Cross-Entropy loss for binary segmentation tasks) [62]-[65]. Here's how the Pixel-wise Cross-Entropy Loss works:

1- Input and Target Masks: Given an input image, the U-Net produces a segmentation mask that represents the predicted class for each pixel. The target mask is the ground truth segmentation mask that corresponds to the correct class for each pixel.

2- Pixel-wise Cross-Entropy: For every pixel in the anticipated mask and the corresponding pixel in the target mask, the cross-entropy loss is figured. This loss estimates the contrast between the anticipated class probabilities and the genuine class (ground truth) for every pixel.

3- Loss Aggregation: The individual pixel-wise cross-entropy values are aggregated to obtain a single loss value that quantifies the dissimilarity between the predicted and target masks.

Mathematically, the pixel-wise cross-entropy loss for a single pixel can be expressed as shown in (**2.4**).

$$L(x) = -\sum_{c=1}^{c} t_c \, log\big(p_c(x)\big) \tag{2.4}$$

Where:

t_c is the ground truth name (0 or 1) for class **c** at pixel **x**.

p_c(x) is the predicted probability of class **c** at pixel **x**.

C is the number of classes (2 for binary segmentation: foreground and background).

The loss is calculated for each pixel and then averaged over all pixels to obtain the final loss value for the entire image. In the case of binary segmentation, where the task is to segment objects from the background, the loss function simplified as shown in **(2.5)**.

$$L(x) = -tlog(p(x)) - (1 - t)log(1 - p(x)) \tag{2.5}$$

Where:

- **t** is the ground truth label (0 or 1) for the pixel.
- **p(x)** is the predicted probability of the pixel belonging to the foreground class.

It's important to note that while pixel-wise cross-entropy is the most common choice, variations of the loss function can be used to address specific challenges. For example, you might incorporate class weights to balance class frequencies or add additional regularization terms to encourage smoother segmentations. Ultimately, the loss function chosen for training U-Net should reflect the characteristics of the segmentation task and the desired behavior of the model's predictions.

2.8 Evaluation Measures

When evaluating the performance of a segmentation model for lumbar spinal stenosis, several evaluation measures are commonly used to assess how well the predicted segmentations match the ground truth segmentations in medical images. These metrics give information on the segmentation findings' accuracy, precision, recall, and other characteristics. Following are some examples of assessment metrics that can be used:

- Dice Similarity Coefficient (DSC): DSC is a popular statistic for determining the overlap between expected and actual segmentation. It has a value ranging from 0 (no overlap) to 1 (perfect overlap). This can be described as followed in **(2.6)**.

$$DSC = \frac{2 \times Area\ of\ Overlap}{Area\ of\ Predicted\ Segmentation + Area\ of\ Ground\ Truth} \quad (2.6)$$

- Jaccard Index (Intersection over Union (IoU))**:** In the field of computer vision, CNNs are the most common approaches for object detection and their location. Metrics such as IoU are commonly used to measure the performance of this task in various use cases such as segmentation, object detection, and object tracking. IoU can also be used for verification in the domain of explainable AI, i.e. for checking whether the recognized area matches the relevant area for recognition of a specific object. Although IoU is able to clearly distinguish between overlapping and non-overlapping boxes, it does not directly indicate whether the boxes overlap almost completely or lie entirely within each other. It is figured as continued in **(2.7)**.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Predicted\ Segmentation + Area\ of\ Ground\ Truth - Area\ of\ Overlap} \quad (2.7)$$

- Precision and Recall: Precision or Confidence (as it is called in Data Mining) denotes the proportion of Predicted Positive cases that are correctly Real Positives. This is what Machine Learning, Data Mining and Information Retrieval focus on, but it is totally ignored in ROC analysis. It can however analogously be called True Positive Accuracy (tpa), being a measure of accuracy of Predicted Positives in contrast with the rate of discovery of Real Positives (tpr). Precision is shown in (2.8).

$$Precision = \frac{TP}{TP + FP} \tag{2.8}$$

Recall is the proportion of Real Positive cases that are correctly Predicted Positive. This measures the Coverage of the Real Positive cases by the +P (Predicted Positive) rule. Its desirable feature is that it reflects how many of the relevant cases the +P rule picks up. It tends not to be very highly valued in Information Retrieval (on the assumptions that there are many relevant documents, that it doesn't really matter which subset we find, that we can't know anything about the relevance of documents that aren't returned). Recall tends to be neglected or averaged away in Machine Learning and Computational Linguistics (where the focus is on how confident we can be in the rule or classifier). However, in a Computational Linguistics/Machine Translation context Recall has been shown to have a major weight in predicting the success of Word Alignment [1]. In a Medical context Recall is moreover regarded as primary, as the aim is to identify all Real Positive cases, and it is also one of the legs on which ROC analysis stands. In this context it is referred to as True Positive Rate (tpr). shown in **(2.9)**.

$$Recall = \frac{TP}{TP + FN} \tag{2.9}$$

Where:

TP: True Positives (correctly predicted positive pixels).

FP: False Positives (incorrectly predicted positive pixels).

FN: False Negatives (positive pixels missed by the prediction).

- F1-Score: The F1-score is the consistent mean of precision and recall. It gives a fair measure of both precision and recall and is especially helpful when the class distribution is imbalanced., as shown in (2.10).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recal}}{\text{Precision} + \text{Recal}} \tag{2.10}$$

- Surface Distance metrics: measurements like Hausdorff distance and mean surface distance assess the spatial differentiations between the anticipated and ground truth surfaces. They demonstrate how well the segmented areas correspond to anatomical designs.

- Sensitivity and Specificity: Sensitivity (additionally called True Positive Rate) measures the capacity of the model to recognize positive cases accurately. Particularity (True Negative Rate) gauges the capacity to accurately recognize negative cases as shown in (2.11) and (2.12).

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{2.11}$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \tag{2.12}$$

- Accuracy: The general exactness of the segmentation, which measures the extent of accurately classified pixels as shown in (2.13).

$$Accuracy = \frac{Number\ of\ Correctly\ Classified\ Pixels}{Total\ Number\ of\ Pixels} \tag{2.13}$$

- Receiver Operating Characteristic (ROC) Curve: ROC curves are utilized to visualize the compromise among sensitivity and specificity. The area under the Curve-ROC (AUC-ROC) gives a general exhibition measure.

3. THE PROPOSED SYSTEM

3.1. Introduction

The research meant to create a (CNN) model to classify spinal images into 'Osteophytes' and all other diseases as 'minor' lesions. Using the TensorFlow and Keras libraries, a sequential model architecture was created with a VGG16 base for feature extraction and fully connected layers for classification.

3.2 The Model Coding

This section shows the coding of the proposed model architecture that will run on a training data set. **Figure 3.1** shows importing of datasets from libraries and loading it to model. The dataset is stored in “read_csv” file. **Figure 3.2** shows a sample of rows of the dataset that was loaded to understand its structure.

```
import matplotlib.pyplot as plt
import matplotlib.patches as patches
from PIL import Image
import numpy as np
import numpy as np
import os
from PIL import Image
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense
from tensorflow.keras.utils import to_categorical
import pandas as pd
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from sklearn.preprocessing import LabelEncoder
from tensorflow.keras.utils import to_categorical
```

```
# Load the CSV file into a pandas DataFrame
data_df = pd.read_csv('train.csv')

# Display the first few rows of the dataset to understand its structure
data_df.head()
```

Figure 3.1: Importing Data from Library

image_id	rad_id	lesion_type	xmin	ymin	xmax
1f1c5c098c35b79052596aae08ac727f	rad1	Osteophytes	712.636842	961.361404	786.587719
1f1c5c098c35b79052596aae08ac727f	rad1	Osteophytes	789.949123	1452.126316	880.707018
1f1c5c098c35b79052596aae08ac727f	rad1	Osteophytes	692.468421	655.473684	746.250877
1f1c5c098c35b79052596aae08ac727f	rad1	Osteophytes	820.201754	1603.389474	910.959649
1f1c5c098c35b79052596aae08ac727f	rad1	Osteophytes	739.528070	1203.382456	820.201754

Figure 3.2: Sample of Rows of the Dataset

Then a set of spinal images were defined in the model as shown in **Figure 3.3**. The process of generating a list of all PNG image paths in the directory takes place.

```
In [127]:
# Path to the directory containing all the images
directory_path = "Spinal/png"

# Generate a list of all PNG image paths in the directory
image_paths = [os.path.join(directory_path, image_name) for image_name in os.listdir(directory_path) if image_name.endswith(

In [128]:
len(image_paths)

Out[128]: 512
```

Figure 3.3: Generating a List of all PNG Images Paths

Defining the visualization process of images then takes place as shown in **Figure 3.4**. Accordingly, the process of running visualizations of images is now ready. Each image data is visualized as shown in **Figure 3.5**, **Figure 3.6**, **Figure 3.7** and **Figure 3.8**.

```
In [129]:  # Define subset_image_paths for the first 3 images
            subset_image_paths = image_paths[:3]

            def visualize_images_with_platform_independent_labels(image_paths, data_df):
                print(f"Number of images to visualize: {len(image_paths)}")

                fig, axes = plt.subplots(nrows=len(image_paths), figsize=(12, 4 * len(image_paths)))
                for ax, path in zip(axes, image_paths):
                    # Read the image
                    image = plt.imread(path)
                    ax.imshow(image, cmap='gray')

                    # Get the image id from the path using os.path.basename() for platform independence
                    image_id = os.path.basename(path).split('.')[0]
                    print(f"Processing image: {image_id}")

                    # Extract rows corresponding to this image_id
                    rows = data_df[data_df['image_id'] == image_id]
                    print(f"Number of rows corresponding to this image: {len(rows)}")
                    for _, row in rows.iterrows():
                        # If bounding box data exists
                        if not pd.isnull(row['xmin']):
                            rect = patches.Rectangle(
                                (row['xmin'], row['ymin']),
                                row['xmax'] - row['xmin'],
                                row['ymax'] - row['ymin'],
                                linewidth=1, edgecolor='r', facecolor='none'
                            )
                            ax.add_patch(rect)
                            ax.text(row['xmin'], row['ymin'], row['lesion_type'],
                                    bbox=dict(facecolor='red', alpha=0.5), fontsize=8, color='white')
                        else:
                            ax.text(50, 50, "No finding",
                                    bbox=dict(facecolor='green', alpha=0.5), fontsize=10, color='white')

                    ax.axis('off')
                    ax.set_title(image_id)

                plt.tight_layout()
                plt.show()
```

Figure 3.4: Defining the Visualization Process of Images

```
# Run the visualization
visualize_images_with_platform_independent_labels(subset_image_paths, data_df)

Number of images to visualize: 3

Processing image: 00073745e02e69432c002b527c565151

Number of rows corresponding to this image: 4

Processing image: 00087195a35bb9948323aa89ccb2a860

Number of rows corresponding to this image: 3

Processing image: 000f985efcb28afd281e3cd1b4d370ee

Number of rows corresponding to this image: 1
```

Figure 3.5: Running Visualization of Images

Figure 3.6: Visualization 1

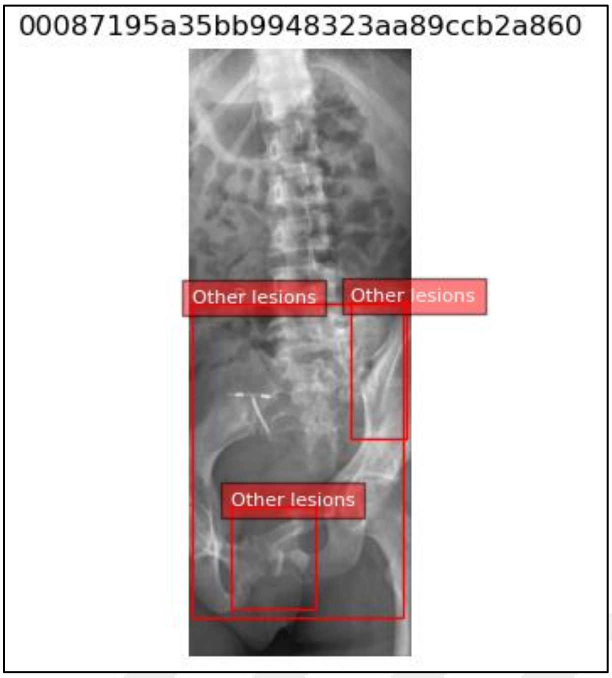


Figure 3.7: Visualization 2

Figure 3.8: Visualization 3

Images need to be cropped and labelled as shown in **Figure 3.9**.

```python
def crop_to_roi(image_path, bbox):
    image = Image.open(image_path)
    return image.crop(bbox)
```

```python
# Augmentation
datagen = ImageDataGenerator(
    rotation_range=10,
    zoom_range=0.1,
    horizontal_flip=True,
    fill_mode='nearest'
)
```

```python
# Cropping images
all_cropped_images = []
all_labels = []

for img_path in image_paths:
    image_id = os.path.basename(img_path).split('.')[0]
    matching_rows = data_df[data_df['image_id'] == image_id]

    if not matching_rows.empty:  # This condition checks if there are matching rows for the image ID
        bbox_row = matching_rows.iloc[0]
        try:
            bbox = [float(coord) for coord in bbox_row[['xmin', 'ymin', 'xmax', 'ymax']].values]
            if not np.isnan(bbox).any():
                cropped_img = crop_to_roi(img_path, bbox)
                all_cropped_images.append(np.asarray(cropped_img))
                all_labels.append(bbox_row['lesion_type'])
        except ValueError:
            pass
```

Figure 3.9: Cropping Images

Accordingly, data of each cropped image is presented as shown in **Figure 3.10**.

```
[134]: print("Number of images:", len(all_cropped_images))
       print("Number of labels:", len(all_labels))

       Number of images: 253

       Number of labels: 253
```

Figure 3.10: Data of Cropped Images

Classification of data is needed to be done on data of cropped images. Minor classes are found. So they are combined into one class of the same classification. This process is shown in **Figure 3.11**.

```
In [135]:
# Combine minor classes into one class
major_class = 'Osteophytes'
all_labels = ['minor' if label != major_class else label for label in all_labels]

# Encode y for binary classification
le = LabelEncoder()
y = le.fit_transform(all_labels)

X = np.array([np.array(Image.fromarray((img * 255).astype(np.uint8)).resize((128, 128))) for img in all_cropped_images])
X = X[..., np.newaxis]  # Add an extra channel dimension
X = X / 255.0

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

In [136]:
y_train.shape

Out[136]: (202,)
```

Figure 3.11: Combining Minor classes Into One Class

The compiled model is shown in **Figure 3.12**.

```
[146]: from tensorflow.keras.applications import VGG16
       from tensorflow.keras.layers import Dense, Flatten, GlobalAveragePooling2D
       from tensorflow.keras.models import Model

       # Load ResNet50 base model
       base_model = VGG16(weights='imagenet', include_top=False, input_shape=(128, 128, 3))

       # Freeze the layers of the base model
       for layer in base_model.layers:
           layer.trainable = False

       # Add custom layers on top of ResNet50
       x = base_model.output
       x = GlobalAveragePooling2D()(x)  # Add a global spatial average pooling layer
       x = Dense(64, activation='relu')(x)  # Add a fully-connected layer
       predictions = Dense(1, activation='sigmoid')(x)  # Add the final output layer

       # Create the final model
       model = Model(inputs=base_model.input, outputs=predictions)

       # Compile the model
       model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

[138]: model.summary()

       Model: "model_2"
```

Figure 3.12: Compiling the Model

The output of compiling the model is shown in **Figure 3.13** and **Figure 3.14**.

```
In [138]:  model.summary()

Model: "model_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_3 (InputLayer)        [(None, 128, 128, 3)]     0

 block1_conv1 (Conv2D)       (None, 128, 128, 64)      1792

 block1_conv2 (Conv2D)       (None, 128, 128, 64)      36928

 block1_pool (MaxPooling2D)  (None, 64, 64, 64)        0

 block2_conv1 (Conv2D)       (None, 64, 64, 128)       73856

 block2_conv2 (Conv2D)       (None, 64, 64, 128)       147584

 block2_pool (MaxPooling2D)  (None, 32, 32, 128)       0

 block3_conv1 (Conv2D)       (None, 32, 32, 256)       295168

 block3_conv2 (Conv2D)       (None, 32, 32, 256)       590080

 block3_conv3 (Conv2D)       (None, 32, 32, 256)       590080

 block3_pool (MaxPooling2D)  (None, 16, 16, 256)       0

 block4_conv1 (Conv2D)       (None, 16, 16, 512)       1180160

 block4_conv2 (Conv2D)       (None, 16, 16, 512)       2359808

 block4_conv3 (Conv2D)       (None, 16, 16, 512)       2359808

 block4_pool (MaxPooling2D)  (None, 8, 8, 512)         0

 block5_conv1 (Conv2D)       (None, 8, 8, 512)         2359808

 block5_conv2 (Conv2D)       (None, 8, 8, 512)         2359808

 block5_conv3 (Conv2D)       (None, 8, 8, 512)         2359808

 block5_pool (MaxPooling2D)  (None, 4, 4, 512)         0

 global_average_pooling2d_2  (None, 512)               0
  (GlobalAveragePooling2D)

 dense_15 (Dense)            (None, 64)                32832

 dense_16 (Dense)            (None, 1)                 65
```

Figure 3.13: Output of Compiling the Model 1

```
=================================================================
Total params: 14747585 (56.26 MB)
Trainable params: 32897 (128.50 KB)
Non-trainable params: 14714688 (56.13 MB)
_________________________________________________________________
```

Figure 3.14: Output of Compiling the Model 2.

Computing weight of each class is then performed as shown in **Figure 3.15**.

```
[139]: from sklearn.utils.class_weight import compute_class_weight

       # Compute class weights
       class_weights = compute_class_weight('balanced', classes=np.unique(y), y=y)

       # Convert class weights to a dictionary
       class_weight_dict = dict(enumerate(class_weights))

[140]: class_weight_dict

[140]: {0: 0.6388888888888888, 1: 2.3}

[141]: import numpy as np

       # Assuming X_train and X_test are your grayscale datasets
       X_train_rgb = np.repeat(X_train, 3, -1)

       X_test_rgb = np.repeat(X_test, 3, -1)

[142]: X_train_rgb.shape

[142]: (202, 128, 128, 3)

[147]: # define the checkpoint
       from tensorflow.keras.callbacks import ModelCheckpoint
       filepath = "weights.best.hdf5"
       checkpoint = ModelCheckpoint(filepath, monitor='val_accuracy', verbose=1, save_best_only=True, mode='max')
```

Figure 3.15: Computing Weight of Each Class

The model is then retrained as shown in **Figure 3.16** (sample of model).

```
26/26 [==============================] - ETA: 0s - loss: 0.6689 - accuracy: 0.6337

Epoch 1: val_accuracy improved from -inf to 0.80392, saving model to weights.best.hdf5

26/26 [==============================] - 12s 450ms/step - loss: 0.6689 - accuracy: 0.6337 - val_loss: 0.5884 - val_accuracy: 0.8039

Epoch 2/10
C:\Users\husse\AppData\Roaming\Python\Python310\site-packages\keras\src\engine\training.py:3000: UserWarning: You are saving your mo
egacy. We recommend using instead the native Keras format, e.g. `model.save('my_model.keras')`.

  saving_api.save_model(
26/26 [==============================] - ETA: 0s - loss: 0.5994 - accuracy: 0.7822

Epoch 2: val_accuracy did not improve from 0.80392

26/26 [==============================] - 13s 497ms/step - loss: 0.5994 - accuracy: 0.7822 - val_loss: 0.7465 - val_accuracy: 0.2941

Epoch 3/10

26/26 [==============================] - ETA: 0s - loss: 0.5628 - accuracy: 0.7178

Epoch 3: val_accuracy did not improve from 0.80392

26/26 [==============================] - 14s 537ms/step - loss: 0.5628 - accuracy: 0.7178 - val_loss: 0.5123 - val_accuracy: 0.7843
```

Figure 3.16: Retraining the Model

The next chapter provides the validation of the model and the results, followed by a discussion section.

4. RESULTS AND DISCUSSION

4.1. Model Validation

This chapter presents the validation of the model to prove its efficiency and accuracy in detecting Spinal disc herniation. **Figure 4.1** shows the training and validation accuracy and loss at each epoch.

```python
# Plot the training and validation accuracy and loss at each epoch
def plot_training_history(history):
    fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
    ax1.plot(history.history['accuracy'])
    ax1.plot(history.history['val_accuracy'])
    ax1.set_title('Model Accuracy')
    ax1.set_xlabel('Epoch')
    ax1.set_ylabel('Accuracy')
    ax1.legend(['Train', 'Test'], loc='upper left')

    ax2.plot(history.history['loss'])
    ax2.plot(history.history['val_loss'])
    ax2.set_title('Model Loss')
    ax2.set_xlabel('Epoch')
    ax2.set_ylabel('Loss')
    ax2.legend(['Train', 'Test'], loc='upper left')

    plt.tight_layout()
    plt.savefig("accuracy_loss_curve.jpg", format='jpg')

    plt.show()


plot_training_history(history)
```

Figure 4.1: The Training and Validation Accuracy and Loss At Each Epoch

After running the model over the training dataset, the model accuracy at each epoch was shown in **Figure 4.2**, and the model loss at each epoch shown in **Figure 4.3**.

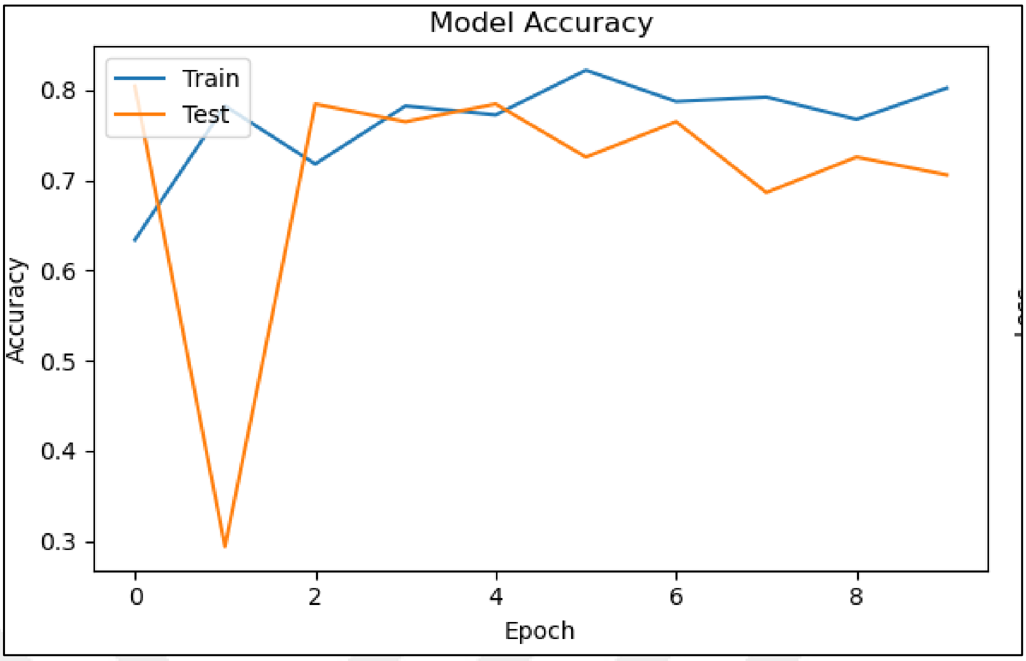


Figure 4.2: Model Accuracy at Each Epoch

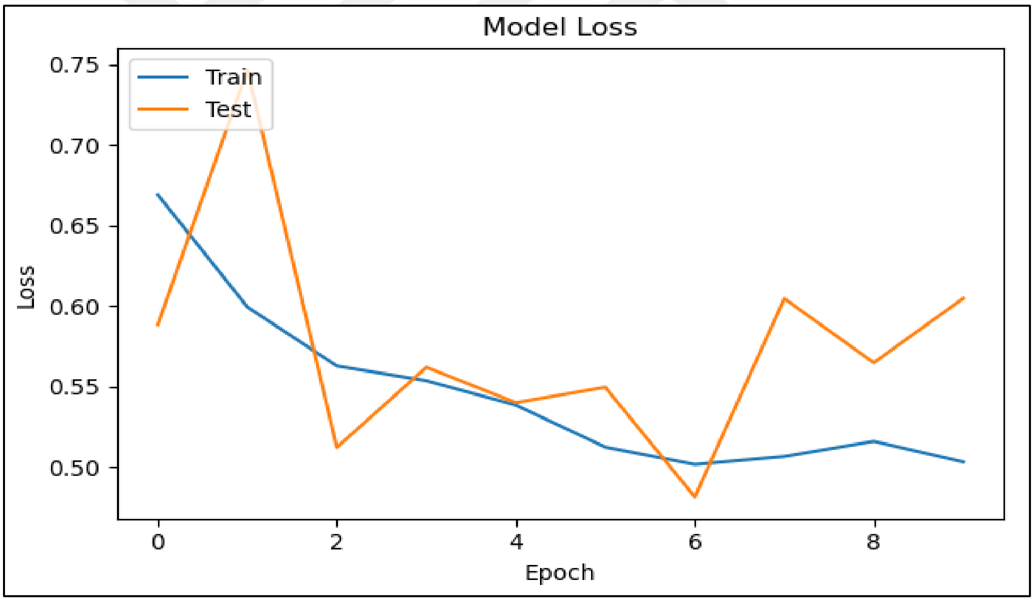


Figure 4.3: Model Loss at Each Epoch

Figure 4.4 shows how the model is evaluated to show accuracy and loss.

```
# Load weights
model.load_weights("weights.best.hdf5")
```

```
# 4. Evaluation

loss, accuracy = model.evaluate(X_test_rgb, y_test)
print(f"Accuracy: {accuracy*100:.2f}%")
```

```
2/2 [==============================] - 0s 32ms/step - loss: 0.6903 - accuracy: 0.8039

Accuracy: 80.39%
```

Figure 4.4: Evaluation of Model Accuracy and Loss

Predictions for the test set is then calculated and samples are released as shown in **Figure 4.5**.

```
# Get predictions for the test set
y_pred = model.predict(X_test_rgb)
predicted_labels = [1 if pred > 0.5 else 0 for pred in y_pred]
true_labels = y_test

# Decode the labels
decoded_predicted_labels = le.inverse_transform(predicted_labels)
decoded_true_labels = le.inverse_transform(true_labels)

# Display some samples from the test set
num_samples_to_show = 4
fig, axes = plt.subplots(nrows=num_samples_to_show, figsize=(5, 5*num_samples_to_show))

for i, ax in enumerate(axes):
    ax.imshow(X_test[i].reshape(128, 128), cmap='gray')
    ax.set_title(f"Predicted: {decoded_predicted_labels[i]}\nTrue: {decoded_true_labels[i]}")
    ax.axis('off')

plt.tight_layout()
plt.show()
```

Figure 4.5: Predictions for the Test Set

The minor predicted is shown in **Figure 4.6**.

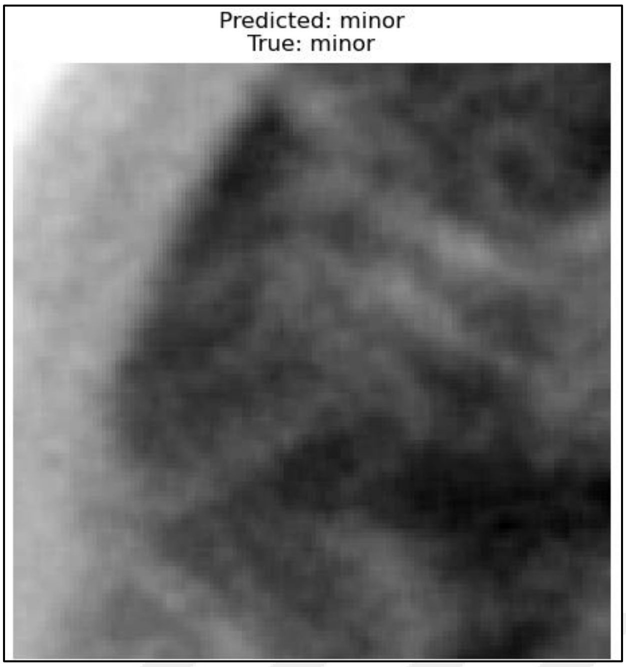


Figure 4.6: The Predicted Minor

The Osteophytes predicted are shown in **Figure 4.7**.

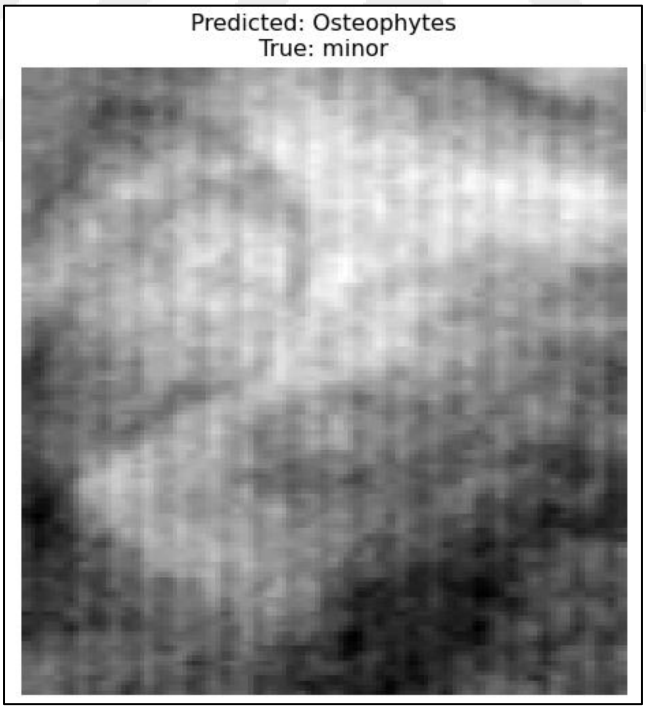


Figure 4.7: The predicted Osteophytes

The predicted True Osteophytes are shown in **Figure 4.8** and **Figure 4.9**.

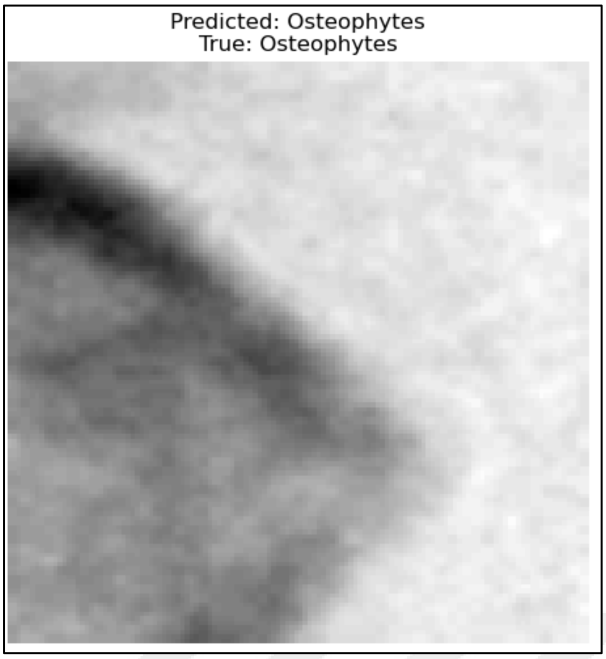


Figure 4.8: Predicted True Osteophytes 1

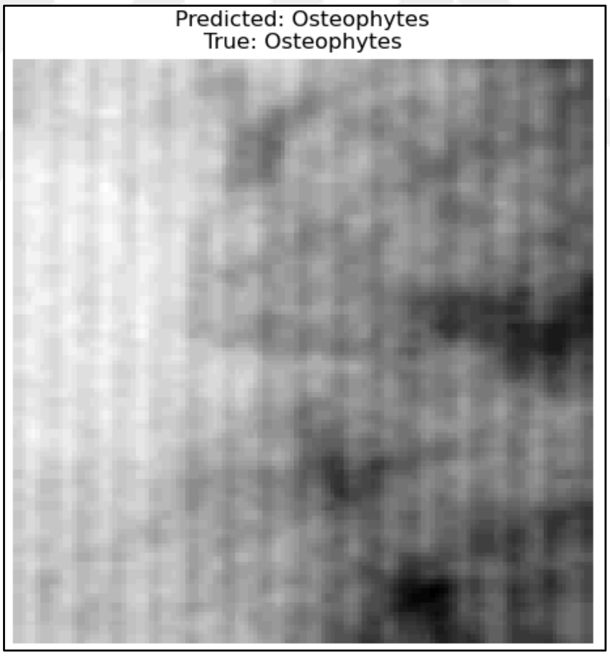


Figure 4.9: Predicted True Osteophytes 2

The classification report is then generated as shown **Figure 4.10**.

```
In [151]:  from sklearn.metrics import classification_report

           # Get predictions for the test set
           y_pred = model.predict(X_test_rgb)

           # # Convert one-hot encoded predictions to label-encoded predictions
           predicted_labels = [ 1 if pred > 0.5 else 0 for pred in y_pred]

           # # Convert one-hot encoded true labels to label-encoded true labels
           # true_labels = np.argmax(y_test, axis=1)

           # Generate and print the classification report
           # Make sure 'le.classes_' corresponds to the correct order of your class labels
           class_report = classification_report(y_test, predicted_labels, target_names=le.classes_, labels=np.arange(len(le.classes_)))
           print(class_report)

           2/2 [==============================] - 2s 606ms/step
                         precision    recall  f1-score   support

            Osteophytes       0.84      0.78      0.81        40
                  minor       0.36      0.45      0.40        11

               accuracy                           0.71        51
              macro avg       0.60      0.61      0.60        51
           weighted avg       0.73      0.71      0.72        51
```

Figure 4.10: Classification Report

The confusion matrix is then generated as shown in **Figure 4.11**.

```
In [152]:  # confusion matrix
           from sklearn.metrics import confusion_matrix
           import seaborn as sns

           # Generate the confusion matrix
           cm = confusion_matrix(y_test, predicted_labels)

           # Plot the confusion matrix
           plt.figure(figsize=(8, 8))
           sns.heatmap(cm, annot=True, cmap='Blues', fmt='g', cbar=False)
           plt.xticks(np.arange(len(le.classes_)) + 0.5, le.classes_, rotation=90)
           plt.yticks(np.arange(len(le.classes_)) + 0.5, le.classes_, rotation=0)
           plt.xlabel("Predicted")
           plt.ylabel("Actual")
           plt.title("Confusion Matrix")
           plt.savefig("confusion_matrix.jpg", format='jpg')
           plt.show()
```

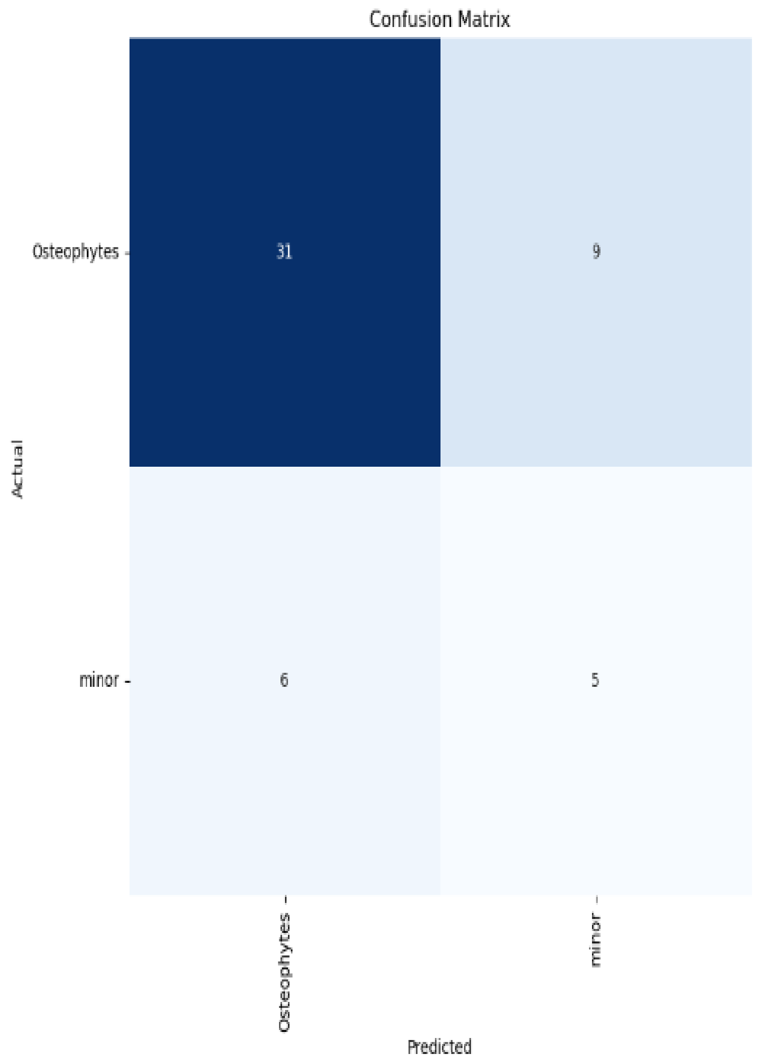


Figure 4.11: Confusion Matrix

Accordingly, counting the frequency of each label is calculated as shown in **Figure 4.12** and **Figure 4.13**.

```
In [ ]:
import numpy as np
import matplotlib.pyplot as plt

# Convert y back to original labels
y_labels = le.inverse_transform(y)

# Convert labels to integers
label_integers = le.transform(y_labels)

# Count the frequency of each label
label_counts = np.bincount(label_integers)

# Get unique labels
unique_labels = np.unique(label_integers)

# Plot the frequencies
plt.figure(figsize=(10, 6))
plt.bar(unique_labels, label_counts)
plt.xlabel('Classes')
plt.ylabel('Count')
plt.title('Count of classes in y')
plt.show()
```

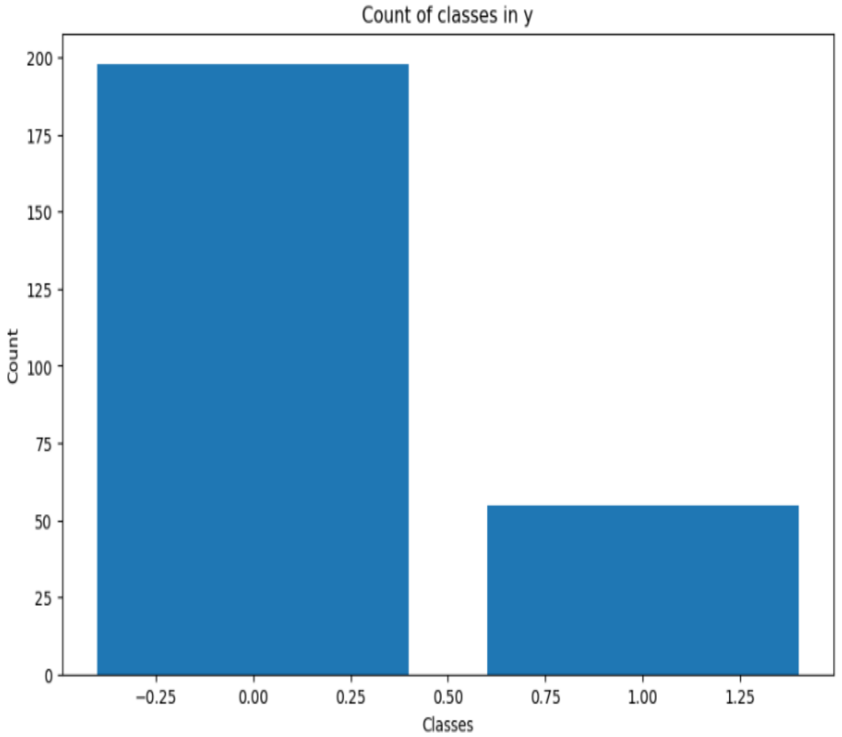


Figure 4.12: Counting the Frequency of Each Label 1

```
# Print the count of each class
for label, count in zip(unique_labels, label_counts):
    print(f'Class {le.inverse_transform([label])[0]}: {count}')

Class Osteophytes: 198

Class minor: 55
```

Figure 4.13: Counting the Frequency of Each Label 2

4.2 Model Evaluation

To evaluate the model efficiency, a survey was conducted among a group of practitioners in the medical field. This survey aims at proving the effectiveness of using deep learning techniques in detecting Spinal Disc Herniation. The survey investigates the importance of using automated systems in improving patient outcomes, which in turn help doctors and radiologists provide enhanced service delivery. The survey investigates how practitioners view data augmentation techniques as valuable ones to enhance the generalizability and robustness of the deep learning system for detecting and classifying spinal disc herniation.

A questionnaire was designed that consists of eight questions and was presented as a google form to practitioners to answer. These questions focus first on the importance of automating detection and classification of Spinal Disc Herniation, and then focus on using deep learning techniques on MRI images.

The questionnaire composed of the following questions:

1- Can a deep learning framework detect the existence of a herniated spinal disc in MRI images? (Yes/No)

2. Is it possible for data augmentation techniques to improve the robustness and generalizability of the deep learning system used to identify and classify spinal disc herniation? (Yes/No)

3. Can the automated system assist radiologists in the diagnosis and treatment planning of spinal disc herniation, leading to improved patient outcomes? (Yes/No)

4- What deep learning architectures, such as (CNNs) or (RNNs), are most suitable for automated detection and classification of spinal disc herniation?

5. How can the deep learning system handle the variability in imaging protocols and artifacts commonly present in MRI images of spinal disc herniation?

6. To what extent does the performance of the automated system depend on the size and quality of the annotated dataset used for training?

7. How does the proposed deep learning-based approach compare to traditional computer-aided diagnosis methods in terms of accuracy, efficiency, and consistency? deep learning more accuracy and efficiency, or deep learning less accuracy and efficiency, or No difference.

8. What are the potential limitations and challenges in implementing the deep learning system in clinical practice?

Eleven practitioners answered the questionnaires and the received data was analyzed to evaluate the model.

Question 1:

Can a deep learning framework detect the existence of a herniated spinal disc in MRI images? (Yes/No)

All practitioners (100%) answered Yes; this reveals that deep learning technology is nowadays likely to be an increasingly crucial element for patients care and improving clinical decision-making, speeding up medical research, and improving patient outcomes. Practitioners argued that deep learning algorithms examine a lot of patient data, including images to increase the precision of diagnoses and improve treatment strategies. They said that due to the high workload, doctors can be affected by stress and fatigue from interpreting complex biomedical images. In addition, they provided that some reports indicated the advantage of the existing computing assistant medical diagnosis and treatment from the artificial neural network (ANN) to deep learning. Deep learning models may be trained to recognize images with high speed and accuracy. Besides, practitioners said that deep learning frameworks have very good chance to shown significant promise in detecting the existence of herniated spinal discs in MRI images. Specifically, (CNNs) and other deep learning architectures surpass in image recognition tasks, and are highly likely to be well-suited for medical image analysis, including identifying spinal disc herniations.

Question 2:

Can data augmentation techniques enhance the generalizability and robustness of the deep learning system for detecting and classifying spinal disc herniation? (Yes/No)

All practitioners (100%) answered Yes; data augmentation techniques play a crucial role in improving the generalizability and robustness of deep learning systems, especially in medical image analysis like detecting and classifying spinal disc herniation. Practitioners revealed they almost certain that data augmentation allows radiologists to create variations of existing images by applying scaling, translation, or changes in brightness. By augmenting the dataset, models are presented to a wider range of variations and scenarios, which help in recognizing different forms of spinal disc herniation. Medical images may have noise, and unclear resolutions. Augmentation techniques resolve these variations to overcome inconsistencies in the input data. These techniques learn to focus on the essential features for classification despite such variations. Augmented data captures different aspects of the same pathology, and provide a more comprehensive understanding to the model. Data augmentation methods enable the generation of additional training samples. With a larger and more varied dataset, almost certain the model can better understand the distinctions and variations present in spinal disc herniation. Models learns to recognize relevant features of spinal disc herniation across different orientations. Practitioners provided that augmentation techniques highly likely can help mitigate dataset biases by creating balanced representations of different classes of spinal disc herniation.

Question 3:

Can the automated system assist radiologists in the diagnosis and treatment planning of spinal disc herniation, leading to improved patient outcomes? (Yes/No)

82 % of practitioners provided “Yes”, that an automated system can assist radiologists in the diagnosis and treatment planning of spinal disc herniation, leading to very good chance for improved patient outcomes. They said that automated systems, such as deep learning models, can assist radiologists by rapidly analyzing potential areas of interest in MRI images, which almost certain speeds up the diagnostic process. This allows radiologists to have more focus on analysis rather

than spending extensive time on routine examinations. Automated systems can accurately detect and highlight abnormalities associated with spinal disc herniation, assisting radiologists in identifying subtle or complex cases. This might leads to earlier detection and intervention. Automated systems can provide quantitative data about the herniation's size, location, and severity, which help in creating treatment plans. These systems also help in monitoring the progression or regression of the herniation over time. Automated systems also enable remote consultations and second opinions from specialists worldwide.

18% of practitioners answered “no”; They point out that automated systems are unlikely to help radiologists diagnose and plan treatment for herniated discs. They doubted that these systems increase the capabilities of radiologists. They added that the experience of interpreting results, making the final diagnosis, developing treatment plans, and clinical judgment of radiologists remains crucial. Ethical considerations regarding patient data privacy and informed consent are essential in the development and deployment of these automated systems.

In addition, the quality and diversity of training data is an essential factor for the success of automated systems. System performance will be compromised if the training data set is biased, incomplete, or does not represent the full range of disc herniation variations, and will result in inaccurate or unreliable results. Incorrect diagnoses or recommendations may result from inaccurate, biased, or insufficient data. There may be unusual or rare cases that will be difficult for automated systems to diagnose or treat. When automated systems make important judgments on their own without human input, ethical questions will arise, including critical issues of informed consent, accountability, and responsibility. There is more than data-driven analysis, but it is the empathy and precise communication by the human element to establish productive patient-physician relationships that are essential in medicine. Complexity may arise when integrating automated technology with existing healthcare procedures and infrastructure. Implementation may be hampered by interoperability and compatibility issues with other systems.

Deep learning models, especially complex neural networks, are often difficult to explain. There may be difficulty in understanding and explaining the decisions made by these models by radiologists, and distrust and reluctance to rely solely on automated results will result. Radiologists relying excessively on automated systems may lead to automation bias where they trust the system's output without critical evaluation. This overreliance can lead to misdiagnosis and incorrect treatment decisions if the system makes an error.

Automated systems are useful for image analysis, but strict rules and regulations with automated systems and medical equipment must be followed. Navigating various regulatory frameworks and ensuring compliance is likely to be difficult for developers and healthcare providers. The field of medicine is developing rapidly. Frequent maintenance and monitoring in order to integrate new information into automated systems is a basic requirement to keep pace with this development, which requires continuous upgrades. There must be collective action by healthcare professionals, legislators, ethicists, and technology specialists to address these issues and ensure that automated systems are designed, used, and managed in a way that optimizes their benefits while minimizing any risks.

Question 4

- *Which deep learning architectures are most suitable for automated detection and classification of spinal disc herniation, such as (CNNs) or (RNNs)?*

About 65% of practitioners reported that CNNs are commonly used for medical image analysis tasks, including detection and classification of spinal disc herniations. (CNNs) have demonstrated remarkable effectiveness in detection and classification tasks related to automated medical imaging, such as spinal disc herniation. Their specialty lies in automatic image recognition and classification using hierarchical feature learning capabilities.

It is very likely that CNNs will be able to identify anomalies in spinal discs by analyzing medical imaging data, such as MRI or CT scans, for the specific purpose of identifying and classifying herniated discs. These networks have the ability to identify patterns and characteristics that indicate a hernia, which can help doctors or radiologists diagnose and treat patients. The preferred architecture is often CNNs due to their efficiency in handling image data. When identifying patterns and structures, CNNs can effectively learn hierarchical features of medical images.

Practitioners argued that CNNs are commonly preferred due to their efficiency in image recognition. CNNs are specifically designed for processing visual data and good at learning hierarchical features from images, making them highly suitable for detecting spinal disc herniation.

Standard CNN architectures with convolutional layers followed by pooling layers have been used for medical image analysis, including spinal disc herniation detection. Models like U-Net, Densely Connected Convolutional Networks (DenseNet), are tailored to the specific characteristics of medical images. Utilizing pretrained models like VGG, ResNet, which were originally trained on large-scale image datasets like ImageNet, has shown promise in medical imaging tasks. Since MRI images are three-dimensional, 3D CNN architectures capture spatial information across multiple slices of MRI volumes. U-Net: is recognized for its application in biomedical image segmentation tasks, U-Net has demonstrated potential in spinal disc segmentation and herniation identification by differentiating aberrant regions in pictures. Residual Networks (ResNets) are these models, especially more deep varieties like ResNet-50 or ResNet-101, can catch many-sided subtleties inside clinical pictures, possibly helping with exact recognition of disc herniation.

Using DenseNet, by associating each layer to each and every other layer, DenseNet encourages include reuse, improving the organization's capacity to learn perplexing examples inside the pictures.

VGG (Visual Geometry Group) while not quite so perplexing as some fresher models, VGG networks with their clear engineering have demonstrated successful in clinical picture arrangement errands because of their deep layering and highlight extraction abilities.

Conversely, recurrent neural networks, or RNNs, are more frequently employed for sequential data, including time series or jobs involving natural language interpretation. RNNs can be used in situations where sequential data pertaining to the evolution or changes in spinal conditions over time need to be analyzed, such as tracking patient histories or monitoring the progression of herniations. However, they might not be the best option for analyzing static images, such as those from medical scans. RNNs might not be the primary choice due to the nature of static images in MRI scans.

Because CNNs are so good at capturing spatial relationships in pictures, they continue to be the go-to option for automated detection and classification tasks in medical imaging. Systems such as U-Net, DenseNet, or specially-crafted CNNs with particular modifications for medical image processing have shown effective in recognizing spinal disc herniation and related conditions.

15% provided that they still do not know what deep learning architectures that exist for automated detection and classification of spinal disc herniation.

20% of practitioners said that it's important to note that the choice of architecture depends on factors such as the nature of the data, available computational resources, and the specific requirements of the task. Since sufficient and well-curated labeled datasets are essential for training these deep learning models effectively.

Question 5:

5. How can the deep learning system handle the variability in imaging protocols and artifacts commonly present in MRI images of spinal disc herniation?

55% provide that deep learning systems can benefit from robust preprocessing techniques and data augmentation. Preprocessing methods, such as normalization and standardization, can enhance model generalization across different imaging protocols. Data augmentation involves applying random transformations to the training data, creating variations that help the model become more resilient to artifacts.

Additionally, transfer learning, using pre-trained models on large datasets, can be effective. By leveraging knowledge gained from diverse datasets, the model becomes more adaptable to different imaging conditions. A regularization technique that enhances the model's ability to handle variation and prevents overfitting is dropout. Data augmentation, normalization techniques, transfer learning, image processing, attention mechanisms, and ensemble models are more efficient.

Practitioners have explained that standardizing image intensity, resolution, and orientation using preprocessing and normalization techniques before adding images to a deep learning model, is one way to reduce variability caused by different imaging processes. At this stage you make sure that all the images are consistent.

To ensure that the deep learning model ignores artifacts and focuses on data relevant to diagnosing spinal disc herniation, it may be possible to identify and remove artifacts from MRI images using an additional model or preprocessing operations. Combining data from other imaging methods, such as (combining CT scan, MRI, or other imaging methods) may help the model better deal with variation and increase the accuracy of its diagnosis of spinal disc herniation. Designing architectures that are resistant to changes and artifacts, modifications to the architecture or attention mechanisms that focus on informational domains is necessary.

15% reported that it is important to emphasize that variation in the effectiveness of these deep learning systems may depend on the specific challenges posed by the imaging data. It is very important to develop a robust deep learning system for disc herniation detection to try out different approaches and comprehensively evaluate the model's performance on diverse data sets.

10% reported that the focus is on mitigating the impact of imaging protocol variations and artifacts through these methods, but having a diverse and representative data set covering a wide range of imaging protocols and artifacts is essential. There is an important role for the quality and diversity of training data in the performance and dissemination of the deep learning system.

20% reported that it is very important for the power and generalizability of deep learning systems in detecting intervertebral disc herniation, the presence of artifacts in MRI images, and variability in imaging protocols.

Question 6:

6To what degree does the performance of an automated system depend on the quality and size of the annotated data set used for training?

100% of practitioners reported that the performance of the automated system was highly dependent on the quality and size of the annotated dataset used for training. A model with a larger and more diverse data set will be able to learn a wider range of patterns and differences, which will make it better able to generalize to new and unseen data. Deep learning models will have the ability to capture a more diverse set of patterns across a larger dataset if available, thus improving performance and better generalization.

Quality is crucial. An example of this is fine-grained annotations that make the model learn meaningful features. In medical imaging, where accuracy is vital, system reliability and diagnostic accuracy depend on a high-quality annotated data set. One of the most important reasons for the model to be misleading is inaccurate or inconsistent annotations, which affects its ability to generalize.

Consistency in how annotations are named is crucial across the dataset. A smaller dataset with high-quality annotations may provide better performance than a larger dataset with poor-quality annotations. It is necessary to strike a balance between dataset quality and size, because a large but poorly annotated dataset may not lead to optimal performance.

The quality and size of the annotated data set used for training has a significant impact on the performance and generalizability of automated systems, especially in medical image analysis tasks such as detecting herniated discs. Larger datasets generally lead to better model performance up to a certain point but accurate and reliable annotation is crucial. Errors or inconsistencies in annotations can mislead the model and impact its ability to learn the correct features. High-quality annotations ensure the model learns from reliable ground truth, improving its performance.

A larger dataset allows machine learning models to learn from diverse examples, potentially improving their performance which is reflected in better understanding. Improved Generalization: More data helps models generalize better, making predictions more accurately on new, unseen data.

The advantages of more data may, however, eventually run out, and the expenses associated with handling and processing huge datasets may surpass the gains in performance. Precise annotations are essential. Annotations of the highest caliber are crucial. Annotations that are inconsistent or inaccurate

Annotations that are consistent throughout the dataset aid the model's comprehension of linkages and patterns. To make sure the model learns patterns that reflect real-world use cases, high-quality datasets should ideally depict real-world scenarios. Predictions that are prejudiced due to biases in the dataset may affect the automated system's fairness and dependability.

When a dataset is adequately sized and has high-quality annotations, it can provide models that are more resilient, flexible in a variety of situations, and able to predict new, unobserved data with greater accuracy.

Question 7

How does the proposed deep learning-based approach compare to traditional computer-aided diagnosis methods in terms of accuracy, efficiency, and consistency? Deep learning more accuracy and efficiency, or deep learning less accuracy and efficiency, or No difference.

91% of practitioners said that the proposed deep learning-based approach gives more accuracy and efficiency compared to traditional computer-aided diagnosis methods. Accuracy suggests that deep learning algorithms can automatically extract complex features from unprocessed data, doing away with the requirement for manually created features in conventional techniques. As a result, complex patterns and relationships within the data are frequently represented with greater accuracy. Deep learning architectures are capable of learning hierarchical representations, collecting subtle information at several levels of abstraction. Examples of these architectures are (CNNs) for images and (RNNs) for sequential data. Efficiency claims that deep learning models are capable of end-to-end learning, which means that they don't need manually designed pipelines to handle raw data and produce predictions right away. This might potentially lead to increased efficiency. Deep learning models may be able to scale efficiently with those numbers, whereas conventional methods may stall or perform worse when dealing with larger amounts of data.

Deep learning models can be effective at prediction once they have been trained, especially for tasks like speech or image recognition, where they have demonstrated remarkable speed and accuracy. Preparing deep learning models can be computationally costly, requiring significant computational assets and time, especially with huge datasets. With smaller datasets and simpler algorithms, traditional methods may be more computationally efficient.

Nevertheless, their effectiveness could come at the expense of exactness, particularly in situations where the information is profoundly complicated or the connections between highlights are nonlinear. 9% argued that deep learning approaches can have less accuracy and efficiency than traditional computer-aided diagnosis methods. This is due to some challenges in which large volumes of labeled data, which are frequently needed for deep learning model training, may not always be available, particularly in the medical field where labeled data can be hard to come by and expensive to acquire.

Furthermore, deep learning models are sometimes referred to as "black boxes," making it difficult to comprehend the logic underlying their predictions. This could be problematic in vital applications where interpretability is essential, such as medical diagnosis. When there isn't enough data available for training or when the data used to train deep learning models doesn't adequately represent the many different scenarios that are found in real-world applications, these models may have trouble. Customary strategies could perform well with more modest datasets and in situations where the connections between highlights are surely known. These models may lack the capacity to handle complex, high-dimensional data and might not adapt as readily to new, unseen patterns.

Question 8:

What are the potential limitations and challenges in implementing the deep learning system in clinical practice?

Challenges provided by practitioners varied. Some were focused on Data privacy and security, interpretability and limited labeled data. Medical data is highly sensitive and subject to strict privacy regulations. Ensuring compliance with these regulations while using patient data for training can be complex and limit the amount of available data. Ensuring data security at every stage of the procedure is essential, ranging from gathering and archiving to training and deriving models. Gaining the confidence of healthcare professionals in a model's predictions requires them to comprehend the logic behind it. It is imperative to devise methods for enhancing the interpretability of deep learning models while maintaining their integrity. The quantity of medical datasets is frequently constrained by issues with privacy and the expense and labor of annotating the data. In the healthcare industry, where labeled data may be hard to come by, it is imperative to investigate methods such as semi-supervised or transfer learning that take advantage of both small amounts of labeled data and larger volumes of unlabeled data to enhance model performance.

Other challenges were that deep learning models frequently function as intricate "black boxes," making it difficult to understand the decisions they make. In clinical settings, it's critical to comprehend and communicate the reasoning behind a diagnosis. Also, the caliber and representativeness of the training data have a major impact on how well deep learning systems function. Model projections may differ depending on data biases, such as institutional or demographic biases. Compliance with healthcare rules, and regulatory requirements are difficult to follow. It is necessary to handle data security and privacy issues. Certain patient groups may not be well-suited for models that were trained on particular populations. The efficacy of the model may be affected by variations in genetics, healthcare practices, and demographics. It is important to provide serious consideration to ethical issues like informed permission, openness regarding AI support for clinical decision-making, and responsible AI use. Deep learning models need constant observation for declines in performance and adjustment to changing.

Preparing and sending deep learning models can request huge computational assets. The accessibility of such assets can be a limit, particularly for more modest medical services foundations. Thorough approval through clinical preliminaries is important to lay out the wellbeing and adequacy of the deep learning framework before far and wide clinical reception. Need a genuine dataset and Dependability of the model against genuine climate factors. Low occasions of dataset, low component of datasets. Low occasions of dataset, low component of datasets are fundamental difficulties of deep learning procedures. When using these datasets for training and inference, it is essential to ensure the privacy of patient data and adhere to regulations. The model's performance and generalizability can be affected by biases and inaccuracies in the training data, such as demographic biases or inconsistent annotations. Deep learning systems frequently necessitate significant infrastructure and computational resources. Absence of adequate data and information is a huge test. To ensure the successful integration of AI technologies into clinical practice, developers, healthcare professionals, and regulatory bodies must continuously collaborate. This iterative and thorough methodology adds to the improvement of strong and dependable instruments that can emphatically affect patient consideration and results.

Other mentioned challenges were Interpretability, Data Quality and Bias, Generalization to New Data, Regulatory Compliance, Integration with Clinical Workflow, Ethical and Legal Concerns, Resource Intensity, Continual Model Updating, Clinician Adoption.

Important obstacles to the application of AI systems in healthcare include those inherent in the field of machine learning, practical implementation hurdles, and taking into account acceptance barriers and required modifications to social norms or workflows. The gold standard for generating evidence should be considered to be robust, peer-reviewed clinical evaluation conducted as part of randomized controlled trials. However, in practice, this may not always be possible or acceptable. Real clinical applicability should be captured by performance measures, and they should be comprehensible to the intended audience.

To sum up from the previous analysis, the proposed system has made a deep learning framework especially custom fitted for the computerized ID and grouping of spinal disc herniation utilizing MRI filters.

There are a number of obstacles in the way of creating an efficient and dependable deep learning system for identifying spinal disc herniations. The intricacy of the spine's physical design, the unconventionality of imaging techniques, and strong algorithms fit for managing the event of relics in MRI picture. Besides, the shortage of enormous, clarified datasets represents one more snag for preparing deep learning models really.

The proposed approach in this research included gathering a thorough dataset of named MRI images from patients with affirmed spinal disc herniation, alongside solid control subjects. The proposed system used cutting edge, deep learning engineering, deep learning models like (CNNs), were investigated and chosen in light of the intricacy of the undertaking and accessible computational assets. The proposed system focused on techniques, such as, expanding information, move learning, and collecting to work on the generalizability and versatility of the proposed system. The viable reception of such a system can possibly work on demonstrative cycle, treatment arranging, and patient results.

Practitioners added that the proposed system succeeded in being a valuable deep learning system as it collected a comprehensive dataset of labeled MRI images from patients with confirmed disc herniation.

4.3 Model Performance

The model's performance, as depicted in the accuracy and loss graphs, indicates a learning trend over epochs. The training and validation accuracy graphs show a convergence trend, albeit with some fluctuations that suggest potential overfitting or instability in learning. The accuracy achieved on the test data is satisfactory, though there is room for improvement. Similarly, the loss graphs show a downward trend, albeit with fluctuations, indicating that the model is learning but may benefit from regularization or other techniques to stabilize training.

The confusion matrix provides insight into the model's classification capabilities, with a significant number of true positives for 'Osteophytes' but a relatively high number of false negatives and false positives for 'minor' lesions. This disparity is further reflected in the precision, recall, and f1-score metrics, where the model shows high precision but lower recall for 'Osteophytes,' and low scores across all metrics for 'minor' lesions. The class imbalance, evident from the count of classes, might contribute to the model's bias toward the majority class.

4.4 Discussion

The classification report reveals a stark contrast between the performance of the two classes. Model precision and recall are shown in **Figure 4.14**. The model demonstrates high precision in identifying 'Osteophytes' (0.84) but struggles with 'minor' lesions, as indicated by the low precision (0.36). This is indicative of the model's propensity to incorrectly predict the 'minor' class. The recall for 'Osteophytes' is relatively high (0.78), suggesting that the model can identify most of the positive cases. However, the recall for 'minor' lesions is low (0.45), highlighting the model's challenges in detecting true positives for this class.

	precision	recall	f1-score	support
Osteophytes	0.84	0.78	0.81	40
minor	0.36	0.45	0.40	11
accuracy			0.71	51
macro avg	0.60	0.61	0.60	51
weighted avg	0.73	0.71	0.72	51

Figure 4.14: Model Precision and Recall

The f1-score of Osteophytes is (0.81), showing effectiveness in detection, while minor detection was 0.40, which is low, which is a challenging issue with this type of diagnosis using deep learning models. The overall accuracy of the model stands at 60%, with a weighted average f1-score of 0.72, which could be considered modest.

These metrics suggest that while the model can identify 'Osteophytes' to a certain degree, it is less effective for 'minor' lesions, likely due to class imbalance and possibly insufficient feature representation for the minority class.

The bar plot of class counts further corroborates the imbalance in the dataset, with 'Osteophytes' being the predominant class. This could have led to the model's bias towards predicting 'Osteophytes' more accurately than 'minor' lesions.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The broad development of clinical information expands the utility of AI and deep learning in the medical care spaces. These days, the utilization of inside and out preparing to deal with clinical pictures has gotten specific consideration.

With the help of artificial intelligence, medical instruments have developed rapidly in recent years and are now widely used to process medical images. Man-made reasoning is various wellsprings of clinical imaging handling like X-beam, Computerized Tomography (CT), and MRI. CT and MRI picture handling errands with a high calculation time prerequisite and calculation speed.

These days, quite possibly of the most basic pattern in the advancement of PC innovation in neuroscience is the handling of clinical pictures and computerized pictures, which are utilized to further develop picture quality, reestablish harmed pictures, distinguish individual components and analyze different illnesses.

The research studied strategies such as augmenting data, transfer learning, and assembling to improve the generalizability and resilience of the proposed system. This research created a therapeutically useful tool that helped radiologists in the correct diagnosis and categorization of spinal disc herniation from MRI data. The effective adoption of such a system has the potential to improve diagnostic process, treatment planning, and patient outcomes.

The following questions were addressed in this thesis and were answered as followed:

1. *Can a deep learning system discover the presence of a herniated spinal disc in MRI images?*

Yes, deep learning frameworks have shown great promise in detecting the presence of herniated disc in MRI images. Specifically, CNNs and other deep learning architectures are best suited for image recognition tasks, and well-suited for analyzing medical images, including identifying herniated discs. Deep learning algorithms can examine a lot of patient data, including images, to increase diagnosis accuracy and improve treatment strategies.

2. *What deep learning architectures, such as CNNs or Recurrent Neural Networks RNNs, are most suitable for automated detection and classification of spinal disc herniation?*

Convolutional Neural Networks (CNNs) have been universally employed in medical image analysis for the identification of spinal disc herniations. Convolutional neural networks (CNNs) have shown notable accuracy in both detection and classification of automated medical imaging including spinal disc herniation. They are a great set of tools that you can use to automatically detecting and classify pictures using deep learning hierarchical feature learning functions.

To identify herniated discs, CNNs could analyze medical imaging data like MRI or CT scans (which show cross-sections of body in detail), and search for patterns indicative of disc abnormalities. These networks are able to statistically identify patterns and features that indicate a hernia allowing doctors or radiologists to dialysis the information to diagnose begin treatment of patients who where referred.

RNNs which are common for sequential data: time series, or utilities with natural language meaning. RNNs can be advantageous in spinal studies when researchers deal with sequential data that change or develop along the time, as example of follow-up of a patient history or progression of disk herniation.

3. *How can the deep learning system handle the variability in imaging protocols and artifacts commonly present in MRI images of spinal disc herniation?*

There are indeed powerful preprocessing and data augmentation techniques which could help deep learning systems Therefore, the pre-processing techniques like standardization and normalization improve the generalizability of the model upon multiple imaging protocols. For data augmentation you would take the training data and literally apply random transformations to it making more examples that train the model to better deal with artifacts. Other modeling or preprocessing operations can be utilized to automatically detect and filter out artifacts from MRI images. By doing this, we ensure that the deep learning model will disregard artifacts, and will look into the data pertaining to spinal disc herniation diagnosis. It may be advantageous in cases with more heterogeneity by adding data from other imaging (e.g., MR/CT fusion or more) if available, since it could improve performance because of higher difficultly in modeling reflecting variations segmentation accuracy for disc herniation. Transfer learning, which is pre-training on large data sets can also be used. Using knowledge from various data sets, the model becomes more robust to diverge imaging conditions.

4. *To what extent does the performance of the automated system depend on the size and quality of the annotated dataset used for training?*

The capabilities of an automated system are in many ways directly dictated by the provided training data. From a broader and richer set of examples the model has a chance to learn pocketed patterns, increasing its ability on deriving new, unseen data points. Before we should head towards solving the problem, lets have a good understanding of data - Deep Learning is using large amount of data to learn how to work, and by which it can learn most common usage patterns thus generalize better and perform well. Quality is crucial. Correct annotations guarantee valuable features for the model to learn. Well annotated data in medical imaging, one of the domains where accuracy is critical, can improve system reliability and diagnostic accuracy. Inaccurate or inconsistent annotations can mislead the model, affecting its ability to generalize.

5. *Is it possible for data augmentation techniques to improve the robustness and generalizability of the deep learning system used to identify and classify spinal disc herniation?*

Data augmentation techniques play a crucial role in improving the generalizability and robustness of deep learning systems, especially in medical image analysis like detecting and classifying spinal disc herniation. Practitioners revealed that data augmentation allows radiologists to create variations of existing images by applying scaling, translation, or changes in brightness. By augmenting the dataset, models are presented to a wider range of variations and scenarios, which help in recognizing different forms of spinal disc herniation. Medical images may have noise, and unclear resolutions. Augmentation techniques resolve these variations to overcome inconsistencies in the input data.

6. *How does the proposed deep learning-based approach compare to traditional computer-aided diagnosis methods in terms of accuracy, efficiency, and consistency?*

The proposed deep learning-based approach gives more accuracy and efficiency compared to traditional computer-aided diagnosis methods. Accuracy suggests that deep learning algorithms can automatically extract complex features from unprocessed data, doing away with the requirement for manually created features in conventional techniques. As a result, complex patterns and relationships within the data are frequently represented with greater accuracy. Efficiency claims that deep learning models are capable of end-to-end learning, which means that they don't need manually designed pipelines to handle raw data and produce predictions right away. This might potentially lead to increased efficiency. Deep learning models may be able to scale efficiently with those numbers, whereas conventional methods may stall or perform worse when dealing with larger amounts of data.

7. *Can the automated system assist radiologists in the diagnosis and treatment planning of spinal disc herniation, leading to improved patient outcomes?*

"Yes", that an automated system can assist radiologists in the diagnosis and treatment planning of spinal disc herniation, leading to improved patient outcomes. They said that automated systems, such as deep learning models, can assist radiologists by rapidly analyzing potential areas of interest in MRI images, which speeds up the diagnostic process. Radiologists can concentrate on analysis as opposed to routine examinations. Machine processes for highlighting abnormalities most likely to be associated with spinal disc herniation can help guide radiologists to specific, more subtle or complicated cases. It helps in the early detection and intervention. The use of automated systems can offer quantitative information on the hernia size, location and severity and help in treatment planning. They also can help to monitor how the hernia is doing over time. Automated systems also allow remote consultations and second opinions with specialists anywhere in the world.

There will be some challenges in the way, which would be systems augmenting radiologist beyond his/her capabilities. Table reads are just the first stop in the imaging journey; critical due diligence is still done by radiologists, who must use their experience and clinical judgment to interpret results, arrive at definitive diagnoses, and forge treatment plans. We must be considerate of the ethical implications on patient privacy and informed consent when designing and implementing such automated systems. Medicine often gets the more complicated situations that may not fit into pre-defined algorithms or patterns. In order to practice this, empathy, adequate communication and relatability rather than statistical analysis are all required in medicine. The clinical relations in the hospital are social associations distinct from departmental and ward workgroup organizations, they often need human incorporation or humanitarian approach, that automated system cannot present.

8. *What are the potential limitations and challenges in implementing the deep learning system in clinical practice, and how can they be addressed?*

Medical data is among the strictest in privacy concerns due to its nature standing out as highly sensitive. Enforcing these regulations when using patient data for training can get tricky and limits the data to use Data security must be guaranteed throughout the entire process, from acquisition and storage to training and model inference.

Compliance with healthcare rules and regulatory requirements is difficult to follow. It is necessary to address data security and privacy issues. Certain patient groups may not be well-suited for models that were trained on particular populations. The efficacy of the model may be affected by variations in genetics, healthcare practices, and demographics. It is important to provide serious consideration to ethical issues like informed permission, openness regarding AI support for clinical decision-making, and responsible AI use. Deep learning models need constant observation for declines in performance and adjustment to changing.

So, in this research, a deep learning system was developed for the automated detection and classification of spinal disc herniation based on MRI data was developed as part of this study. This system used deep learning capabilities to identify and classify spinal disc herniation based on MRI data. The exploration thought about the proficiency of the traditional ways to deal with PC supported finding to the deep learning-based technique. In addition, the robotized framework's clinical utility was evaluated by evaluating its capacity to assist radiologists in selecting the appropriate course of treatment and the board for spinal disc herniation.

By applying such deep learning framework into a remedially helpful instrument, radiologists can utilize it to analyze and group spinal disc herniation from MRI information accurately. The reasonable gathering of such a framework might perhaps deal with characteristic connection, treatment orchestrating, and patient outcomes. The study contributed to the development of computer-aided medical image analysis by demonstrating that deep learning methods can improve the diagnosis and treatment of disc herniation in the spine.

This study might perhaps change the area of spinal disc herniation disclosure and add to the advancement of PC upheld assessment of clinical pictures by using the power of significant learning.

5.2 Recommendations

While the proposed model shows promise in identifying 'Osteophytes,' it may require improvements to classify ‘minor’ lesions more effectively. Addressing class imbalance and refining the model's architecture and training process might lead to enhanced performance and a more balanced classification capability.

In order to approach these issues, future work may necessitate some of the following:

Data augmentation: Apply more aggressive data augmentation techniques in the minority class to balance the dataset.

Adjusting the class weight: Adjusting the class weights more effectively while training the model may help it pay more attention to the minority class.

Enhance model architecture: Consider different model architectures (e.g. deeper networks or alternative pre-trained models) for better feature extraction and generalization.

Regularization: Use regularization techniques, like L2 regularization or dropout, to reduce overfitting thus increasing model stability.

Hyperparameter tuning: By tuning the hyperparameters like using the techniques of grid search or random search we may get better performance of model.

Quality of data: Makes sure that you have a solid training sample representing different conditions and categories as it relates to different MRI scans. First-order Baseline marked information are critical for training good models.

Formalize Pre-processing: Formally, pre-process MRI images from the dataset. Deep learning model input data can go through normalization, resizing (both cropping and padding followed by cropping), noise reduction and artifact removal.

Architecture selection: Choose relevant models such as (CNNs) is best in image- based tasks. Design or architect explicit structures for the MRI analysis task, possibly adding division assignments that fit a model like U-Net.

Transfer learning and pre-trained models How could we leverage the power of models trained on big datasets (metadata), eg ImageNet ones, and finetune them on our single MRI dataset? This is usually a more lean approach since it requires less amount of data and computational resources yet deliver great performance.

Approval and evaluation: To ensure generalizability of the model and avoid overfitting, you can use robust validation methods such as cross-validation or separate validation sets.

Interpretability: Deep learning models can sometimes be viewed as "secret items." Implement procedures (such as Grad-CAM) to visualize and understand parts of the MRI images necessary to guide the model.

Updates and retraining: Medical knowledge and MRI technology are always changing. Constantly update models with new information and retrain them to adapt to emerging patterns or changes in analysis.

Joint effort with medical services experts: Work closely with radiologists and medical services experts to ensure that the deep learning framework may complement their skills and work process rather than replace it. Uniting their criticisms in model turn of events and approval.

REFERENCES

[1]**Dulebohn, S.C., Massa, R.N. and Mesfin, F.B**., "Disc herniation." StatPearls. Treasure Island (FL): StatPearls Publishing. Dostupno na: https://www. ncbi. nlm. nih. gov/books/NBK441822/[01.10. 2019.].

[2]**J. S. Yi, J. G. Cha, J. K. Han, and H. J. Kim**, "Imaging of herniated discs of the cervical spine: Inter-modality differences between 64-slice multidetector CT and 1.5-T MRI," *Korean J Radiol*, vol. 16, no. 4, 2015, doi: 10.3348/kjr.2015.16.4.881.

[3]**J. Ker, L. Wang, J. Rao, and T. Lim**, "Deep Learning Applications in Medical Image Analysis," *IEEE Access*, vol. 6, 2017, doi: 10.1109/ACCESS.2017.2788044.

[4]**E. Salehi, H. Yousefi, H. Rashidi, and H. Ghanaatti**, "Automatic diagnosis of disc herniation in two-dimensional MR images with combination of distinct features using machine learning methods," *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*, Apr. 2019, doi: 10.1109/EBBT.2019.8742052.

[5]**G. Litjens** *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, 2017. doi: 10.1016/j.media.2017.07.005.

[6]**J. T. P. D. Hallinan** *et al.*, "Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI," *Radiology*, vol. 300, no. 1, pp. 130–138, Jul. 2021, doi: 10.1148/RADIOL.2021204289.

[7]**B. J. Erickson, P. Korfiatis, T. L. Kline, Z. Akkus, K. Philbrick, and A. D. Weston**, "Deep Learning in Radiology: Does One Size Fit All?," *Journal of the American College of Radiology*, vol. 15, no. 3, 2018, doi: 10.1016/j.jacr.2017.12.027.

[8]**D. Ravi** *et al.*, "Deep Learning for Health Informatics," *IEEE J Biomed Health Inform*, vol. 21, no. 1, pp. 4–21, Jan. 2017, doi: 10.1109/JBHI.2016.2636665.

[9]**J. Peng and Y. Wang**, "Medical image segmentation with limited supervision: A review of deep network models," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3062380.

[10] **E. Goceri**, "Medical image data augmentation: techniques, comparisons and interpretations," *Artif Intell Rev*, pp. 1–45, Mar. 2023, doi: 10.1007/S10462-023-10453-Z/TABLES/9.

[11] **E. Salehi, H. Yousefi, H. Rashidi, and H. Ghanaatti**, "Automatic diagnosis of disc herniation in two-dimensional MR images with combination of distinct features using machine learning methods," *2019 Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science, EBBT 2019*, Apr. 2019, doi: 10.1109/EBBT.2019.8742052.

[12] **T. Šušteršič, V. Ranković, V. Milovanović, V. Kovačević, L. Rasulić, and N. Filipović**, "A Deep Learning Model for Automatic Detection and Classification of Disc Herniation in Magnetic Resonance Images," *IEEE J Biomed Health Inform*, vol. 26, no. 12, pp. 6036–6046, Dec. 2022, doi: 10.1109/JBHI.2022.3209585.

[13] **W. Mbarki, M. Bouchouicha, S. Frizzi, F. Tshibasu, L. Ben Farhat, and M. Sayadi**, "Lumbar spine discs classification based on deep convolutional neural networks using axial view MRI," *Interdiscip Neurosurg*, vol. 22, 2020, doi: 10.1016/j.inat.2020.100837.

[14] **C. Hou** *et al.*, "An MRI image automatic diagnosis model for lumbar disc herniation using semi-supervised learning," *Complex and Intelligent Systems*, 2023, doi: 10.1007/s40747-023-00981-0.

[15] **M. Cheng** *et al.*, "Multi-scale Information Aggregation Network for Spine MRI Image Segmentation∗," 2022. doi: 10.1145/3573428.3573672.

[16] **W. Mbarki, M. Bouchouicha, F. T. Tshienda, E. Moreau, and M. Sayadi**, "Herniated lumbar disc generation and classification using cycle generative adversarial networks on axial view mri," *Electronics (Switzerland)*, vol. 10, no. 8, 2021, doi: 10.3390/electronics10080982.

[17] **A. Al-Kubaisi and N. N. Khamiss**, "A Transfer Learning Approach for Lumbar Spine Disc State Classification," *Electronics 2022, Vol. 11, Page 85*, vol. 11, no. 1, p. 85, Dec. 2021, doi: 10.3390/ELECTRONICS11010085.

[18] **W. Shi** *et al.*, "Attention Gate Based Dual-Pathway Network for Vertebra Segmentation of X-Ray Spine Images," *IEEE J Biomed Health Inform*, vol. 26, no. 8, pp. 3976–3987, Aug. 2022, doi: 10.1109/JBHI.2022.3158968.

[19] **J. Liu** *et al.*, "Federated learning-based vertebral body segmentation," *Eng Appl Artif Intell*, vol. 116, 2022, doi: 10.1016/j.engappai.2022.105451.

[20] **R. M. Amin, N. S. Andrade, and B. J. Neuman**, "Lumbar Disc Herniation," *Curr Rev Musculoskelet Med*, vol. 10, no. 4, pp. 507–516, Feb. 2023, doi: 10.1007/s12178-017-9441-4.

[21] **S. B. Tibrewal, M. J. Pearcy, I. Portek, and J. Spivey**, "A prospective study of lumbar spinal movements before and after discectomy using biplanar radiography: Correlation of clinical and radiographic findings," *Spine (Phila Pa 1976)*, vol. 10, no. 5, 1985, doi: 10.1097/00007632-198506000-00010.

[22] **M. A. Reicher1, R. H. Gold1, V. V Halbach1, W. Rauschning1', G. H. Wilson1, and R. B. Lufkin1**, "ARRS Executive Council Award Paper MR Imaging of the Lumbar Spine: Anatomic Correlations and the Effects of Technical Variations", Accessed: Jul. 21, 2023. [Online]. Available: www.ajronline.org

[23] **D. L. Kent, D. R. Haynor, E. B. Larson, and R. A. Deyo**, "Diagnosis of lumbar spinal stenosis in adults: A metaanalysis of the accuracy of CT, MR, and myelography," *American Journal of Roentgenology*, vol. 158, no. 5. 1992. doi: 10.2214/ajr.158.5.1533084.

[24] **O. Sortland, B. Mabnaes, and T. Hauge**, "Functional myelography with metrizamide in the diagnosis of lumbar spinal stenosis," *Acta Radiol Suppl (Stockholm)*, vol. NO.355, 1977.

[25] **R. Derby, S. H. Lee, and B. J. Kim**, "Discography," *Interventional Spine E-Book: An Algorithmic Approach*, pp. 291–302, Jan. 2008, doi: 10.1016/B978-0-7216-2872-1.50030-3.

[26] **G. U. Kim, W. T. Park, M. C. Chang, and G. W. Lee**, "Diagnostic Technology for Spine Pathology," *Asian Spine J*, vol. 16, no. 5, 2022, doi: 10.31616/asj.2022.0374.

[27] **R. R. Edelman**, "The history of MR imaging as seen through the pages of radiology," *Radiology*, vol. 273, no. 2. 2014. doi: 10.1148/radiol.14140706.

[28] **K. G. Hartley, B. M. Damon, G. T. Patterson, J. H. Long, and G. E. Holt**, "MRI techniques: A review and update for the orthopaedic surgeon," *Journal of the American Academy of Orthopaedic Surgeons*, vol. 20, no. 12. 2012. doi: 10.5435/JAAOS-20-12-775.

[29] **M. T. Modic** *et al.*, "Acute low back pain and radiculopathy: MR imaging findings and their prognostic role and effect on outcome," *Radiology*, vol. 237, no. 2, 2005, doi: 10.1148/radiol.2372041509.

[30] **A. Manoliu** *et al.*, "Diffusion Tensor Imaging of Lumbar Nerve Roots: Comparison Between Fast Readout-Segmented and Selective-Excitation Acquisitions," *Invest Radiol*, vol. 51, no. 8, pp. 499–504, Aug. 2016, doi: 10.1097/RLI.0000000000000260.

[31] **R. M. Byrne, A. K. Aiyangar, and X. Zhang**, "A Dynamic Radiographic Imaging Study of Lumbar Intervertebral Disc Morphometry and Deformation In Vivo," *Sci Rep*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-019-51871-w.

[32] **K. P. Botwin and R. D. Gruber**, "Lumbar spinal stenosis: Anatomy and pathogenesis," in *Physical Medicine and Rehabilitation Clinics of North America*, 2003. doi: 10.1016/S1047-9651(02)00063-3.

[33] **F. Natalia** *et al.*, "Development of Ground Truth Data for Automatic Lumbar Spine MRI Image Segmentation," in *Proceedings - 20th International Conference on High Performance Computing and Communications, 16th International Conference on Smart City and 4th International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2018*, 2019. doi: 10.1109/HPCC/SmartCity/DSS.2018.00239.

[34] **S. Altun and A. Alkan**, "Lumbar Spinal Stenosis Analysis with Deep Learning Based Decision Support Systems," *Gazi University Journal of Science*, 2022, doi: 10.35378/gujs.1116423.

[35] **A. S. Al-Kafri** *et al.*, "Boundary Delineation of MRI Images for Lumbar Spinal Stenosis Detection Through Semantic Segmentation Using Deep Neural Networks," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2908002.

[36] **A. S. Al Kafri** *et al.*, "Segmentation of Lumbar Spine MRI Images for Stenosis Detection Using Patch-Based Pixel Classification Neural Network," in *2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings*, 2018. doi: 10.1109/CEC.2018.8477893.

[37] **N. C. Lehnen** *et al.*, "Detection of degenerative changes on mr images of the lumbar spine with a convolutional neural network: A feasibility study," *Diagnostics*, vol. 11, no. 5, 2021, doi: 10.3390/diagnostics11050902.

[38] **J. T. P. D. Hallinan** *et al.*, “Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI,” *Radiology*, vol. 300, no. 1, 2021, doi: 10.1148/radiol.2021204289.

[39] **F. Natalia and S. Sudirman**, “Classification of Sagittal Lumbar Spine MRI for Lumbar Spinal Stenosis Detection Using Transfer Learning of a Deep Convolutional Neural Network,” in *Lecture Notes in Networks and Systems*, 2022. doi: 10.1007/978-981-16-6309-3_16.

[40] **M. A. Knaub, D. S. Won, R. McGuire, and H. N. Herkowitz**, “Lumbar spinal stenosis: indications for arthrodesis and spinal instrumentation.,” *Instructional course lectures*, vol. 54. 2005.

[41] **S. Ruiz-España and D. Moratal**, “OncoSpineSeg: A Software Tool for a Manual Segmentation of Computed Tomography of the Spine on Cancer Patients,” in *Computed Tomography - Advanced Applications*, 2017. doi: 10.5772/intechopen.68552.

[42] **P. A. Yushkevich** *et al.*, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, 2006, doi: 10.1016/j.neuroimage.2006.01.015.

[43] **H. Li** *et al.*, “Automatic lumbar spinal MRI image segmentation with a multi-scale attention network,” *Neural Comput Appl*, vol. 33, no. 18, 2021, doi: 10.1007/s00521-021-05856-4.

[44] **S. Altun and A. Alkan**, “LSS-net: 3-dimensional segmentation of the spinal canal for the diagnosis of lumbar spinal stenosis,” *Int J Imaging Syst Technol*, vol. 33, no. 1, 2023, doi: 10.1002/ima.22807.

[45] **A. S. Lundervold and A. Lundervold**, “An overview of deep learning in medical imaging focusing on MRI,” *Zeitschrift fur Medizinische Physik*, vol. 29, no. 2. 2019. doi: 10.1016/j.zemedi.2018.11.002.

[46] **A. Krizhevsky, I. Sutskever, and G. E. Hinton**, “ImageNet classification with deep convolutional neural networks,” *Commun ACM*, vol. 60, no. 6, 2017, doi: 10.1145/3065386.

[47] **A. Esteva** *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, 2017, doi: 10.1038/nature21056.

[48] **A. Neubert** *et al.*, “Automated detection, 3D segmentation and analysis of high resolution spine MR images using statistical shape models,” *Phys Med Biol*, vol. 57, no. 24, 2012, doi: 10.1088/0031-9155/57/24/8357.

[49] **C. Janiesch, P. Zschech, and K. Heinrich**, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, no. 3, 2021, doi: 10.1007/s12525-021-00475-2.

[50] **T. Zhou, S. Ruan, and S. Canu**, “A review: Deep learning for medical image segmentation using multi-modality fusion,” *Array*, vol. 3–4, 2019, doi: 10.1016/j.array.2019.100004.

[51] **S. Suganyadevi, V. Seethalakshmi, and K. Balasamy**, “A review on deep learning in medical image analysis,” *Int J Multimed Inf Retr*, vol. 11, no. 1, 2022, doi: 10.1007/s13735-021-00218-1.

[52] **R. Janssens, G. Zeng, and G. Zheng**, “Fully automatic segmentation of lumbar vertebrae from CT images using cascaded 3D fully convolutional networks,” *Proceedings - International Symposium on Biomedical Imaging*, vol. 2018-April, pp. 893–897, May 2018, doi: 10.1109/ISBI.2018.8363715.

[53] **S. Barın and G. E. Güraksın**, “An automatic skin lesion segmentation system with hybrid FCN-ResAlexNet,” *Engineering Science and Technology, an International Journal*, vol. 34, 2022, doi: 10.1016/j.jestch.2022.101174.

[54] “Introduction to Convolutional Networks | by Manish aggarwal | Nerd For Tech | Medium.” https://medium.com/nerd-for-tech/convolutional-networks-b54335f4e21f (accessed Aug. 09, 2023).

[55] **Purwono, A. Ma’arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq**, “Understanding of Convolutional Neural Network (CNN): A Review,” *International Journal of Robotics and Control Systems*, vol. 2, no. 4, 2022, doi: 10.31763/ijrcs.v2i4.888.

[56] **J. Shi** *et al.*, “Improvement of damage segmentation based on pixel-level data balance using vgg-unet,” *Applied Sciences (Switzerland)*, vol. 11, no. 2, 2021, doi: 10.3390/app11020518.

[57] **Naveen**, “Difference between Sigmoid and Softmax activation function?,” *Nomidl*, 2022.

[58] **O. Ronneberger, P. Fischer, and T. Brox**, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015. doi: 10.1007/978-3-319-24574-4_28.

[59] **H. Zunair and A. Ben Hamza**, “Sharp U-Net: Depthwise convolutional network for biomedical image segmentation,” *Comput Biol Med*, vol. 136, 2021, doi: 10.1016/j.compbiomed.2021.104699.

[60] **J. Zhang, Y. Zhang, Y. Jin, J. Xu, and X. Xu**, “MDU-Net: multi-scale densely connected U-Net for biomedical image segmentation,” *Health Inf Sci Syst*, vol. 11, no. 1, 2023, doi: 10.1007/s13755-022-00204-9.

[61] **O. Ronneberger**, “Invited Talk: U-Net Convolutional Networks for Biomedical Image Segmentation,” 2017. doi: 10.1007/978-3-662-54345-0_3.

[62] **S. Jadon**, “SemSegLoss: A python package of loss functions for semantic segmentation[Formula presented],” *Software Impacts*, vol. 9, 2021, doi: 10.1016/j.simpa.2021.100078.

[63] **J. R. Clough, N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schnabel, and A. P. King**, “A Topological Loss Function for Deep-Learning Based Image Segmentation Using Persistent Homology,” *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 12, 2022, doi: 10.1109/TPAMI.2020.3013679.

[64] **Y. Chen** *et al.*, “A survey of loss function of medical image segmentation algorithms,” *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, vol. 40, no. 2, 2023, doi: 10.7507/1001-5515.202206038.

[65] **R. El Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah**, “High-level prior-based loss functions for medical image segmentation: A survey,” *Computer Vision and Image Understanding*, vol. 210, 2021, doi: 10.1016/j.cviu.2021.103248.

[66] **T. Sustersic, V. Rankovic, V. Milovanovic, Kovacevic V, Rasulic L, Filipovic N**, “A Deep Learning Model for Automatic Detection and Classification of Disc Herniation in Magnetic Resonance Images,” *IEEE Journal Biomedical Health Inform*, vol. 26, no. 12, pp. 6036-6046, 2022. doi: 10.1109/JBHI.2022.3209585.

[67] **Y. Wang** et al.**,** “Deep learning-based detection and classification of lumbar disc herniation on magnetic resonance images”, *JOR SPINE*, Wiley Online Library, vol. 6, no. 3, 2023.

[68] **M. A. Alomari** et al. “Deep learning-based detection and classification of lumbar disc herniation on magnetic resonance images,” *Journal of Orthopaedic Research*, vol. 41, no. 5, pp.1264-1276, 2023.

[69] **C. W. Tsai** et al, “A Deep Learning Model for Automatic Detection and Classification of Disc Herniation in Magnetic Resonance Images,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, Issue 5, pp. 2954-2965, 2023.

[70] **Y. Zhang** et al. “Deep learning-based detection and classification of lumbar disc herniation on magnetic resonance images,” *JOR SPINE*, vol.1, no. 3, pp. 121-130, 2023.

[71] **J. H. Lee** et al. “Automated diagnosis of lumbar disc herniation using convolutional neural networks,” *Journal of digital imaging*, vol. 30, no. 4, pp. 491-498, 2017.

[72] **X. Li** et al. “Automated diagnosis of lumbar disc herniation using deep learning,” *Journal of healthcare engineering*, 2018.

[73] **R. M. Amin, N. S. Andrade, B. J. Neuman**, “Lumbar disc herniation,” *Curr Rev Musculoskelet Med.*, vol.10, no. 4, pp. 507-516, 2017, doi: 10.1007/s12178-017-9441-4.

[74] **Henrik Weber**, “Lumbar Disc Herniation: A Controlled, Prospective Study with Ten Years of Observation,”, *SAS Journal*, vol. 3, no. 1, pp. 30-40, 2009, doi: 10.1016/S1935-9810(09)70005-2.

[75] **J. Y. Tsai, I. Y. Hung, Y. L. Guo**, et al. “Lumbar disc herniation automatic detection in magnetic resonance imaging based on deep learning,” *Front Bioeng Biotechnol*, 2021.

[76] **L. W. Mysliwiec, J. Cholewicki, M.D. Winkelpleck, G.P. Eis**, “MSU classification for herniated lumbar discs on MRI: toward developing objective criteria for surgical selection,” *Eur Spine J*, vol. 19, no. 7, 2010, doi: 10.1007/s00586-009-1274-4.

[77] **M d'Ercole, G. Innocenzi, F. Ricciardi, S. Bistazzoni**, “Prognostic value of Michigan State University (msu) classification for lumbar disc herniation: is it suitable for surgical selection?,” *Int J Spine Surg*, vol.15, no. 3, pp. 466-470, 2021. doi: 10.14444/8068.

[78] **A.K. Kaliya-Perumal, C.A. Luo, Y.C. Yeh, Y.F. Tsai, M.J. Chen, T.T. Tsai,** “Reliability of the Michigan State University (MSU) classification of lumbar disc herniation,” *Acta Ortop Bras*, vol. 26, no. 6, pp.411-414, 2018. doi:10.1590/1413-785220182606201444.

[79] **D. Shen, G. Wu, HI. Suk,** "Deep Learning in Medical Image Analysis," *Annu Rev Biomed Eng*, vol. 19, pp. 221-248, 2017. doi: 10.1146/annurev-bioeng-071516-044442.

[80] **N.C. Lehnen, R. Haase, J. Faber,** et al. "Detection of degenerative changes on MR images of the lumbar spine with a convolutional neural network: a feasibility study," *Diagnostics (Basel)*, vol. 11, no. 5, 2021. doi: 10.3390/diagnostics11050902.

[81] **R.S. Alomari, J.J Corso, V. Chaudhary, G. Dhillon**, "Computer-aided diagnosis of lumbar disc pathology from clinical lower spine MRI," *Int J Comput Assist Radiol Surg*, vol. 5, no.3, pp. 287-293, 2010.

[82] **J.T. Lu, S. Pedemonte, B. Bizzo**, et al. "Deep spine: automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning," *ArXiv*, vol. 85, pp. 403-419, 2018.

[83] **P. Rajpurkar**, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," A*rXiv*, vol. 3, 2017.

[84] **V. Markotić, T. Pojužina**, **D. Radančević, M. Miljko, V. Pokrajčić**, "The Radiologist Workload Increase; Where Is the Limit?: Mini Review and Case Study," *Psychiatr Danub*, vol. 33, pp. 768-770, 2021.

[85] **M.A. Mazurowski, M. Buda, A. Saha, M.R. Bashir**, "Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI," *J Magn Reson Imaging*, vol. 49, no.4, pp. 939-954, 2019. doi: 10.1002/jmri.26534.

[86] **A.W. Moawad, D.T. Fuentes, M.G. ElBanan**, et al. "Artificial Intelligence in Diagnostic Radiology: Where Do We Stand, Challenges, and Opportunities," *J Comput Assist Tomogr*, vol. 46, no. 1, pp. 78-90, 2022. doi: 10.1097/RCT.0000000000001247.

[87] **N.N. Knezevic, K.D. Candido, J.W.S. Vlaeyen, J. Van Zundert, S.P. Cohen**, "Low back pain," *Lancet*, vol. 398, no. 10294, pp. 78-92, 2021. doi: 10.1016/S0140-6736(21)00733-9.

[88] **T. Paolucci, C. Attanasi, W. Cecchini, A. Marazzi, S.V. Capobianco, V. Santilli**, "Chronic low back pain and postural rehabilitation exercise: a literature review," *J Pain Res*, vol. 12, pp. 95-107, 2018. doi: 10.2147/JPR.S171729.

[89] **A. Qaseem, T.J. Wilt, R.M. McLean**, et al "Noninvasive Treatments for Acute, Subacute, and Chronic Low Back Pain: A Clinical Practice Guideline from the American College of Physicians," *Ann Intern Med.*, vol. 166, no. 7, pp. 514-530, 2017. doi: 10.7326/M16-2367.

[90] **X. He, Y. Yin, M. Sharma, G. Brahm, A. Mercado, S. Li,** "Automated Diagnosis of Neural Foraminal Stenosis Using Synchronized Superpixels Representation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9901, pp. 335-343, 2016.

[91] **Y. Cai, S. Leung, J. Warrington, S. Pandey, O. Shmuilovich, S. Li**, "Direct spondylolisthesis identification and measurement in MR/CT using detectors trained by articulated parameterized spine model," *SPIE Medical Imaging*, vol. 10133, 2017. doi:10.1117/12.2254072.

[92] **A. Yabu, M. Hoshino, H. Tabuchi**, et al. “Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images,” *Spine J.*, vol. 21, no. 10, pp. 1652-1658, 2021. doi: 10.1016/j.spinee.2021.03.006. Epub 2021 Mar 13. PMID: 33722728.

[93] **K.U. LewandrowskI, N. Muraleedharan, S.A. Eddy**, et al “Feasibility of Deep Learning Algorithms for Reporting in Routine Spine Magnetic Resonance Imaging,” *Int J Spine Surg*, vol. 14, no. 3, pp. 86-97, 2020. doi: 10.14444/7131.

[94] **Z.H. Su, J. Liu, M.S. Yang**, et al. “Automatic Grading of Disc Herniation, Central Canal Stenosis and Nerve Roots Compression in Lumbar Magnetic Resonance Image Diagnosis,” *Front Endocrinol (Lausanne)*, vol. 13, 2022. doi: 10.3389/fendo.2022.890371.

[95] **A. Jamaludin, T. Kadir, A. Zisserman**, “SpineNet: Automated classification and evidence visualization in spinal MRIs,” *Med Image Anal.*, vol. 41, pp. 63-73, 2017. doi: 10.1016/j.media.2017.07.002.

[96] **R.S. Alomari, J.J. Corso, V. Chaudhary, G. Dhillon**, “Toward a clinical lumbar CAD: herniation diagnosis,” *International Journal of Computer Assisted Radiology and Surgery* , vol. 6, no. 1, pp. 119–126, 2011. doi: 10.1007/s11548-010-0487-7.

[97] **S.G. Beyaz, A.M. Ülgen, B. Kaya, M.E. İnanmaz, T. Ergönenç, A. Eman, B. Doğan**, “A Novel Combination Technique: Three Points of Epiduroscopic Laser Neural Decompression and Percutaneous Laser Disc Decompression With the Ho:YAG Laser in an MSU Classification 3AB Herniated Disc,” *Pain Practice*, vol. 20, no. 5, pp. 501-509, 2020. doi: 10.1111/papr.12878.

[98] **S. Ren, K. He, R. Girshick, J. Sun**, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137-1149, 2015.

[99] **K. He, X. Zhang, S. Ren, J. Sun**, “Deep Residual Learning for Image Recognition,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016. doi:10.1109/cvpr.2016.90.

[100] **S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He**, “Aggregated Residual Transformations for Deep Neural Networks,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987-5995, 2017.

[101] **P.E. Shrout, J.L. Fleiss**, “Intraclass correlations: uses in assessing rater reliability,” *Psychol Bull*, vol. 86, no. 2, pp. 420-428, 1979. doi: 10.1037//0033-2909.86.2.420.

[102] **L.L. Wiltse, P.E. Berger, J.A. McCulloch**, “A system for reporting the size and location of lesions in the spine,” *Spine (Phila Pa 1976)*, vol.22, no. 13, pp.1534-1537, 1997. doi: 10.1097/00007632-199707010-00023.

[103] **D. Won, H.J. Lee, S.J. Lee, S.H. Park**, “Spinal Stenosis Grading in Magnetic Resonance Imaging Using Deep Convolutional Neural Networks,” *Spine (Phila Pa 1976)*, vol. 45, no. 12, pp. 804-812, 2020. doi: 10.1097/BRS.0000000000003377.

[104] **R.J. McDonald, K.M. Schwartz, L.J. Eckel**, et al. “The effects of changes in utilization and technological advancements of cross-sectional imaging on

radiologist workload," *Academic Radiology*, vol. 22, no. 9, pp. 1191-1198, 2015. doi: 10.1016/j.acra.2015.05.007.

[105] **D.F. Fardon, P.C. Milette**, "Combined Task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology. Nomenclature and classification of lumbar disc pathology. Recommendations of the Combined Task Forces of the North American Spine Society, American Society of Spine Radiology, and American Society of Neuroradiology," *Spine (Phila Pa 1976)*, vol. 26, no. 5, pp. 93-113, 2001. doi: 10.1097/00007632-200103010-00006.

RESUME

Mustafa Isam AL-AJAJ

EDUCATION:

- Bachelor’s degree: 2008, University of Baghdad, Administration and Economic College, department of Statistics.
- Master’s degree Department of Statistics and Data Science, Istanbul Gedik University, Istanbul, Turkey in 2024.

PROFESSIONAL EXPERIENCE:

- Statistical analysis.
- Programming in Python and PHP.
- Website management and design.
- Installing local networks.
- Programming and installation of local networks and their settings.
- Programming and installation of servers