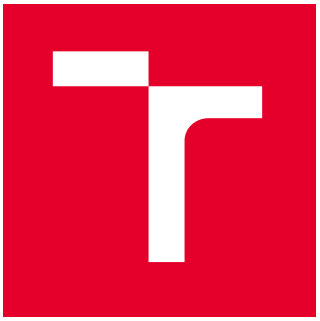


BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

ANALYSIS OF OSTEOLYTIC SPINAL TUMORS IN PATIENTS WITH MULTIPLE MYELOMA USING CT DATA

ANALÝZA OSTEOLYTICKÝCH NÁDORŮ PÁTEŘE U PACIENTŮ S MNOHOČETNÝM MYELOMEM NA DATECH Z CT

BACHELOR'S THESIS

BAKALÁŘSKÁ PRÁCE

AUTHOR — **Miriam Čurillová**

AUTOR PRÁCE

SUPERVISOR — **Ing. et Ing. Michal Nohel**

VEDOUCÍ PRÁCE

BRNO 2024

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

Bachelor's Thesis

Bachelor's study program **Biomedical Technology and Bioinformatics**

Department of Biomedical Engineering

Student: Miriam Čurillová ***ID:*** 231091

Year of study: 3 ***Academic year:*** 2023/24

TITLE OF THESIS:

Analysis of osteolytic spinal tumors in patients with multiple myeloma using CT data

INSTRUCTION:

1) Learn about spinal tumors in patients with multiple myeloma and their diagnosis using spectral CT. 2) Propose an appropriate methodology for the characterization of spinal tumors and their evaluation. 3) Become familiar with the available data and the possibilities of different parametric images. Familiarize yourself with the available annotations and make the necessary modifications. 4) Implement the proposed analysis approach in the chosen programming environment and apply it to the developed database. 5) Perform statistical evaluation of the results. Analyze the influence of parameters and annotation inaccuracies on the results of the analysis. 6) Appropriately present and discuss the results obtained, possibilities and problems of tumor analysis.

RECOMMENDED LITERATURE:

[1] LEE, Seungeun, So-Yeon LEE, Sanghee KIM, Yeon-Jung HUH, Jooyeon LEE, Ko-Eun LEE a Joon-Yong JUNG. Differentiating Multiple Myeloma and Osteolytic Bone Metastases on Contrast-Enhanced Computed Tomography Scans: The Feasibility of Radiomics Analysis. Diagnostics [online]. 2023, 13(4) [cit. 2023-07-25]. ISSN 2075-4418. Dostupné z: doi:10.3390/diagnostics13040755

[2] MCCOLLOUGH, Cynthia H., Shuai LENG, Lifeng YU a Joel G. FLETCHER. Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications. Radiology [online]. 2015, 276(3), 637-653 [cit. 2023-07-25]. ISSN 0033-8419. Dostupné z: doi:10.1148/radiol.2015142631

Date of project specification: 5.2.2024 ***Deadline for submission:*** 29.5.2024

Supervisor: Ing. et Ing. Michal Nohel

Consultant: Mgr. Ing. Marek Dostál, PhD.

doc. Ing. Jana Kolářová, Ph.D.
Chair of study program board

ABSTRACT

This bachelor's thesis focuses on analysis of osteolytic lesions in patients with multiple myeloma. The first step in achieving our goal was to research this disease, its diagnostic criteria, possible complications and available treatment. The practical part consisted of a few individual tasks. A statistical analysis was done on a dataset consisting of CT scans of patients with diagnosed multiple myeloma as well as individuals with no spinal pathologies. After extracting and reducing the number of features, we completed an analysis of obtained data. We came to a conclusion that there are features that vary significantly among the two groups. After analyzing the whole vertebral bodies, analysis of lesions in follow-up scans was completed, where their volume was analyzed.

KEYWORDS

multiple myeloma, computed tomography, osteolytic lesions, feature extraction, radiomics, statistical analysis

ABSTRAKT

Táto bakalárska práca sa zameriava na analýzu osteolytických lézií u pacientov s mnohopočetným myelómom. Prvým krokom k dosiahnutiu nášho cieľa bolo štúdium tohto ochorenia, jeho diagnostických kritérií, možných komplikácií a dostupnej liečby. Praktická časť pozostávala z niekoľkých samostatných úloh. Štatistická analýza bola vykonaná na súbore údajov pozostávajúcom z CT skenov pacientov s diagnostikovaným mnohopočetným myelómom, ako aj osôb bez akejkoľvek patológie chrbtice. Po extrakcii a redukcii počtu príznakov sme dokončili analýzu získaných dát. Dospeli sme k záveru, že existujú príznaky, ktoré sa medzi týmito dvoma skupinami výrazne líšia. Po analýze celých tiel stavcov bola vykonaná analýza lézií na kontrolných snímkach, kde bol analyzovaný ich objem.

KĽÚČOVÉ SLOVÁ

mnohopočetný myelóm, počítačová tomografia, osteolytické lézie, extrakcia príznakov, rádiomika, štatistická analýza

ROZŠÍRENÝ ABSTRAKT

Táto bakalárska práca sa zaoberá analýzou osteolytických lézií u pacientov s mnohopočetným myelómom na CT dátach. Práca pozostáva z teoretickej a praktickej časti. V teoretickej časti sme sa venovali samotnému ochoreniu, využitiu počítačovej tomografie pri jej diagnostike a základom rádiomiky. Mnohopočetný myelóm je druh rakoviny kostnej drene, ktorý postihuje najmä stavce, hrudnú kosť a panvu. Pacienti s mnonohopočeným myelómom často trpia neznesiteľnými bolesťami, ktoré sú jedným z hlavných príznakov ochorenia. Medzi ďalšie prejavy choroby patrí anémia, hyperkalcémia a tvorba osteolytických lézií. Tieto lézie boli objektom našej analýzy. Lézie sa na CT obrazoch vyznačujú tmavším zafarbením, čo otvára rôzne možnosti analýzy daných obrazov. Aj keď medicína ponúka viacero možností zmiernenia príznakov mnohopočetného myelómu, v súčasnej dobe ide stále o nevyliečiteľnú chorobu. Pri diagnostike mnohopočetného myelómu sa postupuje podľa presne určených krokov a neodmysliteľnú úlohu hrajú zobrazovacie systémy ako CT, MRI a PET/CT. Rádiomika je úzko spätá s lekárskymi zobrazovacími technikami. Ponúka rôzne prístupy ako získať z obrazových dát informácie, ktoré môže byť pre ľudské oko neviditeľné, avšak môžu byť rozhodujúcim faktorom pre stanovenie správnej diagnózy a nastavenie vhodnej liečby.

Praktická časť práce je rozdelená do dvoch celkov. Pred samotnou analýzou bolo potrebné zozbierať vhodné dáta a upraviť ich pre potreby našej analýzy. Pracovali sme s dvoma rôznymi datasetmi. Prvý dataset obsahoval klinické dáta. Dáta boli získané z Fakultnej nemocnice Brno, poskytnuté Klinikou radiologie a nukleární medicíny. V rámci stáže boli dáta anonymizované, aby sa zabezpečila ochrana osobných údajov pacientov. Dataset pozostával z dát od desiatich osôb bez patológie chrbtice a desiatich pacientov s mnohopočetným myelómom. Neskôr boli zozbierané aj dáta niekoľkých pacientov s kontrolnými vyšetreniami, na ktorých je možné pozorovať vývoj choroby v čase a posudzovať účinnosť liečby. Okrem samotných CT skenov sme mali k dispozícií rôzne parametrické obrazy a segmentačné masky. Väčšina skenov boli celotelové CT obrazy, niektoré však obsahovali len skeny krčnej a hrudnej časti chrbtice.

Druhým použitým datasetom bola verejne dostupná databáza CT skenov bedrovej časti chrbtice. Dataset pozostával z 185 CT skenov a ich segmentačných masiek stavcov, ktoré nezahrňovali výbežky ani okosticu. Na rozdiel od klinických dát, skeny neobsahovali žiadne patológie a pacietni s akoukoľvek anomáliou chrbtice (implantáty, lézie, nádory) boli z datasetu vyradení.

Prvá časť je venovaná analýze tiel postihnutých ako aj zdravých stavcov. Kvôli výpočetnej náročnosti bolo výhodné si pred samotnou analýzou stavcov jednotlivé stavce vyextrahovať zo skenov. Takto sme zabezpečili, že namiesto veľkého objemu dát sme pracovali s menšími celkami, čo bolo časovo výhodnejšie. Na extrakciu

stavcov z obrazov boli použité dva samostatné postupy, ktoré boli nakoniec navzájom porovnávané. Prvý prístup spočíval v manuálnej extrakcii stavcov použitím ich priestorových súradníc. Tento postup bol časovo náročný a nevedeli sme zaručiť, že z obrazu dostaneme presne nami požadovanú časť bez toho, aby sme extrahovali aj výbežky alebo okosticu.

V druhom prístupe sme na segmentáciu tiel stavcov využili model strojového učenia založený na nnU-Net modele. Verejne dostupná databáza bola v pomere 80:20 rozdelená na trénovacie a testovacie dáta. Účinnosť modelu bola vyhodnotená na testovacích dátach s využitím metriky Dice a Housdorffovej vzdialenosti. Na testovacích dátach sme dosiahli relatívne vysokú presnosť, s mediánovými hodnotami Dice 0.980 a Housdorffovej vzdialenosti 2.159. Model bol následne aplikovaný na klinické dáta a výsledky boli vyhodnotené vizuálne. Pri stavcoch obsahujúcich anomálie bol pozorovaný pokles presnosti, čo bol predpokladaný výsledok. Správne segmentované stavce boli použíté pre ďalšiu analýzu.

Ďalším krokom bola štatistická analýza jasových hodnôt stavcov. Účelom bolo zistiť, či je medzi zdravými a postihnutými stavcami štatisticky významný rozdiel. V tejto časti sme používali len stavce vysegmentované naučeným modelom strojového učenia. V porovnávaní sme využili okrem konvenčných CT skenov aj dostupné parametrické obrazy, ako sú obrazy s potlačením vápnika a monoenergetické obrazy. Pre lepšiu predstavu bolo zhotovených niekoľko krabicových grafov na porovnanie zdravých a postihnutých stavcov. V grafoch boli jasne pozorovateľné rozdiely medzi týmito dvoma skupinami. Využitím štatistického testu bola potvrdená štatistická významnosť týchto rozdielov. S vedomím toho, že tieto dve skupiny sa od seba štatisticky významne líšia, sme pristúpili k ďalšej analýze.

Na získanie rádiomických príznakov bola použitá knižnica Pyradiomics, ktorá je často využívaná na analýzu medicínskych obrazových dát. Príznaky boli extrahované zo stavcov osôb bez akejkoľvek patológie chrbtice, ako aj zo stavcov pacientov, ktorým bola choroba v minulosti diagnostikovaná. Za účelom lepšieho porovnania boli vyselektované stavce s viditeľnými patológiami. Použili sme stavce segmentované oboma použitými prístupmi a porovnali sme, ako sa budú výsledky analýzy medzi sebou odlišovať.

Na príznakoch bola vykonaná analýza hlavných komponentov, aby sme zistili, či sa od seba dve skupiny stavcov svojimi príznakmi odlišujú. Vizualizáciou bolo potvrdené, že náš predpoklad je správny a vytvorili sa dva jasne odlíšiteľné zhluky. V prípade stavcov segmentovaných pomocou modelu strojového učenia boli zhluky lepšie ohraničené a v prípade zdravých stavcov sme pozorovali aj rozdiel medzi hrudnými a bedrovými stavcami. S využítím random forestu sme boli schopní stanoviť významnosť jednotlivých príznakov, ktoré by boli využiteľné pri trénovaní klasifikátora. Aj v tejto analýze sme pozorovali rozdielnu váhu príznakov medzi manuálne

extrahovanými stavcami a stavcami segmentovanými modelom strojového učenia.

Posledným celkom praktickej časti bola analýza kontrolných skenov. Pre účely analýzy boli vybraní dvaja pacienti. Pri oboch pacientoch sme si vyselektovali tri lézie, ktoré boli dobre ohraničené a pozorovali sme ich zmenu. Skeny boli vytvorené s odstupom troch mesiacov, čo je pri takýchto pacientoch štandardná doba medzi kontrolnými vyšetreniami. Prvým krokom bola analýza objemu jednotlivých lézií. Pri léziách získaných od jedného pacienta sme pozorovali kladnú aj zápornú zmenu objemu a rôzny percentuálny nárast alebo zmenšenie. Podobný scenár sa opakoval aj pri léziách extrahovaných zo skenu druhého pacienta. Na základe tohto nemôžeme spoľahlivo posúdiť účinnosť liečby. Okrem objemu lézií boli analyzované aj zmeny v priemere a smerodatnej odchylke jasových hodnôt. Podobne ako predtým sme pozorovali rôzne zmeny u rovnakého pacienta. Okrem skutočnej zmeny objemu a jasových hodnôt lézií môžu byť zmeny zapríčinené aj nesprávnou segmentáciou hraničných oblastí. Na záver sme vyhotovili zoznam príznakov s najväčšou percentuálnou zmenou medzi prvým a kontrolným skenom.

Veríme, že práca bola spracovaná zrozumiteľne a všetky kroky vedúce k výsledkom sú riadne popísané a pochopiteľné. Jednotlivé kroky sú ľahko zreprodukovateľné. Analýza obrazov hrá v medicíne významnú rolu a dúfame, že aj táto práca môže byť spôsobom, ako čitateľovi prilížiť možnosti využitia rádiomických príznakov pri diagnostike rôznych patológií. Ďalším možným krokom by bolo natrénovanie klasifikátora na vyhodnotenie stavu stavca na základe našich výsledkov.

ČURILLOVÁ, Miriam. *Analysis of osteolytic spinal tumors in patients with multiple myeloma using CT data*. Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Biomedical Engineering, 2024, 70 p. Bachelor's Thesis. Advised by Ing. et. Ing. Michal Nohel

Author's Declaration

Author:	Miriam Čurillová
Author's ID:	231091
Paper type:	Bachelor's Thesis
Academic year:	2023/24
Topic:	Analysis of osteolytic spinal tumors in patients with multiple myeloma using CT data

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno ....

author's signature*

*The author signs only in the printed version.

ACKNOWLEDGEMENT

I would like to thank the advisor of my thesis, Ing. et Ing. Michal Nohel, for his guidance, valuable comments, patience, and professional insights, and The University Hospital Brno for providing clinical data.

Contents

List of Figures

Introduction

Cancer stands as one of leading causes of mortality worldwide. It is often detected through the presentation of symptoms. Its early detection and intervention is crucial. Multiple myeloma is a type of cancer that starts in the bone marrow, where blood cells are made. Specifically, it affects a type of cell called plasma cells. These cells normally help fight infections, but in multiple myeloma, they grow out of control. This can lead to problems like weak bones, anemia, and trouble fighting infections. Though it's a serious condition, treatments like medications, chemotherapy, and sometimes stem cell transplants can help manage it and improve quality of life. [1, 2]

Advancements in technology have transformed cancer diagnosis, with radiomics standing out as an innovative area. This field harnesses sophisticated imaging technologies such as MRI, CT scans, and PET scans to produce highly detailed body images. Radiomics goes beyond simple imaging, using complex algorithms and software to process and evaluate these images. It uncovers hidden patterns or characteristics that often go unnoticed by humans, enhancing tumor detection and characterization. By extracting information from medical images, radiomics offers the potential to tailor treatments, predict outcomes, and enhance overall cancer management. [3]

This bachelor's thesis focuses on utilizing radiomics in the statistical analysis of multiple myeloma lesions on CT scans. We use data from patients being treated for multiple myeloma as well as individuals with no spinal pathologies. Comparing these two groups gives us a better insight into what radiomic features are crucial to consider and which ones are, on the other hand, not important and can be skipped. In this bachelor's thesis, we work with two different datasets - a publicly available dataset and a dataset consisting of clinical data. These CT scans were used to extract vertebral bodies, which are later used for the analysis itself. Statistical analysis was performed to explore the differences between healthy and affected vertebrae. The last part of the thesis deals with follow-up scans where we examine the changes in the lesions over a period of three months.

1 Spine anatomy

The human spine, known interchangeably as the vertebral column or backbone, is composed of small bones known as vertebrae. These vertebrae are named according to their position along the length of the spinal column. In all, the spine encompasses 33 to 34 vertebrae, which are organized into five groups: [4]

- 7 cervical vertebrae,
- 12 thoracic vertebrae,
- 5 lumbar vertebrae,
- 5 sacral vertebrae fused into 1 sacrum,
- 4 to 5 coccygeal vertebrae fused into 1 coccyx.

Cervical vertebrae serve to support the head and enable its mobility. Thoracic vertebrae form connections with the ribs, while lumbar vertebrae, the spine's largest and most robust bones, lie between the rib cage and the pelvis. The sacrum connects with the hip bones, and the coccyx, known as the tailbone, marks the vertebral column's concluding segment. Additionally, certain muscles of the pelvic floor are anchored to the coccyx. [4]

Ligaments connect all the vertebrae to create a flexible backbone that offers sufficient support for the entire body. Intervertebral discs separate adjacent vertebrae, acting as cushions and facilitating movement between them. Moreover, the spine serves as a protective encasement for the spinal cord housed within the vertebral canal. A healthy spine has four inherent curves, named after the vertebrae that form them: thoracic kyphosis, cervical lordosis, lumbar lordosis, and sacral kyphosis. [4]

1.1 Vertebrae anatomy

Vertebrae exhibit variations in shape and size based on their location along the spine. Despite these variances, they share a largely similar structure. Each vertebra consists of three components: a body, a vertebral arch, and processes. The body of the vertebra bears the weight of the body, with its size and thickness increasing progressively down the spinal column. Situated in the posterior portion of each vertebra, the vertebral arch consists of four parts: the right and left pedicles, and the right and left laminae. The spacious opening between the vertebral arch and body houses the spinal cord. Additionally, seven processes arise from most of the vertebrae, serving as vital sites for muscle attachment. These processes also facilitate the formation of slightly movable joints between adjacent vertebrae [6]. Fig. 1.1 shows the vertebral column from anterior, right lateral and posterior view. Fig. 1.2 shows parts of a typical vertebrae.

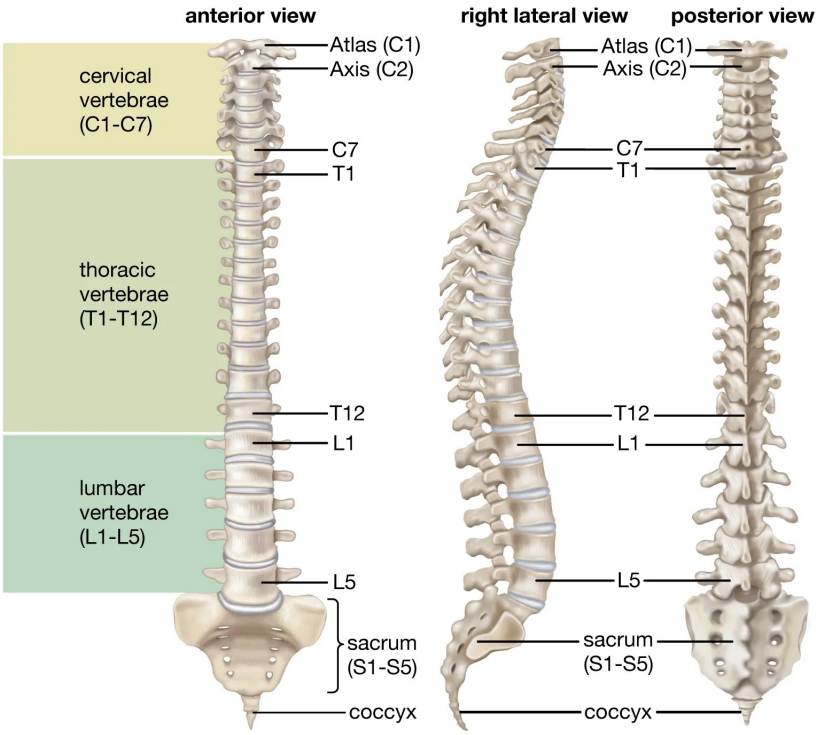


Fig. 1.1: Vertebral column anatomy [5].

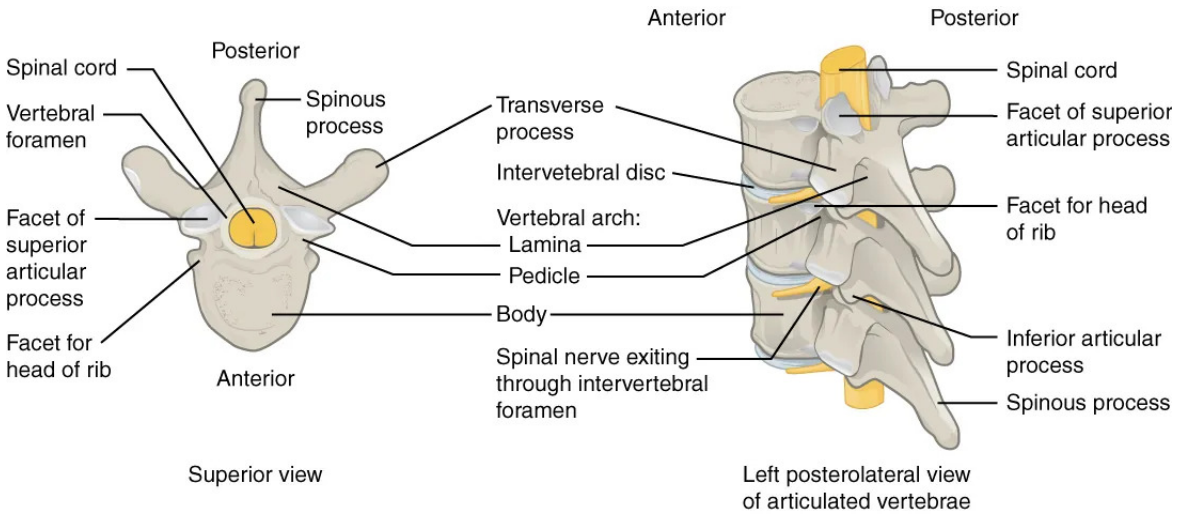


Fig. 1.2: Parts of a typical vertebrae [6].

1.2 Spinal diseases

The term spinal disease refers to a range of conditions or disorders that affect the backbone. While some spinal diseases garner more recognition than others, they all cause a level of discomfort to the affected individual. Here, we will highlight a few

of these conditions, focusing on those most familiar to the general population.

Scoliosis

Scoliosis refers to a sideways curvature of the spine, commonly identified during adolescence. The cause of most childhood scoliosis is unknown. While certain instances may begin mildly, the curvature can significantly exacerbate as children continue to grow. Severe scoliosis cases can result in disability, giving rise to various medical complications, such as compromised lung function due to reduced chest space. [7]

Lumbar spinal stenosis

Lumbar spinal stenosis involves the narrowing of the spinal canal in the lower back region, which results in the compression of nerves extending from the lower back down to the legs. Stenosis may arise from various factors, including injuries or osteoarthritis. Additionally, some individuals may develop spinal stenosis due to congenitally narrower spinal canals. [8]

Spinal tumors

Just like any other kind of tumor, the term spinal tumor refers to an abnormal growth of tissue in the area of the spine. Tumors can exhibit either benign or malignant characteristics. Those originating directly within the spine or spinal cord are termed primary tumors. Conversely, cancerous tumors elsewhere in the body can spread to the spine, resulting in secondary tumors. [9]

Bone tumors are categorized into osteoblastic, osteolytic, and mixed types. Osteolytic lesions signify an imbalance in bone remodeling, where bone breakdown surpasses bone formation. Consequently, bones become fragile and susceptible to pathological fractures. This condition significantly impacts patients' quality of life, often accompanied by severe bone pain. On the other hand, osteoblastic lesions develop due to an excess formation of bone tissues. This leads to bones being very dense and deformed. In these instances, bone formation happens faster than bone resorption, causing an abnormal proliferation of the affected bone tissue. [10, 11]

2 Multiple myeloma

2.1 Overview

Multiple myeloma (MM) is a form of bone marrow cancer and is the most prevalent cancer affecting the skeletal system. MM triggers the development of osteolytic bone lesions, areas of compromised bone tissue that weaken the bone and increase fracture risk. Common symptoms of MM typically involve severe bone pain, hypercalcemia, spinal cord compression, and pathological fractures. [1]

MM can affect any bone, but the dominant areas of occurrence are bones containing a higher quantity of red marrow. Consequently, bones such as the vertebral bodies, skull, pelvis, and ribs are particularly susceptible to MM. [1]

2.2 Myeloma bone disease

Multiple myeloma bone disease (MMBD) is characterized by impaired bone formation and increased bone resorption. One of the causes is that the osteoblast activity is severely decreased or completely absent and the bone destruction is increased due to increased osteoclast activity. To make matters worse, bone scans in patients with MM can often underestimate the severity of the disease and the lesions are extremely hard to heal. MM bone lesions typically present with a distinctive 'punched-out' appearance on X-rays. [1]

2.3 Prevalence, presentation and diagnosis

Clinical presentation of MM varies among patients, making it challenging to identify initially. In addition to this, approximately 11 % of all MM patients are initially asymptomatic which only makes it easier to miss [1]. In such instances, the disease is typically diagnosed through routine laboratory tests. However, this approach may result in delays, potentially leading to treatment initiation when it's already too late. Because of this, there are a few criteria when diagnosing MM. Even if a patient doesn't exhibit all criteria, or if the presentation is uncertain, prompt treatment initiation is vital to reduce the risk of end-organ damage. [1]

In nearly all cases, MM is preceded by an asymptomatic premalignant phase known as monoclonal gammopathy of undetermined significance (MGUS) [12]. MGUS is present in about 3-4 % of the population above the age of 50. According to [12], having a BMPC (bone marrow plasma cell percentage) of at least 60 % means that

the individual has a notably high risk of progression. With this knowledge, the protocol dictates that these patients should be considered to have multiple myeloma and offered appropriate treatment. One of the most prevalent initial symptoms is bone pain, experienced by roughly two-thirds of patients [12].

The vast majority of MM patients are elderly. According to [1] a median age at diagnosis is 69 and a median age at death is 74. Even though the treatment of MM has made a long way over the last few years, the disease still remains incurable.

2.4 Diagnosis criteria

As mentioned above, the early diagnosis is crucial in any type of cancer. MM patients often time present with a set of symptoms, also referred to as CRAB features (calcemia, renal failure, anemia, bone lesions). [13]

2.4.1 Hypercalcemia

Hypercalcemia in multiple myeloma is caused by bone destruction from osteolytic tumor lesions. Calcium released from damaged bone tissue enters the extracellular fluid. Additionally, renal insufficiency, common among MM patients, leads to inadequate calcium excretion, resulting in calcium accumulation in the body. Symptoms associated with hypercalcemia include nausea, vomiting, confusion, depression, anorexia, constipation, and eventually coma. In MM patients, hypercalcemia typically indicates uncontrolled disease or resistance to current therapy, although not all MM patients experience hypercalcemia. Treatment for hypercalcemia generally involves hydration and either a bisphosphonate or RANKL inhibitor. Addressing the underlying cause is crucial, beginning with fluid replenishment followed by administration of an anti-hypercalcemic medication. [13]

2.4.2 Renal failure

Renal failure (RF) stands as one of the most prevalent and severe complications associated with MM. The destruction of the bone marrow microenvironment leads to excessive production of monoclonal immunoglobulins. RF may result from an abundance of these immunoglobulins or other factors such as hypercalcemia, infection, etc.[14]

RF is a medical emergency, which requires measures to be taken as soon as possible. Because of the severity of the issue and the potential damage it can cause

to the whole body, various potential mechanisms of RF are usually addressed simultaneously. General measures include hydration with urinary alkalization. Chemotherapy should also be started as soon as possible, however, the drugs used must not be excreted through the kidneys. Despite implementing these general measures, improvement may not always be evident. In such instances, renal replacement therapy or dialysis may be required. [14]

2.4.3 Anemia

According to [15] around 60-70 % of MM patients have anemia at the time of their diagnosis and almost all of them will experience it at some point during the course of the disease. The most common causes of anemia in MM are anemia of chronic disease, relative erythropoietin deficiency, and myelosuppressive effects of chemotherapy. Treatment options usually include red blood cell transfusions and recombinant human erythropoietin. [15]

2.4.4 Bone lesions

In a healthy skeleton, there is a balance between the breakdown of old bone tissue (osteoclasts) and the building of new bone tissue (osteoblasts). MM disrupts this balance by producing osteoclast-activating factors and inhibiting the formation of osteoclasts. This leads to uncontrollable breakdown of the bone. As previously mentioned, multiple myeloma weakens bones, making them more susceptible to fractures. The bones most frequently affected include the spine, pelvis, ribs, skull, and the long bones of the limbs. Multiple myeloma bone disease can also lead to spinal cord compression due to pressure exerted on the spinal cord [16]. According to [16], 70 % of MM patients have bone loss in the spine.

The initial stage of lytic lesions is called focal lesions. Focal lesions are abnormal areas in the bone marrow. These areas signal the development of lytic lesions. Typically, lytic lesions appear within the subsequent 18-24 months. Even if a patient exhibits no symptoms, the presence of more than one focal lesion of at least 5 mm in size on an MRI scan constitutes a myeloma-defining event, and the patient should be treated for active disease. [16]

Lytic lesions are areas where the bone tissue has been destroyed, leaving a hole in the bone. Some lytic lesions in specific bones may necessitate surgical intervention to fortify and stabilize the bone. The International Myeloma Working Group (IMWG) defines the minimal amount of bone damage that requires surgery. The two requirements are: [16]

- more than one focal lesion, measuring 5 mm or more in size on MRI or
- one or more lytic bone lesions detected on CT scan.

3 Computed tomography

3.1 Introduction

Wilhelm Röntgen's discovery of X-rays marked a significant advancement in diagnostic imaging. Despite its undeniable advantages, over time, the limitations of X-ray technology became increasingly apparent. Early applications of X-rays in imaging involved positioning an X-ray source opposite a radio-sensitive film, with the subject placed between them. However, this setup significantly compromised the diagnostic value of the scan. Information along the projection axis was eliminated, resulting in the loss of vital details about the subject's internal structure. [17]

Radon transform, formalized by Johann Radon, offers a means to mathematically reconstruct a 2D function from an infinite series of projections of that function. This method enables the extraction of information about the internal structure of the imaged subject. [17]

The first person to produce an image by using different projection angles was William Oldendorf in 1963. He came up with a completely new technique of recording the data as the source and the film were rotated around the subject. This innovative approach led to the creation of what is now recognized as a sinogram. However, Oldendorf's work did not involve the reconstruction of the sinogram into images. [17]

Credit for this achievement goes to Godfrey Hounsfield. He was the first one to implement the reconstruction of the image by using the measured projections as solutions to linear equations describing the attenuation of a discretized volume of the scanned body along the path of the beam. The research of the principles of the CT machine did not stop there. Shortly after the invention of the CT machine as we know it today, it was discovered that using multiple scanning energies is also beneficial. By implementing this, the physical properties of the material being scanned, such as density and atomic number can be obtained. Originally, it was observed that it is possible to decompose a CT scan into a set of N basis materials, given that you measure the data at N different energies. When talking about dual-energy CT, naturally, $N = 2$. But we can also go higher and increase the number of different energies. So, if $N > 2$, we are talking about multi-spectral CT. It was later found that in reality, you can decompose the image into $N + 1$ basis materials. [17]

3.2 Computed tomography principle

As previously stated, the fundamental principle of CT revolves around photon attenuation. When a photon interacts with the subject under examination, two outcomes are possible: absorption within the medium or scattering away from the original beam trajectory. Consequently, a beam of photons experiences attenuation as it traverses through a medium. This decrease of energy is characteristic of the medium. We can describe this using Beer-Lambert law [17]:

$$I = \int I_0\left(E\right) e^{-\int \mu(E,s)ds} dE, \tag{3.1}$$

where I represents the intensity of the photon beam detected by the sensor after it passes through the imaged subject, $I_0(E)$ stands for the original intensity of the beam at each energy level, $\mu(E,s)$ represents the linear attenuation coefficient, indicating the likelihood of photon interaction per unit distance at a specific energy level, $\int$ ds represents the integral along the path traversed by the photon beam, $\int$ dE is the integral across the energy range within which the photon source emits photons. [17]

The likelihood of photon interaction with matter depends on two things - the material the photon passes through and the photon's energy level. This enables us to distinguish between materials within a scan. To grasp this concept better, we should first look into the definition of a HU (Hounsfield unit). [17]

HU is a dimensionless unit used in CT. It expresses the attenuation for the material scanned, scaled relative to the difference between the linear attenuation coefficients of water and air. Biological materials usually range anywhere from -700 HU for a lung to 1800 HU for bone tissue [17]. It is defined as:

$$HU = 1000\frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}, \tag{3.2}$$

where µstands for the linear attenuation coefficient of the material, μ_{water} and μ_{air} stand for the linear attenuation of water and air.

As mentioned above, computed tomography measures the transmission of the imaging source at many different angles. This can be achieved either by rotating the subject or the X-ray tube and detector. This approach produces a set of projection data per imaging angle. The collection of the data is called a sinogram. A sinogram is later used to reconstruct a 2-D image. [17]

The practical application of a CT scan follows the principles we've previously outlined. The patient is positioned on a table within the CT machine gantry, where an X-ray source and detector are situated, facing each other. These components rotate around the patient, capturing projection data from various angles. After each rotation, the table moves slightly through the gantry, and another set of projection

data is acquired. This process is repeated until the desired area of the body has been scanned. [17]

In technical terms, the CT machine includes several key components, with the three primary ones being the X-ray tube, the detector, and the collimators. The X-ray tube functions as the source of photons, while collimators shape the photon beam to the desired geometry for imaging. Additional filters may be used to eliminate lower energy photons, which are less diagnostically relevant. [17]

3.3 Dual energy computed tomography

As the name suggests, dual energy computed tomography uses two different X-ray energy spectra. This allows us to record energy-dependent changes in attenuation. Different materials exhibit a specific change in attenuation at different energy levels. This fact allows us to better characterize the examined tissue. The implementation of different energy spectra can be accomplished through various methods, classified as either dual-source or single-source dual-energy computed tomography techniques. Dual source CT uses two detectors and two sources, positioned opposite each other in corresponding pairs of source-detector setups. This method allows high-energy and low-energy spectra to be scanned at the same time [18, 19]. A typical setup for dual source CT is shown in Fig. 3.1.

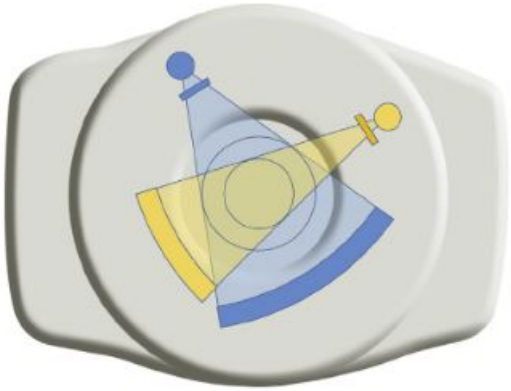


Fig. 3.1: Dual source CT [18].

Unlike dual-source dual energy CT, single-source approach uses a single pair source-detector. There are different types of single-source dual energy CT. Probably the most basic and the easiest to understand uses rapid kVp switching, shown in Fig. 3.2. The source is capable of rapidly switching between high and low energy. The sampling frequency can be as fast as 50 microseconds. [18]

Another possibility is to use a layered detector. In this configuration, the detectors comprise two scintillator layers, each with the highest sensitivity at different photon energies. During the scan at high energy, the top layer of the detector absorbs low-energy photons, while the bottom layer absorbs the remaining photons. It's worth noting that there are additional methods in dual-energy computed tomography that we have not discussed. [18]

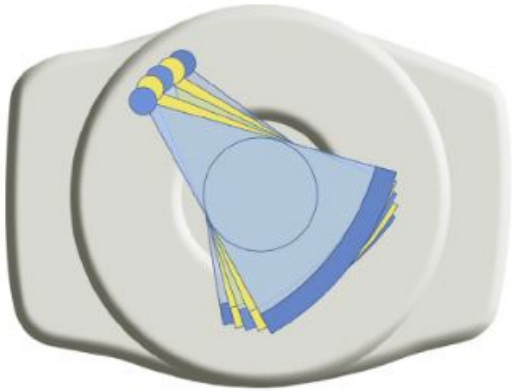


Fig. 3.2: Single source CT, using rapid kVp switching [18].

Dual energy computed tomography can be used to acquire so called virtual monoenergetic images, or VMI for short. Virtual Monoenergetic Images (VMI) in Computed Tomography (CT) involve reconstructed images that replicate the anatomical appearance as though the X-ray beams employed in imaging were of a singular energy level. The capability of virtual monoenergetic imaging enables clinicians to produce images as though they were obtained using X-rays with a specific energy level, commonly known as a "monoenergetic" beam. VMIs have some undeniable advantages, such as material differentiation, dose reduction, and artifact reduction [20] [21].

The technique of virtual non-calcium (VNCa) has gained popularity for its capacity to eliminate calcium from anatomical structures, leading to improved evaluation of various pathological conditions that might be obscured in standard CT scans [22]. Its primary benefit lies in significantly enhancing the image contrast of structures typically obscured by calcium deposits, such as calcified vessels or bone marrow. VNCa imaging is effective in portraying a wide range of traumatic, inflammatory, infiltrative, and degenerative disorders affecting both the spine and the appendicular skeleton. VNCa imaging represents a major advancement in Dual-Energy CT (DECT), allowing the visualization of conditions and disorders that typically involve the use of pricier and more time-consuming methods like MRI, PET/CT, or bone scans [22].

Another post-processing technique is virtual non-contrast imaging (VNC). Virtual Non-Contrast (VNC) imaging provides a non-invasive option that removes the necessity for contrast injection while still delivering diagnostically valuable images. The method utilizes the distinct attenuation characteristics of different tissues across varied energy spectrums to distinguish and digitally remove the contrast medium from the captured images. This procedure produces virtual non-contrast images, portraying the anatomy as though no contrast agent had been administered [23].

4 Radiomics in medical imaging

Radiomics is closely tied to medical imaging, offering a quantitative approach to describing medical images. While medical images may contain information invisible to the human eye, this data can hold diagnostic significance and frequently serves as a crucial factor in timely and appropriate patient treatment. [3]

Radiomics aims to enhance the existing data using mathematical analysis. According to the Image Biomarker Standardisation Initiative (IBSI), features can be categorized into a few classes: intensity based-statistical, intensity histogram-based, intensity-volume histogram-based, morphological features, local intensity, and texture matrix-based features. [3]

There are a few important steps that need to be taken in order to get the most information from medical image data. Every step is an important part of the radiomics pipeline.

4.1 Image segmentation

The first thing that we need to decide on is the region of interest (ROI) in 2-dimensional data or the volume of interest (VOI) in 3-dimensional data. This basically specifies the region from which desired features will be extracted and computed. Image segmentation can be performed manually, semi-automatically, or fully automatically. Despite the widespread use of manual and semi-automatic methods, they come with inherent drawbacks. Not only are they time-intensive, but they also introduce observer bias, potentially leading to subsequent issues. [3]

Considering these drawbacks, fully automated image segmentation seems like the best option. And in many cases it holds true. However, there are instances where the generalizability of these algorithms is a limitation. This leads to incorrect results when applying the algorithms to different datasets. [3]

4.2 Feature extraction

Following the initial two stages, we proceed to feature extraction. Features are numerical values utilized to quantify the grayscale level within our designated ROI/VOI. However, an essential step preceding feature extraction is feature selection. The potential number of features extractable from an image spans from a few to theoretically limitless. Selecting a higher number of features results in fewer cases in each group and vice versa. The challenge lies in determining not only the appropriate number of features but also in excluding redundant and irrelevant ones. According to [3], some rules define the optimal number of features for a given sample size.

However, none of these rules have been properly tested and there is no evidence in any literature that they work. [3]

There are a few steps that have been performed in several clinical studies and that proved to be accurate. The first step would be to eliminate any non-reproducible features. Features that suffer from any observer bias are not informative and it is best to exclude them right in the beginning. Another step would be to calculate the correlation coefficients. This should ensure that the features that we are left with are reproducible and robust. Another thing to keep in mind is that we want to select the best features for the respective task. This means that we may be left with many features, but only a few of them might be relevant in reaching our goal. [3]

4.3 Features used in image processing

Understanding the different classes of radiomic features isn't mandatory for reading radiomics research papers or conducting related studies. Nonetheless, understanding the fundamental principles can assist in interpreting results and selecting features tailored to specific applications. [24]

Radiomic features can be broadly categorized into statistical features, which include histogram-based and texture-based attributes; model-based characteristics; transform-based properties; and shape-based attributes. [24]

Histogram features

The most basic statistical descriptors rely on the overall gray-level histogram and encompass gray-level mean, maximum, minimum, variance, and percentiles. As these features stem from the analysis of individual pixels or voxels, they are referred to as first-order features. Advanced features encompass skewness and kurtosis. [24]

Texture features

Texture analysis and radiomics help uncover hidden information within images that the human eye cannot see. It includes features such as absolute gradient, gray-level co-occurrence matrix, gray-level run-length matrix, and others. Texture analysis is a helpful tool in the diagnosis and prognosis of cancer. Some would even call it a "virtual biopsy". [25, 26]

Model-based features

Model-based analyses seek to interpret spatial gray-level data to define and describe objects or shapes. A model that generates texture is created and adjusted to the region of interest. Its calculated parameters are utilized as radiomic features. For example, the autoregressive model operates on the premise that a pixel's gray level depends on the weighted sum of its neighboring pixels: left, top-left, top, and top-right. [24]

Transform-based features

Transform-based features in radiomics involve manipulating the image data through various mathematical transformations or operations to extract meaningful information. These transformations can include techniques like wavelet transforms, Fourier transforms, or other mathematical functions applied to the image pixel values. The resulting transformed data or coefficients can then be used as features to capture specific characteristics or patterns within the image, aiding in the analysis and interpretation of medical imaging data for diagnostic or prognostic purposes. [24]

Shape-based features

Geometric attributes known as shape-based features characterize the form of regions of interest. While some features, like 2D and 3D diameters and their ratios, are straightforward in concept, others involving mesh-based approaches using polygons such as triangles and tetrahedrons add complexity. Features like compactness and sphericity depict how the shape of a region of interest deviates from that of a circle (in 2D) or a sphere (in 3D). [24]

4.4 Feature reduction

After implementing feature extraction methods, we are usually left with a large number of features, not all of them being relevant to reach our goal. There are a few approaches that are useful in selecting only the most relevant set of features. Let's take a brief look at some of them.

Feature selection using statistical tests

If we want to talk about statistical tests, it is best to make some terms clear, that will help our understanding:
Null hypothesis - This hypothesis asserts that there is no significant difference between the sample and population, or among different populations, represented as

H0. [27]

Alternate hypothesis - The alternative hypothesis represents the statement that contradicts the null hypothesis. It's symbolized as H1. [27]

Critical value - It marks a point on the test statistic scale where we reject the null hypothesis. The higher the critical value, the less likely it is that two samples belong to the same distribution. [27]

p-value - The p-value, short for "probability value", denotes the probability of a result occurring solely by chance. In hypothesis testing, it aids in either supporting or rejecting the null hypothesis. A lower p-value indicates stronger evidence for rejecting the null hypothesis. [27]

Now that we have set a foundation of basic understanding, we can better understand how statistical testing works. A statistical test serves as a method to decide if a random variable follows the null or alternative hypothesis. It essentially shows us whether there are notable distinctions between a sample or multiple samples and a population. Descriptive statistics such as mean, median, mode, range, or standard deviation are used for this purpose, with the mean often being the preferred metric. After running the statistical test, it provides a numeric outcome, which is then compared against the p-value. If the obtained value surpasses the p-value, the null hypothesis is accepted; otherwise, it is rejected. [27]

There are many statistical tests to be chosen from, but we will go over only a few of them.

Correlation test - A correlation test evaluates the extent of connection between variables, bearing in mind that this assessment necessitates variables to be continuous. Several techniques are available for conducting correlation tests, including Covariance, Pearson correlation coefficient, Spearman rank correlation coefficient, among others. [27]

Z-test - A Z-test examines whether the means of two given samples are from the same distributions, but it's not applicable for sample sizes below 30. A Z-test can be either a one-sample Z-test or a two-sample Z-test. The One-Sample Z-test determines whether the sample mean significantly deviates from a known or hypothesized population mean. Conversely, the two-sample Z-test compares two independent variables. [27]

T-test - Similar to the Z-test, a t-test compares the means of two provided samples, but it's employed particularly when the sample size is less than 30. The t-test works on the premise of a normal distribution within the sample and can be employed as either a one-sample or a two-sample test. [27]

Data visualization can be a helpful tool in feature reduction as well. However, it helps to reduce the number of dimensions using statistical tests beforehand. Once the data's dimensionality has been decreased using statistical tests, the step of data

visualization becomes highly significant. [24]

Principal component analysis

Principal component analysis (PCA), stands as a highly effective technique within feature compression. Classified as a biometric technique, it falls under the statistical domain in this field. Its core objective is to reduce extensive spatial dimensions into more compact and manageable spaces. [28]

Introduced by the renowned mathematician Karl Pearson in 1901, PCA finds utility in predictive modeling, particularly in face recognition, and in exploratory data analysis systems. Acknowledged as a genuine eigenvector-based multivariate analysis, PCA is considered to be the most streamlined method in this domain. Its purpose centers on reducing the dimensions of transformed data by employing fundamental components to reach a compressed representation resembling shadow-like images, thereby presenting a condensed yet informative depiction of the original data. [28]

PCA excels in discerning data variations by unveiling their underlying composition. The quantity of original variables, like face images, often matches or surpasses the number employed in PCA transformations. [28]

Random forest classifier

A random forest is a machine learning technique that combines multiple decision trees to perform tasks like classification or regression. Instead of relying on a single tree, it creates an ensemble of trees, each trained on different subsets of the data and features. By allowing individual trees to vote or contribute predictions, the random forest produces more accurate and robust results, especially in handling complex datasets with noise and high dimensionality. This method is widely used across various fields due to its effectiveness in making predictions and its ability to reduce overfitting compared to single decision trees. [29]

5 Current trends in feature extraction methods

As talked about previously, many different features can be extracted from an image. This fact leads to different researchers and research groups choosing different paths, which can lead to different results and can give us an insight into how feature extraction works. Let's take a brief look at their approaches.

When it comes to tumor imaging methods, magnetic resonance imaging (MRI) is considered to be one of the most effective. MRI images have high contrast properties that are very much appreciated in tumor imaging [30].

Various authors used features extracted from MRI to detect malignant tumors and the stage of the cancer. There are different strategies proposed by different authors. Reedy et al. in their study use MRI images to detect bone cancer [31]. First of all, they remove the noise from the image [31]. Later, they used mean pixel intensity to predict the stage of the cancer. [31]

Sinthia and Sujatha take a different approach [32]. They focus on edge recognition. Using Sobel edge identification, they are able to distinguish the tumor zone. [32]

Similarly to this manner, Asuntha et al. also implement edge detection [33]. However, they go a step further and extract more features that are used to identify bone cancer [33].

Avula et al. use mean pixel power to distinguish the malignant growth from MRI images. This mean intensity value is crucial for distinguishing between cancerous and non-cancerous areas. Based on experiments conducted with 400 images, they have found that the mean intensity values indicative of malignant tumors typically range from 234 to 250 [34].

As mentioned in [35], feature extraction can also be used to distinguish between malignant and benign growths. For this purpose, the authors used texture features together with K-means clustering algorithm.

Paper written by Ashish Sharma et al. combines feature extraction with a classification model [36]. They state that the textures of cancerous bones differ from the healthy ones and the pixel of the cancerous bone is more scattered. They perform an experiment that leads to the realization that texture features are not sufficient to achieve their goal. Based on their findings they combine the texture features with skewness and entropy. The value of entropy is much lower in the cancerous region compared to healthy tissue. [36]

However, as we read in MRI is not the only medical imaging technique that uses the principles of radiomics. The authors of [36] focus on detecting bone cancer

through a machine learning model that relies on feature extraction from X-ray images. In their study, they discovered that using support vector machine (SVM), the GLCM (gray-level co-occurence matrix) texture features by themselves are not enough to classify bone cancer. Thus, additional features were used to detect and categorize cancerous bones, such as histogram of oriented gradients (HOG) features. HOG assesses pixel shape and orientation within local image cells, aiding in the identification of cancerous regions. The experiment revealed that integrating the HOG feature with the GLCM texture feature resulted in an accuracy score of 92.68 %, surpassing the 87.80 % achieved without the HOG feature. [36]

The authors of [37] discuss the importance of radiomics in oncology, focusing on spine surgery. They mention the undeniable advantages that radiomics offers in detecting pathologies and treatment prediction. However, they are aware of the limitations radiomics face, such as reproducibility, model overfitting, and class imbalances. Many of these setbacks can be solved with a good understanding of the available dataset. They come to a conclusion that in the field of personalized medicine, spinal radiomics presents clinicians with a promising tool for decision-making. Yet, it's crucial to undertake additional validation on high-quality datasets before integrating these models into clinical practice. Similar to many machine learning applications, future efforts in spinal radiomics are necessary to enhance the transparency of these algorithms. This will enable better clinical integration and the creation of innovative clinical insights. [37]

6 Available datasets

6.1 Publicly available dataset

The LumVBCanSeg dataset [38], a publicly available collection focused on Lumbar Vertebral Body Cancellous Bone Segmentation, was utilized for both training and testing purposes. This extensive dataset comprises 185 lumbar CT scans obtained from various CT scanners, including those manufactured by Philips and Siemens. All scans were sourced from ShengJing Hospital of China Medical University. The dataset intentionally excludes cases involving vertebral fractures, metal implants, bone tumors, and foreign materials. To ensure consistency, all data underwent resampling to achieve an isotropic resolution of 1×1×1 mm. In addition to the CT scans, the dataset includes corresponding segmentation masks. Annotations cover the bodies of five lumbar vertebrae, L1 to L5, and are sequentially labeled from 1 to 5 [38]. As seen in Fig. 6.1, the mask does not include the periostenum and processes.

Fig. 6.1: An example of lumbar CT scan with the corresponding segmentation mask from publicly available dataset.

6.2 Clinical data

In order to collect clinical data, we completed an internship at The University Hospital Brno (Clinic of radiology and nuclear medicine). The data was gathered from patients being treated for multiple myeloma. Data was anonymized to make it usable for the purposes we intended. For the bachelor's thesis, we worked with a dataset of 20 patients - 10 individuals with no spinal pathologies and 10 patients with multiple myeloma. Out of the healthy patients, 5 have a full body scan and 5 only have scans of their thoracic spine. This will be important for our tasks later on. In addition to these scans, we have acquired scans of 5 MM patients who have one follow-up scan available. The follow-up scans will be later used for volume analysis of the lesions.

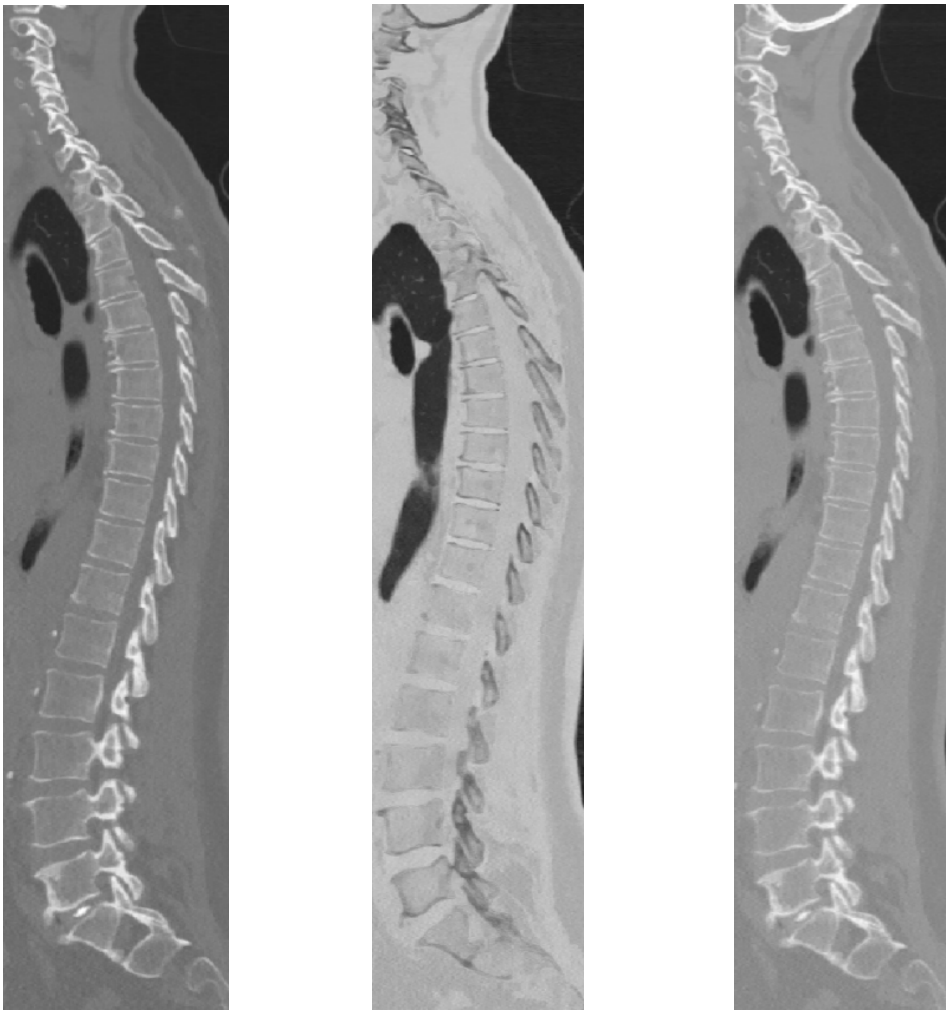
Data were acquired following approval from the Ethics Committee, under the application registration number NU23J-08-00027. All patients gave their consent after being informed. The data were acquired utilizing the Philips Healthcare IQon spectral CT system in collaboration with The University Hospital Brno, Clinic of radiology and nuclear medicine. The scanning parameters included a peak tube voltage of 100 kV, tube current of 10 mA, matrix size of 512×512, and a slice thickness of 0.9 mm using a sharp reconstruction kernel and hybrid iterative reconstruction technique (iDose4, set to level 4). Scans of the myeloma patients covered the region from the head to the knee. The scans were examined using a dedicated workstation (Intellispace Portal version 12.1; Philips Healthcare) by two independent readers, with at least one being board certified. Patients were diagnosed with MM based on elevated monoclonal immunoglobulin in the blood and an increased plasma cell count in the bone marrow.

Before completing feature reduction, we wanted to become more familiar with the dataset. For this purpose, we used the Medical Imaging Interaction Toolkit (MITK) workbench for data visualization. MITK [39] is an open-source software framework specifically designed for developing interactive medical image processing applications. MITK provides a platform that facilitates the visualization, analysis, and processing of medical image data.

Parametrical images

For each patient, there are a few different scans. First of all, we have conventional CT scans, which are available in both NIfTI and DICOM data formats. We decided to work with NIfTI files which proved to be easier to work with. Then, for each patient, we have three virtual monoenergetic images, acquired at 40 keV, 80 keV, and 120 keV. Lastly, we have images with calcium suppression. These images are available with different levels of calcium suppression (25%, 50%, 75%, 100%). Having

this much information gives us extensive possibilities for data analysis. Examples of parametrical images are shown in Fig. 6.2.



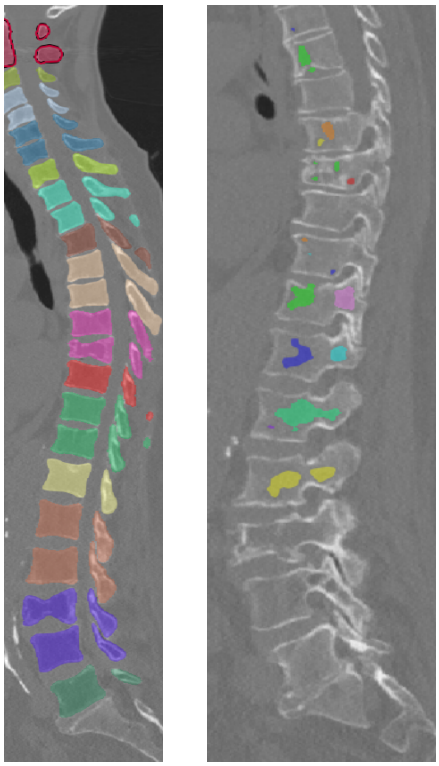
(a) (b) (c)

Fig. 6.2: Example of parametrical images, MM patient, (a) - conventional CT, (b) - 50 % calcium suppression, (c) - virtual monoenergetic image at 120 keV.

Spine labels and lesion labels

Adding to CT images, the dataset also contains spine labels from different segmentation algorithms, for both healthy individuals and MM patients. Used segmentation algorithms are further discussed in [40]. Each of these labels has a unique ID that makes it easy to extract only one vertebra. Using MITK workbench, we can visualize the data which gives me a better understanding of its nature and helps with future analysis. In multiple myeloma patients, each lesion has an assigned ID,

where the segmentation masks were corrected by a radiologist, to ensure correct segmentation. Both spine and lesion labels are shown in Fig. 6.3.



(a) (b)

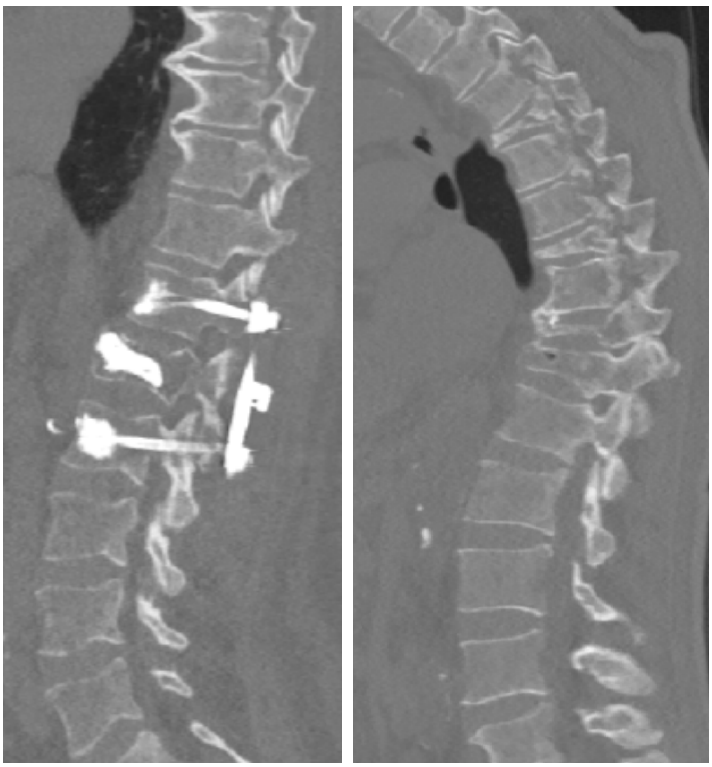
Fig. 6.3: Example of spine labels (a) and lesion labels (b).

Examples of problematic patients

Even though most of the scans and respective segmentation were accurate, there were a few individuals that may cause problems later down the path in our analysis. Because of this, we have decided to only use the scans which were of good quality and the segmentation seems to be done correctly.

One of the myeloma patients has a spine implant. In Fig. 6.4, we can see that the implant shows much more than the other structures. This can lead to inaccurate features. This is why these vertebrae were excluded from further analysis.

Another example would be patients with vertebrae so affected by MM that they literally seem to be falling apart, their structure is very damaged. As we can see in Fig. 6.4, some of the vertebrae appear very thin and are almost non-existent. These were also excluded and not used in our analysis.



(a) (b)

Fig. 6.4: Example of problematic vertebrae, (a) - metallic implant, (b) - damaged vertebrae.

7 Analysis of vertebrae

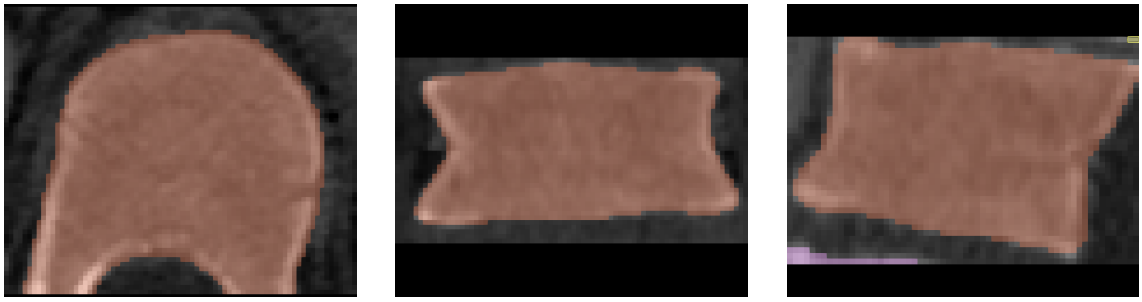
7.1 Dataset preparation

Before doing any analysis, the first step was to prepare the dataset. Even though we have more than enough data, working with whole-body scans is time-consuming and computationally demanding. Because of this, we extracted only the regions which will be later used to extract features. Because the goal was to analyze only the vertebral bodies without processes, the vertebrae had to be extracted accordingly. There were two approaches that we took, each one offering some advantages while also having some imperfections.

7.1.1 Manual extraction

The first approach was to manually extract only the body of the vertebrae. This was not only time-consuming but also unreproducible. The extraction was accomplished by forming a bounding box around the vertebral body and extracting it.

The problem was that the processes were hard not to be included in the bounding box. It was almost impossible to ensure that each vertebra has the same amount of processes extracted with them, or not have any portion of it at all. Another issue might be that the periostenum appears at higher intensity and this may also cause an error. Even though we are aware of these issues and the errors they can cause, we still decided to go ahead and continue our tasks. Fig. 7.1 shows an example of manually extracted vertebrae, with partially extracted processes.



(a) (b) (c)

Fig. 7.1: Vertebra extracted with processes, (a) axial view, (b) coronal view, (c) sagittal view.

After extracting these vertebrae, we started noticing these irregularities. We tried to only use the vertebrae that were extracted correctly. However, even when

they are selected correctly, we can't ensure that all of them were extracted with the same portion of processes.

7.1.2 Using nnU-Net deep learning model

nnU-Net [41] model was trained for segmentation of trabecular tissue of the lumbar spine. In contrast with manually extracting the vertebral bodies, using nnU-Net model for segmentation method only segments the trabecular part of the vertebrae and does not include any part of the processes. This means that we do not have to take into account human error when extracting the vertebrae. The model was trained on the publicly available dataset LumVBCanSeg [38]. Its accuracy was calculated using the Dice coefficient and Hausdorff distance. Later, it was tested on our clinical data.

Model implementation

nnU-Net represents an automated semantic segmentation technique crafted to adapt to a wide array of datasets. This implies that, even in the absence of specialized knowledge, the model can be effortlessly trained and applied across various applications. It particularly excels in semantic segmentation, showcasing proficiency in processing both 2D and 3D images across diverse input modalities and channels. Its adaptive nature extends to accommodating variations in voxel spacings and anisotropies, demonstrating resilient performance even in cases with substantial class imbalances. nnU-Net employs supervised learning, necessitating the submission of a set of training cases tailored to the specific application. [42]

As previously stated, the training dataset only contains CT scans of the lumbar region. Consequently, to align with the desired outcomes, the data from the clinical dataset had to undergo cropping to achieve the intended size. This involved creating a bounding box with the specified dimensions. Fig. 7.2 shows the original data after being cropped, showing both spine without any pathologies and multiple myeloma lesions.

The dataset was divided in 80:20 ratio, with 80 % of the dataset being used for training, leaving the 20 % for testing purposes. This means that we got 143 scans in the training set and 37 scans were used for testing purposes. The optimal architecture and learning parameters for nnU-Net were automatically determined. The configuration was specified as 3D full resolution, with an initial learning rate of 0.01 that gradually decreased during training. The batch size was set to 2, patch size was 128×128×128 px, featuring a kernel size of 3×3×3.

A total of 1000 learning epochs were carried out. Planning and testing was done on a school computer CUDA 2 with following parameters:

- graphics card: Nvidia Titan Xp
- processor: Intel Xeon E5-2603v4
- RAM: 64 GB
- operating system: Linux Ubuntu 22.04

Training of the model took place on a dedicated computer at the Metacenter, using the codes from [43].

(a) (b)

Fig. 7.2: Scans after being cropped, (a) - no spinal pathologies, (b) - MM patient.

The lumbar spine segmentation was assessed using the Dice coefficient [44], a metric employed to measure the similarity between two sets. This coefficient can be computed for binary or multilabel classification. In our scenario, we opted for the multilabel approach due to the nature of the available data. The result is determined using the following formula:

$$Dice(P,T) = \frac{1}{N}\sum_{i=1}^{N}\frac{2\,|P_i \cap T_i|}{|P_i| + |T_i|}, \tag{7.1}$$

where P stands for predicted values, T represents the ground truth and i is an index of the N^{th} vertebrae. The final dice score ranges between 0 and 1 and has no unit.

The Hausdorff distance [45] serves as a metric to evaluate the similarity between two sets of points within a specified metric space. This distance is computed

by identifying the maximum distance from any point in one set to the nearest point in the other set. It functions as an indicator of dissimilarity or mismatch between the two sets and finds applications in various fields such as computer vision, image analysis, pattern recognition, and shape matching. Additionally, it can be applied in either binary or multilabel form. In the assessment of our model, the multilabel Hausdorff distance was employed:

$$HD(A,B) = \max(\max_{a\in A}\min_{b\in B} d(a,b), \max_{b\in B}\min_{a\in A} d(a,b)[px], \tag{7.2}$$

where $HD(A,B)$ represents the Hausdorff distance of sets A and B, the symbol max represents the maximum, the symbol min represents the minimum, $d(a,b)$ represents the distance of point a and point b, where it can be for example the L2 norm, the symbol $a \in A$ indicates that the variable a takes values from the set A and the symbol $b \in B$ indicates that the variable b takes values from the set B. In our case, we calculated the HDs for each vertebral mask separately and then calculated their average value to evaluate the success of the segmentation model.

Results

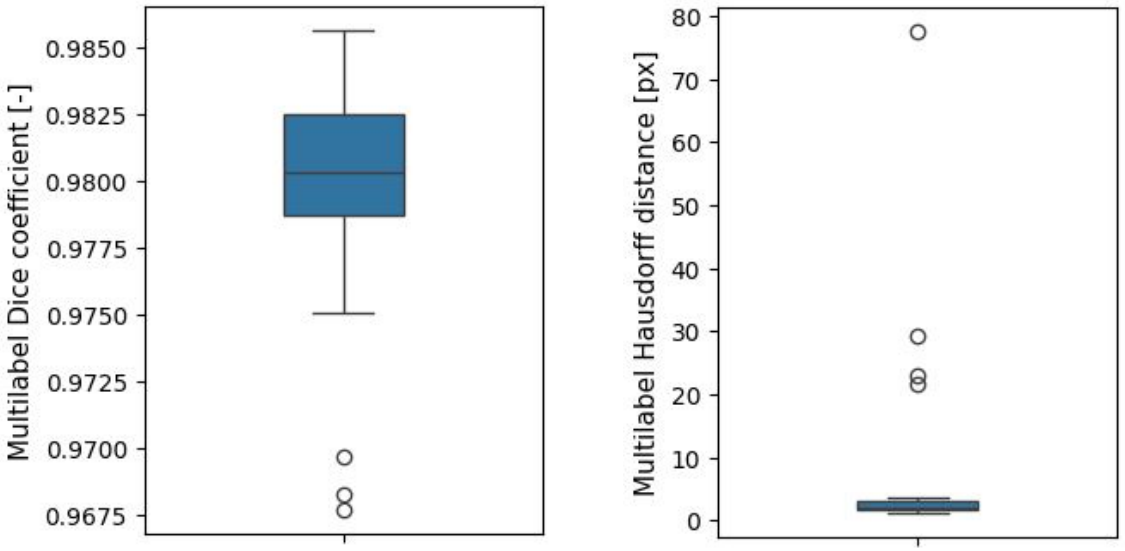
The performance of the trained model was assessed using the testing database, comprising 37 images. After a brief examination of the outcomes, it can be concluded that the model demonstrated satisfactory performance. Table 7.1 provides a summary of the computed metrics. Further inspection of the results revealed a few outliers, characterized by significant differences in both metrics compared to the remainder of the testing dataset.

Metric	Dice coefficient	Hausdorff distance
Minimum value	0.526	1.166
Maximum value	0.986	173.816
Mean	0.949	26.189
Median	0.980	2.159
Standard deviation	0.103	48.527

Tab. 7.1: Calculated metrics.

Fig. 7.3 depicts the outcomes through box-and-whisker plots. It is evident that other than a few outliers, the model exhibited satisfactory performance. To enhance visualization, the six most prominent outliers were excluded.

After using the neural network on our clinical data, we received segmentation masks for the lumbar portion of the spine for each patient that had whole body CT scan available. This made it easier to select vertebra for our analysis.

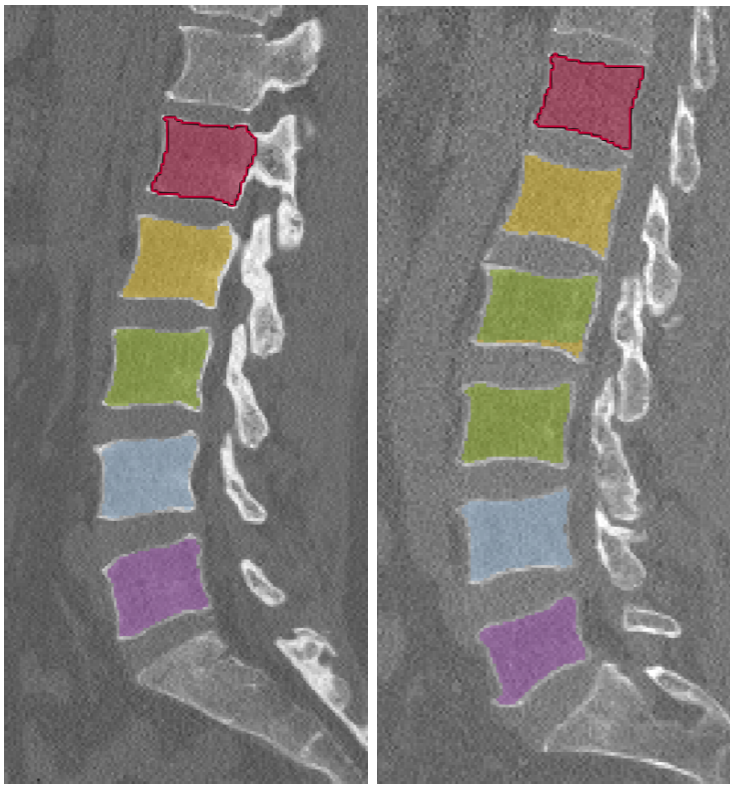


(a) (b)

Fig. 7.3: Box-and-whisker plots showing the results, (a) - Dice coefficient, (b) - Hausdorff distance.

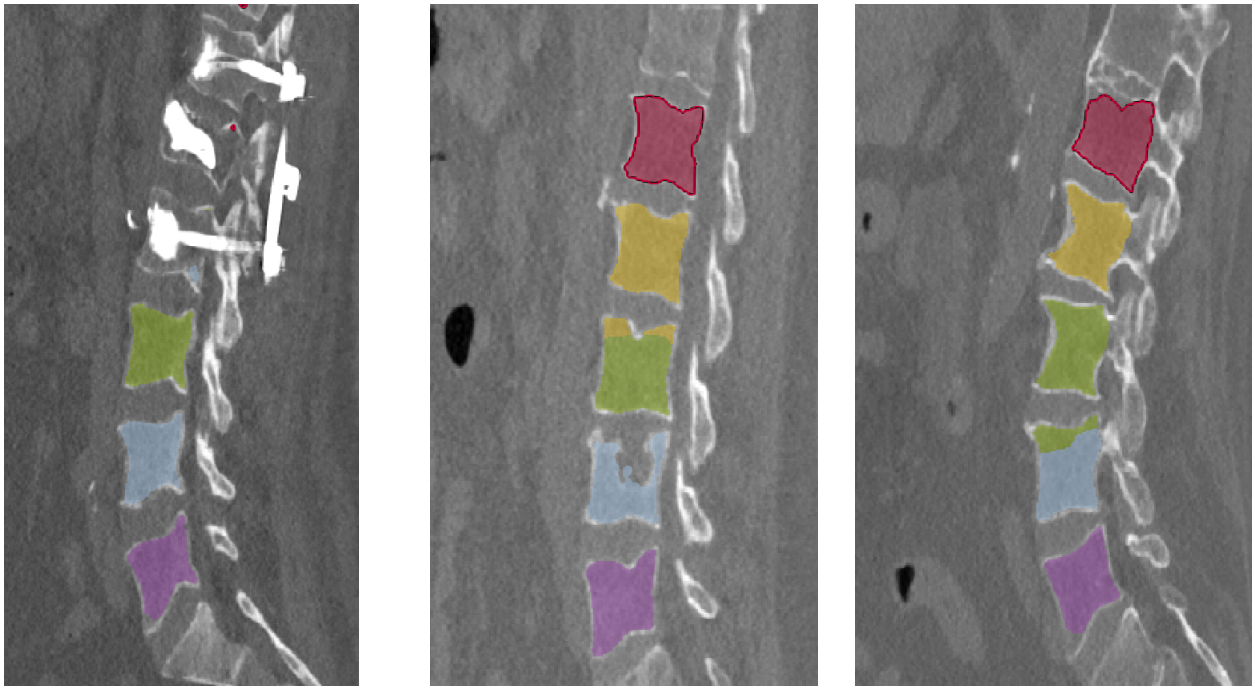
Fig. 7.4 shows us two results of the segmentation on the testing dataset. The failed segmentation shows that instead of 5 lumbar vertebrae, 6 vertebrae were segmented, including one thoracic vertebra, and that two of the vertebrae were assigned the same label.

The model was then applied on the clinical data. The results were evaluated visually. However, because the neural network was only trained using healthy individuals, and scans containing any abnormalities were omitted, we saw much more accurate results with the healthy vertebrae. When it came to MM patients, we saw a decline in the accuracy of the neural network. In the examples below we can see examples of incorrect vertebrae segmentation in MM patients. The segmentation inaccurately captured vertebrae with anomalies, mostly from MM patients, while the healthy vertebrae were segmented correctly, see Fig. 7.5 and Fig. 7.6.



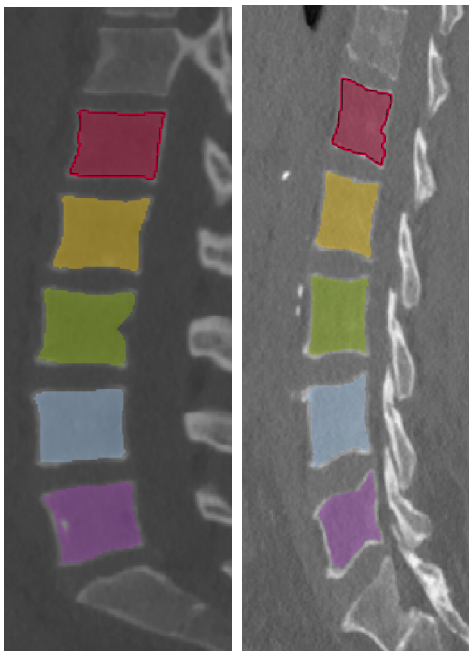
(a) (b)

Fig. 7.4: Results of the segmentation by nnU-Net on the testing dataset, (a) - successful segmentation, (b) - failed segmentation.



(a) (b) (c)

Fig. 7.5: Incorrect segmentation examples from clinical dataset.

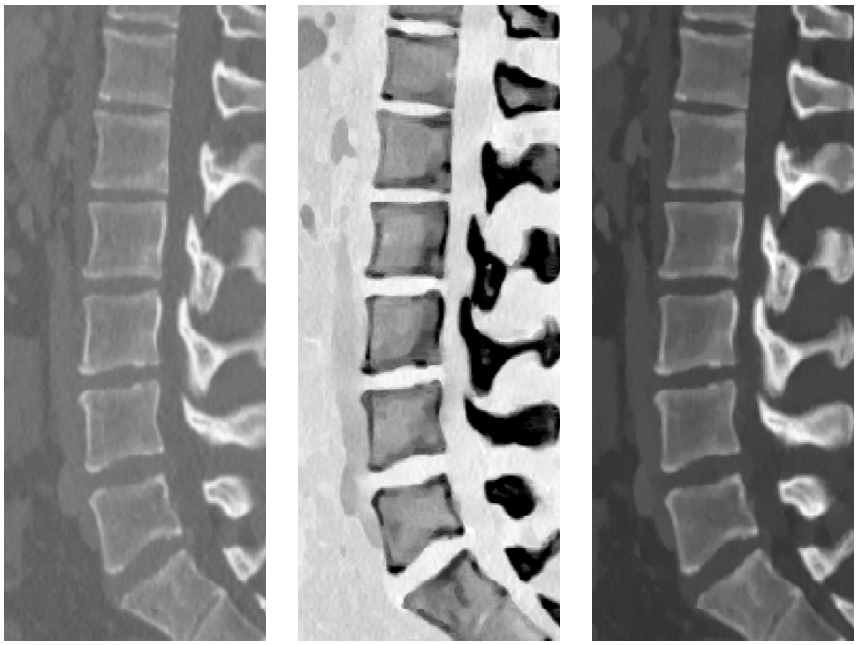


(a) (b)

Fig. 7.6: Correct segmentation examples from clinical dataset.

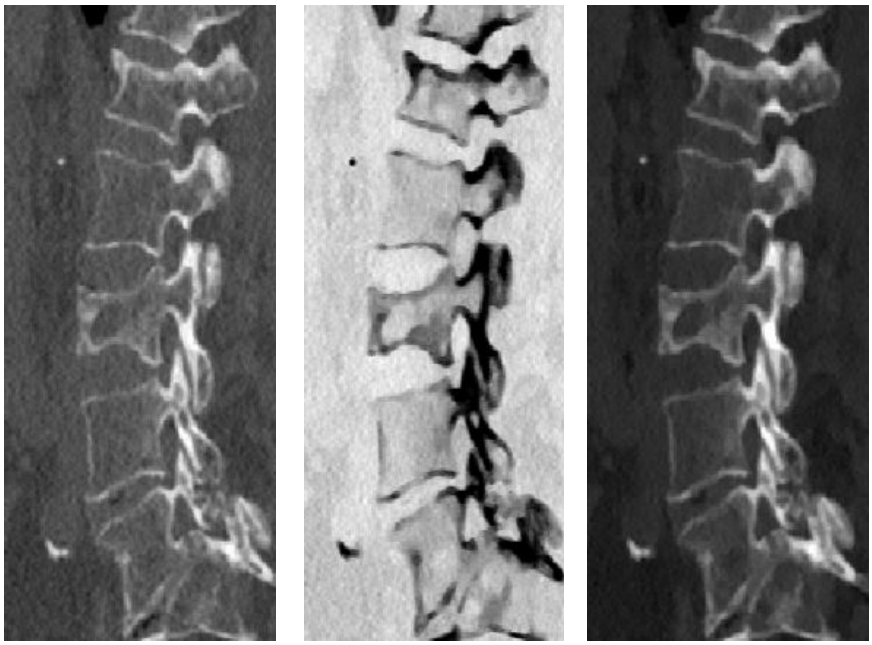
7.2 Statistical analysis

In this part of the thesis, we perform basic statistical analysis. We use the lumbar vertebrae segmented using nnU-Net. Statistical features (mean, median, standard deviation) were extracted from the trabecular part of the vertebral body. These features were compared between conventional CT images and different parametrical images. The results were visualized using box-and-whisker plots. An example of the available parametrical images using only the lumbar part of the spine is shown in Fig. 7.7 and Fig. 7.8.



(a) (b) (c)

Fig. 7.7: Example of different types of images, no spinal pathologies, (a) - conventional CT, (b) calcium suppress at 25 %, (c) - virtual monoenergetic image at 40 keV.



(a) (b) (c)

Fig. 7.8: Example of different types of images, multiple myeloma patient, (a) - conventional CT, (b) calcium suppress at 25 %, (c) - virtual monoenergetic image at 40 keV.

Before moving on to comparing chosen statistical metrics, the values of the whole vertebrae were compared. Conventional CT, images with 25 % calcium suppression and virtual monoenergetic images acquired at 40 keV were selected for the comparison. Three healthy and three affected vertebrae were chosen from conventional CT scans as well as parametrical images. The results are visualized in Fig. 7.9, Fig. 7.10, Fig. 7.11. We can see that, as expected, there is a difference between the two groups. All outliers were omitted for better visualization. The vertebrae were selected from two patients, one without MM and one who was previously diagnosed with MM.

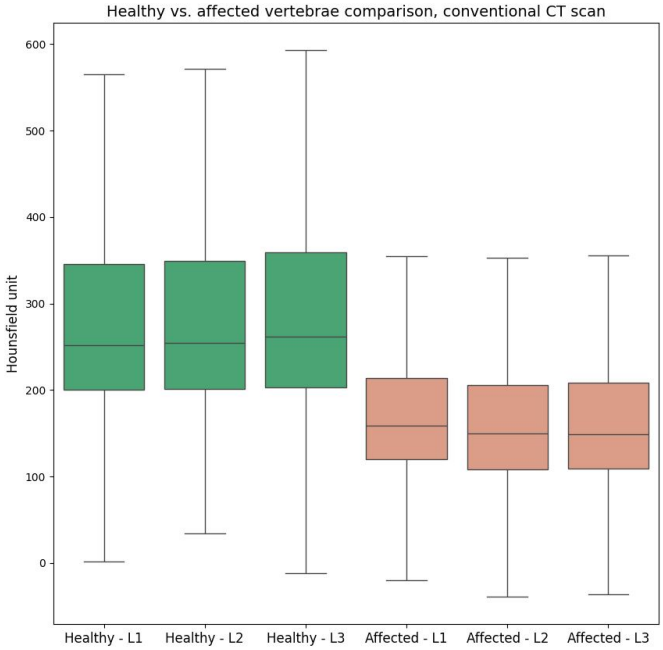


Fig. 7.9: Comparison of healthy and affected vertebrae, lumbar vertebrae, conventional CT scan.

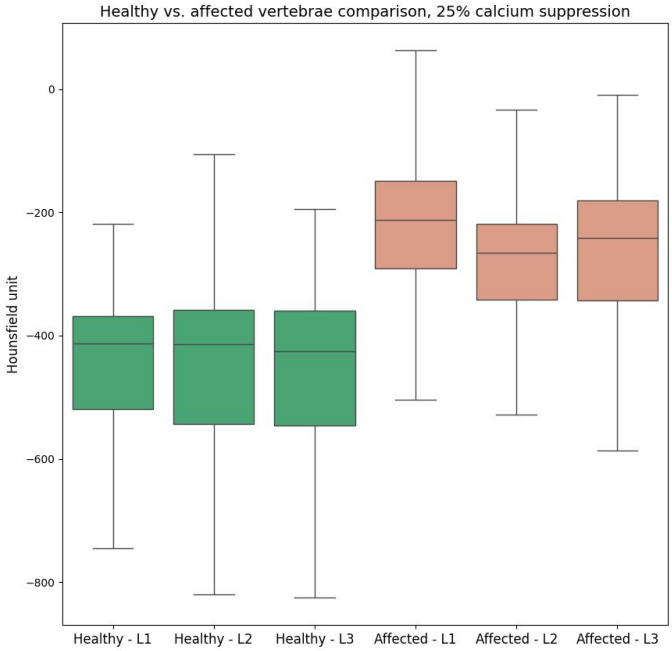


Fig. 7.10: Comparison of healthy and affected vertebrae, lumbar vertebrae, 25 % calcium suppression.

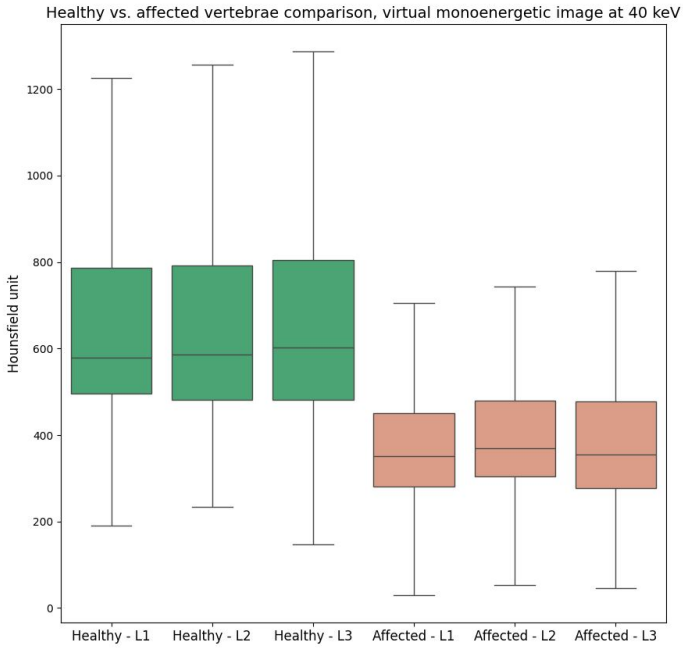
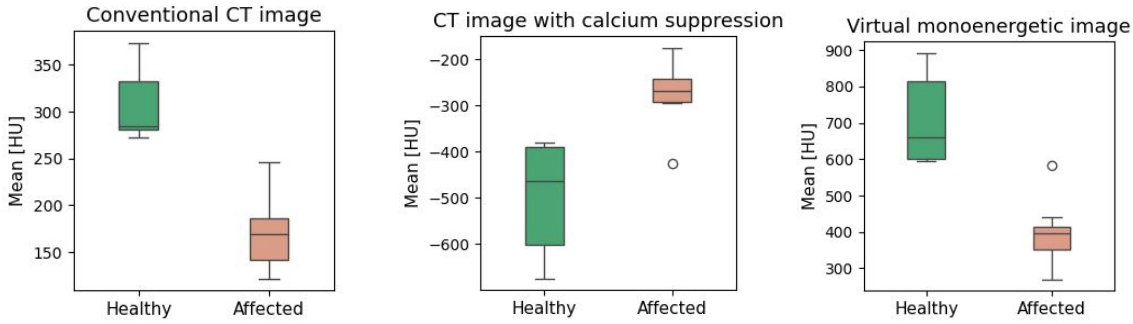


Fig. 7.11: Comparison of healthy and affected vertebrae, lumbar vertebrae, virtual monoenergetic image at 40 keV.

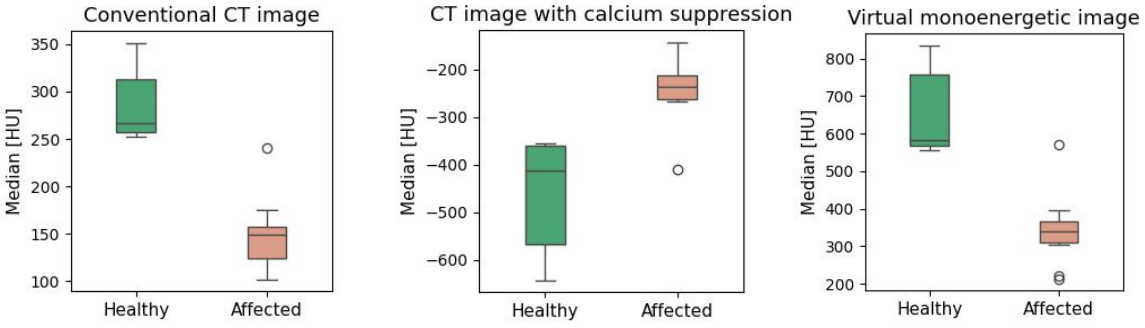
After the initial analysis, 10 healthy and 10 affected vertebrae were selected from the scans using the nnU-Net segmentation. Mean, median, and standard deviation were extracted from each one of them. All the metrics were compared between the pairs selected from the same type of image. Results were again visualized using box-and-whisker plots.



(a) (b) (c)

Fig. 7.12: Box-and-whisker plots showing the difference in mean, (a) - conventional CT image, (b) - CT image with 25% calcium suppression, (c) - virtual monoenergetic image at 40keV.

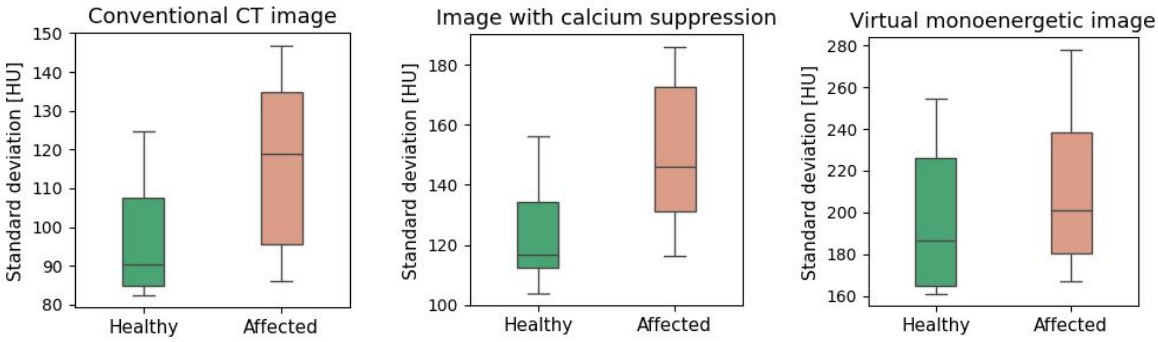
In Fig. 7.12 we see the comparison of mean values among all types of available scans. There is a clear difference between healthy and affected vertebrae in all three cases.



(a) (b) (c)

Fig. 7.13: Box-and-whisker plots showing the difference in median, (a) - conventional CT image, (b) - CT image with 25% calcium suppression, (c) - virtual monoenergetic image at 40keV.

Fig. 7.13 visualizes the differences between the two groups using median values. Just like in Fig. 7.12, there are clear differences and the two groups do not overlap.



(a) (b) (c)

Fig. 7.14: Box-and-whisker plots showing the difference in median, (a) - conventional CT image, (b) - CT image with 25% calcium suppression, (c) - virtual monoenergetic image at 40keV.

Fig. 7.14 visualizes the differences in standard deviation in healthy and affected vertebrae. Unlike in comparing mean and median values, we see that the two groups significantly overlap, which means that the values are similar.

After comparing the box-and-whisker plots between scans with no spinal pathologies and MM patients, slight differences are visible, which was expected. However,

to test whether the differences are statistically significant, further analysis is necessary. To test it, the Mann-Whitney U test was performed to determine if there is a statistically significant difference between the two groups. Mann Whitney U test is a non-parametric statistical. The assumptions for this test are that the two groups of data must be continuous, the groups have no relation to each other, the data is not normally distributed and the sample size is usually more than 5 [46]. This suited our case. The parameters of the statistical test were:

- H_0: There is no difference in the values between healthy and affected vertebrae.
- H_a: The values between the two groups are different.
- p-value = 0.05.

The statistical test was done nine times, to compare all the pairs of respective statistical values. In eight cases, the p-value was less than 0.05, which meant that the null hypothesis was rejected. In the case of comparing the standard deviation of virtual monoenergetic images, the p-value was 0.3447. That meant that the null hypothesis could not be rejected. Taking into account that in a vast majority of cases, the null hypothesis was rejected, we can safely say that, overall, there is a statistically significant difference between the two groups.

Metric	**Conventional CT**	**Calcium suppression**	**VMI**
Mean	0.0002	0.0006	0.0002
Median	0.0002	0.0006	0.0004
Standard deviation	0.0376	0.0211	0.3447

Tab. 7.2: p-value of performed statistical tests, testing the statistical significance of HU of healthy and affected vertebrae.

7.3 Feature analysis

In the first part of the thesis we focused on comparing healthy vertebral bodies with vertebral bodies affected by MM. This was accomplished by extracting radiomic features from the vertebrae that were extracted before. Pyradiomics was used for this task [47]. Pyradiomics is a Python package designed for extracting a large array of quantitative imaging features from medical images. It's particularly used in the field of radiomics, which involves the extraction and analysis of a vast number of quantitative imaging features from medical images, such as MRI, CT scans, or PET scans. These features can include shape, intensity, texture, and

more, providing a comprehensive characterization of regions within these images. [47]

Pyradiomics provides a framework to standardize the extraction of these features, allowing researchers and clinicians to analyze and interpret medical images more quantitatively and systematically. It's often used in research related to tumor characterization, treatment response assessment, and disease prognosis, among other medical applications. [47]

Using Pyradiomics, we extracted radiomic features from all the respective vertebrae. This left us with an overwhelming volume of features. However, not all of them would be useful for reaching our goal.

7.3.1 Manually extracted dataset

First of all, we decided to use PCA and visualize the result to see, whether there would be two unique clusters forming. For this part of our thesis, only the conventional CT scans were used. Before completing PCA, the data had to be standardized to ensure the same variance among in each variable. PCA is very sensitive to large differences between the variances of variables, as large variances will dominate over the variables with smaller ranges [48]. As seen in Fig. 7.15, even though there are a few outliers that differ from the rest of the dataset, there is a difference between the features of the healthy and the affected vertebrae. This led us to believe that there has to be a set of features that are significant when deciding whether the vertebra contains lesions or not.

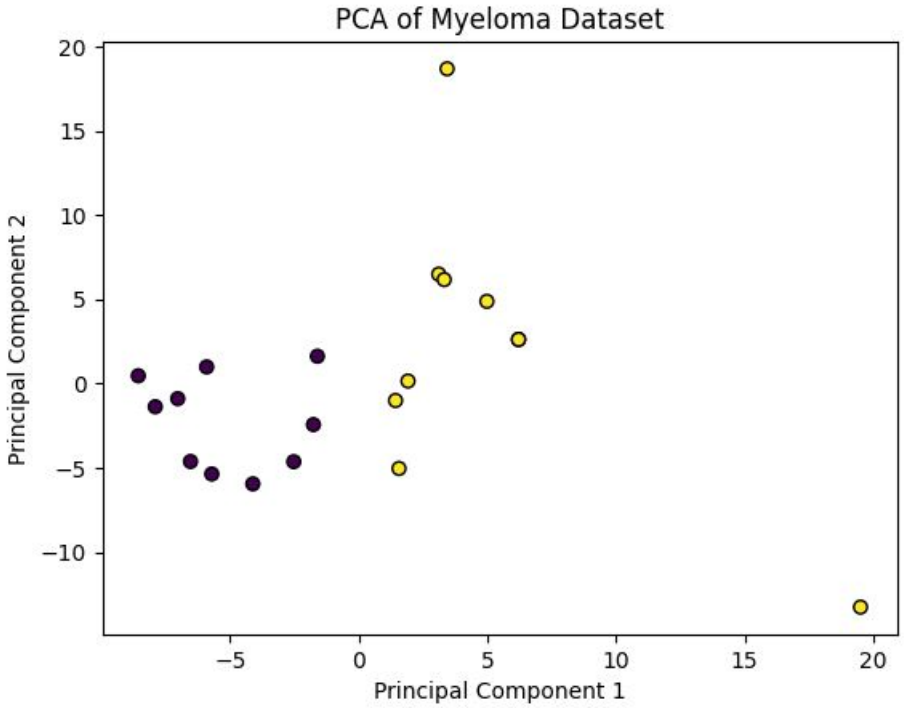


Fig. 7.15: PCA result visualization, manually segmented vertebrae, healthy - purple, MM patients - yellow.

Vertebrae vary in shape and size according to their position in the spine. In the next step, PCA was performed containing precisely selected vertebrae, so that the dataset would contain 5 healthy thoracic and 5 healthy lumbar vertebrae and 5 thoracic and 5 lumbar vertebrae affected by multiple myeloma. The point of this was to explore whether we would get four different clusters forming, one for each group. That would mean that the values of the features are not only influenced by whether the vertebra is healthy or not, but also by its position in the spine. Cervical vertebrae were not included in the analysis. Because of their shape and size, it was difficult to correctly extract them from the images, which could lead to incorrect results.

As seen in Fig. 7.16, both groups of healthy vertebrae can be clearly distinguished. On the other hand, both types of vertebrae that were affected by MM are grouped together, without any clear difference.

Vertebrae extracted using nnU-Net

Using the output from the neural network, we were able to perform an analysis of vertebral bodies again, using more accurate data. While segmenting the vertebral body manually was prone to error, this time we are sure that the segmentation does not

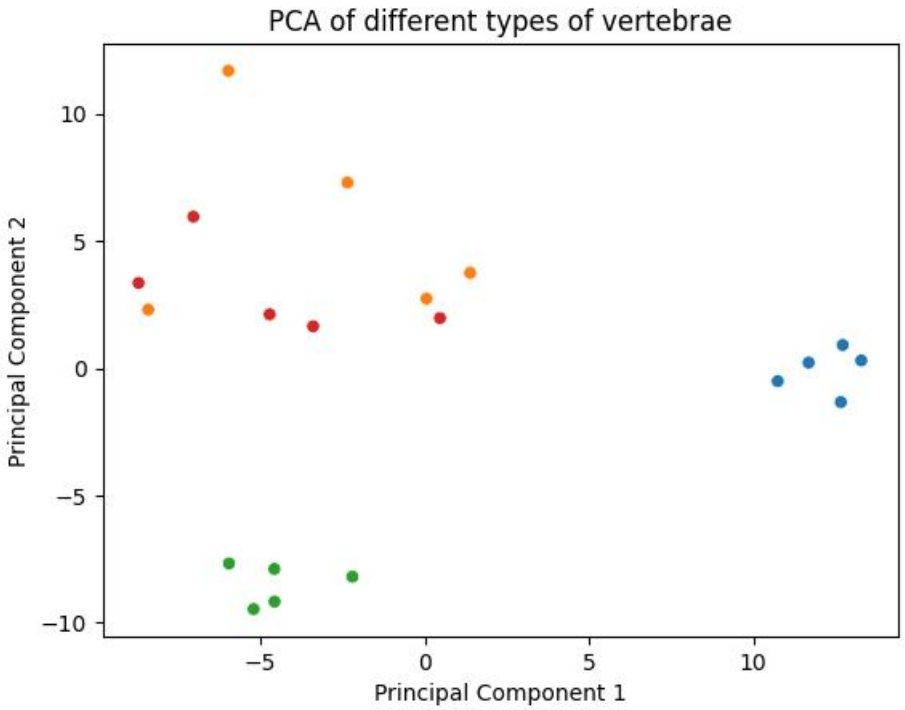


Fig. 7.16: PCA result visualization of different types of vertebrae, red - affected lumbar vertebrae, orange - affected thoracic vertebrae, blue - healthy thoracic vertebrae, green - healthy lumbar vertebrae.

involve vertebral processes, which can lead to different results than what we got before. The segmentation also filters out periosteum which we were not able to achieve manually. We performed the same analysis as we did before, with the manually extracted vertebral bodies. First of all, PCA was used to ensure that there are two clusters forming of healthy and affected vertebrae. Fig. 7.17 shows the outcomes of PCA.

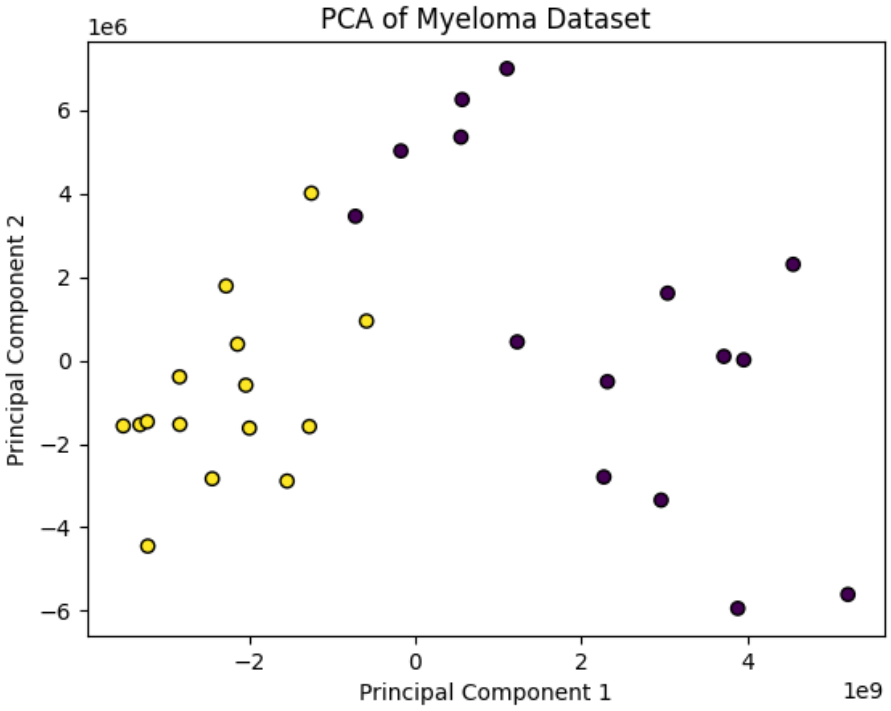


Fig. 7.17: PCA result visualization, segmentation done by nnU-Net, healthy - purple, MM patients - yellow.

7.3.2 Feature importance

Following the initial analysis, where we confirmed that there are differences in the values of features between vertebrae affected by MM and vertebrae without any pathologies, we looked deeper into feature reduction. After using Pyradiomcs, we were left with more than 100 features. The number had to be reduced to continue our analysis. This meant eliminating correlated or otherwise unnecessary features. There are many ways to reduce the number of features, with each one of them having its pros and cons. We explored a few statistical tests and correlation analysis. This proved to work, but not to an extent we were hoping for. We were still left with too many features which made working with them unnecessarily slow and confusing at times.

After doing more research we came across a random forest classifier. In a random forest model, feature importance refers to a technique used to determine the relevance or contribution of each feature (or variable) in making predictions. It helps in understanding which features have the most significant impact on the model's accuracy or output. This evaluation helps in pinpointing the most influential features for prediction, potentially aiding in feature selection, enhancing model efficiency, and offering insights into which factors carry more weight in the model's decisions [49]. We set the limit at 0.03 which means we were left with 15 most significant

features. As we can see in Fig. 7.18, two features carry the most significant amount of information.

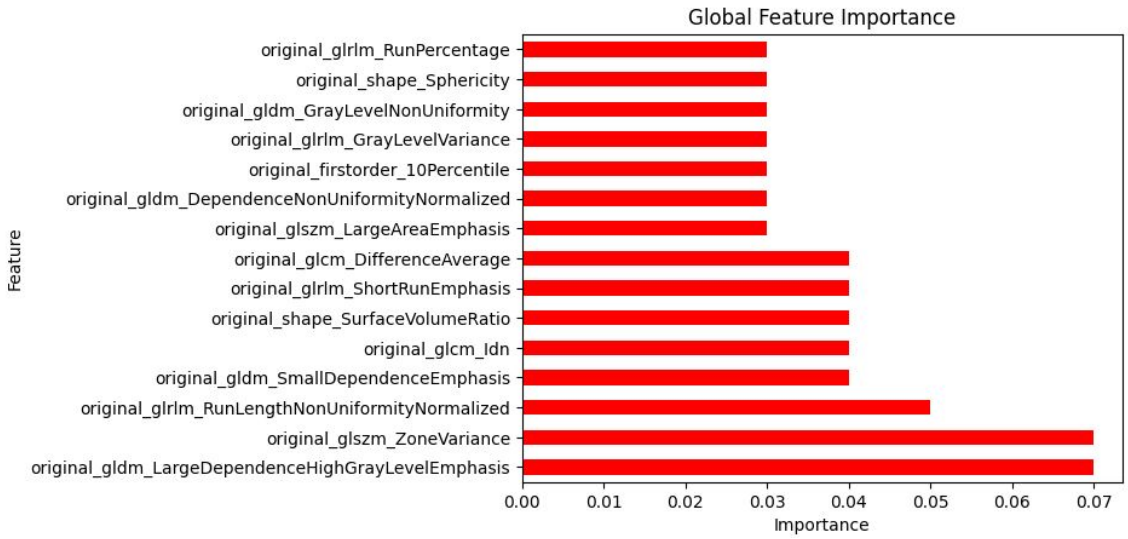


Fig. 7.18: Feature importance, manually extracted dataset, 15 features.

After performing the same analysis on lumbar vertebrae extracted using masks from nnU-Net, the results were visualized in Fig. 7.19. Just like in the dataset before, 15 most significant features were displayed.

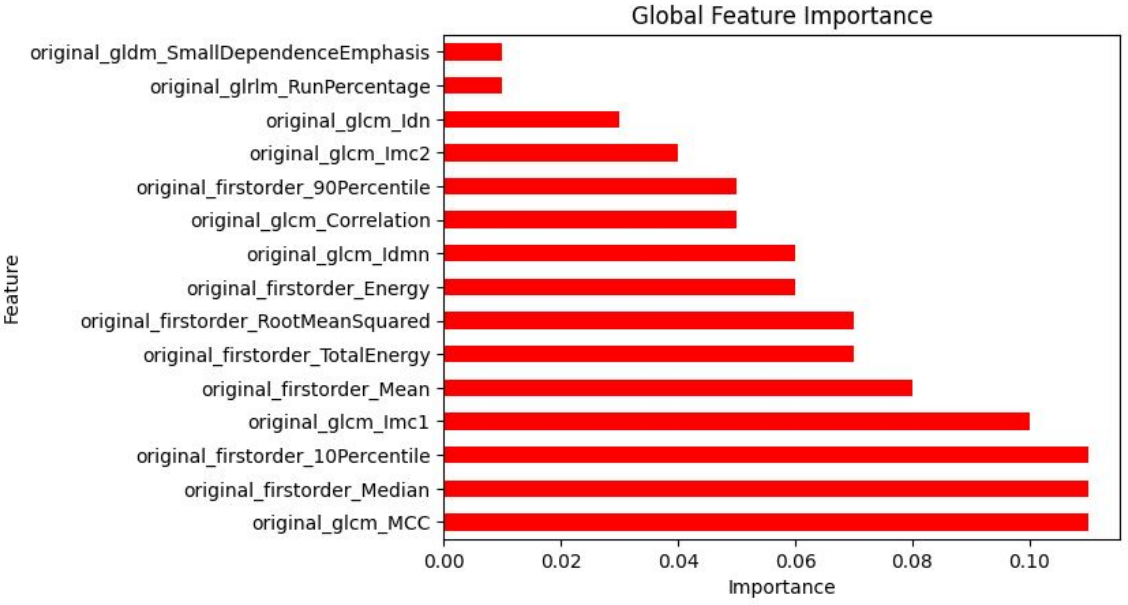


Fig. 7.19: Feature importance, lumbar vertebrae. 15 features.

After comparing Fig. 7.18 and Fig. 7.19, we see that there are a few differences. First of all, the features are different among the two datasets. Only 3 out of the 15

features are the same. Secondly, the importance of the features also varies. While in the dataset containing lumbar vertebrae the most significant features have an importance of approximately 0.11, in the manually extracted dataset it was only 0.07. This also means that there is a difference between the two datasets.

8 Analysis of lesions in follow-up scans

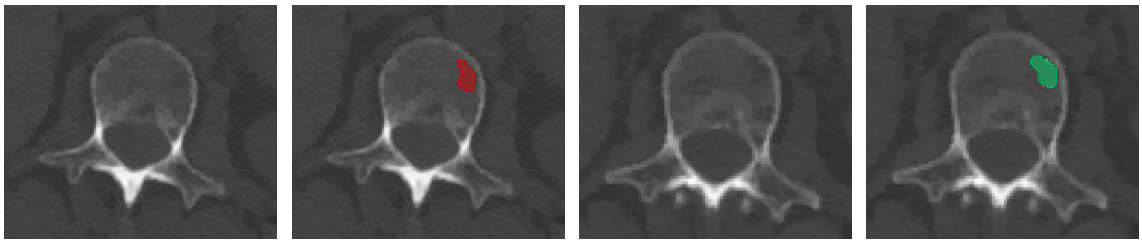
Besides the dataset mentioned, some of the patients were scanned multiple times during the course of their treatment. These scans are also available to us. This allows us to look at how the lesions change over time and how or if this correlates with the extracted radiomic features. We do not know whether the patient's condition is improving or declining. However, we can use the metadata from the DICOM files to see how much time has passed between the two scans. Only two patients were used for this analysis and three lesions were analyzed per patient. We tried to choose lesions that were correctly marked in the original scan as well as the follow-up scan, and that maintained their integrity, meaning that the lesion did not divide into multiple pieces. According to [50], MM patients should have check-up appointments every one to three months and should continue for the life of the patient.

In the first case, the time between the scans was approximately three months, which falls in the recommended range. Volume analysis was performed to better understand how the size of the lesions changed between the scans. We can see that despite the lesions being chosen from the same patient, not all of them grew. Tab. 8.1 summarizes the change in the lesion volume over time. In Fig. 8.1 we see the comparison of the same lesion in the first and second scans. Even though there is not a massive difference visible, in Tab. 8.1 (Lesion A), we see that the lesion has grown significantly.

	Scan 1 - volume [mm³]	**Scan 2 - volume [mm³]**	**Difference [%]**
Lesion A	302.980	502.497	65.852
Lesion B	74.757	87.082	16.487
Lesion C	136.117	125.818	-7.566

Tab. 8.1: Volume analysis, first patient.

However, the change was not significant. This can mean that the lesion is not correctly segmented in the second scan, or that the patient is responding to the treatment. We can only speculate what has caused this change.



(a) (b) (c) (d)

Fig. 8.1: Comparison of lesion A, first patient, virtual monoenergetic image at 40 keV, (a) - first scan, (b) - first scan with segmentation mask, (c) - second scan, (d) - second scan, segmentation mask.

	Scan 1	**Scan 2**
	Mean (standard deviation)	**Mean (standard deviation)**
Lesion A	-3.143 (56.697)	-6.555 (43.769)
Lesion B	34.186 (55.986)	31.229 (46.124)
Lesion C	-65.0 (22.011)	-41.118 (13.987)

Tab. 8.2: Difference in mean and standard deviation between two scans from the same patient, conventional CT scan.

Tab. 8.2 summarizes the change in mean and standard deviation of the brightness values of lesions from both scans. We can see that in all three cases, there was either a positive or a negative change. The results can mean a few things. A smaller standard deviation can mean that the area is more uniform and that the brightness values of the pixels are more alike that in the first scan, either brighter or darker. The change in mean values could be interpreted as change in the shade of the analyzed volume. Darker areas can mean that the disease is progressing, brighter areas can represent the formation of new bone tissue. However, because the periostenum appears brighter than the rest of the bone, change in mean values can be caused by incorrect segmentation of the lesion, where the mask includes a portion of periostenum. These inconsistencies make it more difficult to correctly interpret the results.

After analyzing the volume, the next step was to analyze how the features changed. Percentage increase or decrease was calculated for the two most changed features. The results were summarized in Tab. 8.3.

The same analysis was performed on another patient. Just like in the previous case, the time between the scans was approximately three months. Lesions were chosen based on the same criteria as before. In 8.4 we see the results of volume analysis of the lesions from the second patient. Tab. 8.4 shows the differences in volume and we can see that the change in volume was not significant.

	Maximum increase	Maximum decrease
Lesion A	original firstorder skewness	original glcm cluster prominence
Lesion B	original gldm large dependence low gray level emphasis	original glszm large area high gray level emphasis
Lesion C	original shape flatness	original glcm cluster shade

Tab. 8.3: Feature analysis, first patient.

	Scan 1 - volume [mm³]	**Scan 2 - volume [mm³]**	**Difference [%]**
Lesion A	554.370	751.526	35.564
Lesion B	851.834	999.199	17.300
Lesion C	172.460	159.646	-7.430

Tab. 8.4: Volume analysis, second patient.

	Scan 1	**Scan 2**
	Mean (standard deviation)	**Mean (standard deviation)**
Lesion A	27.062 (73.198)	17.044 (69.648)
Lesion B	23.185 (69.323)	23.333 (73.831)
Lesion C	64.267 (48.977)	61.682 (58.780)

Tab. 8.5: Difference in mean and standard deviation between two scans from the same patient, conventional CT scan.

	Maximum increase	Maximum decrease
Lesion A	original gldm large dependence low gray level emphasis	original shape maximum 2D diameter column
Lesion B	original glcm cluster shade	original glszm large area high gray level emphasis
Lesion C	original firstorder skewness	original glcm cluster shade

Tab. 8.6: Feature analysis, second patient.

For better understanding, we should look at what each of the features means [47]:

- original firstorder skewness - quantifies the degree of asymmetry in the distribution of values relative to the mean,
- original glcm cluster prominence - assesses the skewness and asymmetry within the GLCM. Elevated values suggest greater asymmetry around the mean, whereas lower values indicate a concentration around the mean with reduced fluctuation,
- original gldm large dependence low gray level emphasis - measures the ratio within the image of smaller-sized zones with lower gray-level values in their joint distribution,
- original glszm large area high gray level emphasis - assesses the percentage within the image of smaller-sized zones with elevated gray-level values in their

joint distribution,

- original shape flatness - illustrates the correlation between the largest and smallest principal components within the shape of the ROI,
- original glcm cluster shade - quantifies both the skewness and uniformity of the GLCM. Elevated cluster shade values indicate increased asymmetry around the mean,
- original shape maximum 2D diameter column - refers to the maximum pairwise Euclidean distance among vertices of the tumor surface mesh within the row-slice (typically coronal) plane.

We can see that some of the features are occurring more than once. However, the features are not the same among the two groups of lesions, those that grew and those that shrunk. This can be caused by a number of different factors. The most probable one is that the lesions were not marked correctly by their masks. This means that even thought the lesion may have grown, some pixels were left out of the masks which caused that it seems that the lesion shrunk. We can only speculate what would happen if the lesions were marked correctly. Based on our analysis, we can not make a reliable conclusion.

Conclusion

This thesis aimed to analyze osteolytic lesions in multiple myeloma patients. Before completing the analysis, it was crucial to study the disease itself to get a better understanding of the problem. We looked at the causes, prevalence and diagnostics criteria. Next, we explored the role of computed tomography in diagnosing MM. After completing the theoretical part of the thesis, patient data was collected during an internship at University Hospital Brno at the Clinic of radiology and nuclear medicine. During the internship, data was anonymized so it would be suitable to use for the purposes we intended.

The practical part was divided into a few individual subsections. The first part dealt with analyzing the whole vertebral bodies. The first step was a statistical analysis of the vertebrae. Using box-and-whisker plots to visualize the results showed a clear difference in statistical features between the two groups. A statistical test was used to prove that the differences are statistically significant. This knowledge, that there is a difference, was set as the foundation for the subsequent analysis.

The next comparison was based on the extracted radiomic features. Two approaches of extraction were taken. At first, the vertebrae were extracted manually using bounding boxes. This proved to have a few disadvantages. The biggest one was that we could not always extract the vertebrae without any portion of the processes. PCA was completed to ensure that there are two clusters forming to prove that there is a difference between the healthy and affected vertebrae.

After the initial analysis, a different approach to extraction was taken. To ensure that only the trabecular part of the vertebrae was segmented. nnU-Net was trained for this purpose. The network was trained and tested on a publicly available dataset containing only the lumbar part of the spine. The accuracy of the trained model was evaluated using the Dice coefficient and Hausdorff distance. For the Dice coefficient, the median was 0.980, the mean value was 0.949 and the standard deviation was 0.103. The Hausdorff distance has reached the median value 2.159, the mean value of 26.189 and the standard deviation of 48.527. Looking at the final values we can say that except for a few outliers, the model has performed well on the testing dataset.

The trained model was then applied to clinical data, which was prepared beforehand to only include the lumbar spine. We saw that the clinical data was segmented correctly, except for a few vertebrae containing inconsistencies, such as osteolytic lesions and metallic implants, which were not included in the training database. Same as with the manually extracted dataset, PCA was performed. PCA showed two clusters with smaller dispersion than before.

Both databases that were created were used to explore feature importance values.

Knowing which features were the most important when deciding whether the vertebrae are healthy or affected could be later used to train a classifier that would divide the vertebrae into two classes. After applying random forest, we were able to visualize the importance of each individual feature. However, for better understanding, only 15 features with the highest importance were visualized. By comparing the two datasets, we saw a clear difference in the importance of each feature. This supports our hypothesis, that the manual extraction could introduce an error further down the road, even more.

After completing the previous part of analyzing the vertebral bodies, we used the available follow-up scans to see how the volume of the lesions changes over time. The volume of the lesions was analyzed in two patients with follow-up scans to see whether the volume increased or decreased in between the scans. Three lesions were analyzed per patient. We saw that even if the lesions were extracted from the same scan, some of them grew and some of them shrunk. The rate at which the volume changed also varied. Besides analyzing the volume of the lesions, we looked at features that increased or decreased the most. These features were also different among the patients.

Bibliography

[1] Rebecca Silbermann and G. David Roodman. Myeloma bone disease. *Journal of Bone Oncology*, 2(2):59–69, 2013. URL: https://linkinghub.elsevier.com/retrieve/pii/S2212137413000201, doi:10.1016/j.jbo.2013.04.001.

[2] Tobore Onojighofia Tobore. On the need for the development of a cancer early detection, diagnostic, prognosis, and treatment response system. *Future Science OA*, 6(2), 2020-02-01. URL: https://www.tandfonline.com/doi/full/10.2144/fsoa-2019-0028, doi:10.2144/fsoa-2019-0028.

[3] Janita E. van Timmeren, Davide Cester, Stephanie Tanadini-Lang, Hatem Alkadhi, and Bettina Baessler. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights into Imaging*, 11(1), 2020. URL: https://insightsimaging.springeropen.com/articles/10.1186/s13244-020-00887-2, doi:10.1186/s13244-020-00887-2.

[4] Valerie C. Scalton. *Essentials of Anatomy and Physiology*. F.A. Davis Company, fifth edition edition, 2006.

[5] Vertebral column. URL: https://www.britannica.com/science/vertebral-column.

[6] *J. Gordon Betts, Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, Peter DeSaix*. OpenStax, Houston, Texas, 2013. URL: https://openstax.org/books/anatomy-and-physiology/pages/1-introduction.

[7] Scoliosis, 2023. URL: https://www.mayoclinic.org/diseases-conditions/scoliosis/symptoms-causes/syc-20350716.

[8] Peter J. Moley. Lumbar spinal stenosis, 2022. URL: https://www.msdmanuals.com/home/bone,-joint,-and-muscle-disorders/low-back-and-neck-pain/lumbar-spinal-stenosis.

[9] Spinal tumors. URL: https://www.aans.org/Patients/Neurosurgical-Conditions-and-Treatments/Spinal-Tumors.

[10] Filipa Macedo, Katia Ladeira, Filipa Pinho, Nadine Saraiva, Nuno Bonito, Luísa Pinto, and Francisco Gonçalves. Bone metastases. *Oncology Reviews*, 2017-03-03. URL: https://www.frontierspartnerships.org/articles/10.4081/oncol.2017.321, doi:10.4081/oncol.2017.321.

[11] G David Roodman and Rebecca Silbermann. Mechanisms of osteolytic and osteoblastic skeletal lesions. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4625391/, doi:10.1038/bonekey.2015.122.

[12] S Vincent Rajkumar et al. International myeloma working group updated criteria for the diagnosis of multiple myeloma. *The Lancet Oncology*, 15(12):e538–e548, 2014. URL: https://linkinghub.elsevier.com/retrieve/pii/S1470204514704425, doi:10.1016/S1470-2045(14)70442-5.

[13] Kimberly Nooan, Kathleen Colson, and Melissa Cochran. Patient care spotlight: Current management of hypercalcemia associated with multiple myeloma. URL: https://www.dana-farber.org/for-physicians/clinical-resources/hematologic-malignancies/advances-newsletter/2013-issue-2/patient-care-spotlight-hypercalcemia-multiple-myeloma.

[14] Sumana Kundu, Surajkumar B Jha, Ana P Rivera, Gabriela V Flores Monar, Hamza Islam, Sri Madhurima Puttagunta, Rabia Islam, and Ibrahim Sange. Multiple myeloma and renal failure. *Cureus*. URL: https://www.cureus.com/articles/86261-multiple-myeloma-and-renal-failure-mechanisms-diagnosis-and-management, doi:10.7759/cureus.22585.

[15] Multiple myeloma and anemia. URL: https://www.myeloma.org/multiple-myeloma-anemia.

[16] Bone disease. URL: https://www.myeloma.org/bone-disease.

[17] Richard Garnett. A comprehensive review of dual-energy and multi-spectral computed tomography. *Clinical Imaging*, 67:160–169, 2020. URL: https://linkinghub.elsevier.com/retrieve/pii/S0899707120302928, doi:10.1016/j.clinimag.2020.07.030.

[18] Reza Forghani, Bruno De Man, and Rajiv Gupta. Dual-energy computed tomography. *Neuroimaging Clinics of North America*, 27(3):371–384, 2017. URL: https://linkinghub.elsevier.com/retrieve/pii/S1052514917300187, doi:10.1016/j.nic.2017.03.002.

[19] Parang S Sanghavi and Bhavin G Jankharia. Applications of dual energy ct in clinical practice. *Indian Journal of Radiology and Imaging*, 29(03):289–298, 2021-07-22. URL: http://www.thieme-connect.de/DOI/DOI?10.4103/ijri.IJRI_241_19, doi:10.4103/ijri.IJRI_241_19.

[20] Tommaso D'Angelo, Giuseppe Cicero, Silvio Mazziotti, Giorgio Ascenti, Moritz H. Albrecht, Simon S. Martin, Ahmed E. Othman, Thomas J. Vogl, and Julian L. Wichmann. Dual energy computed tomography virtual monoenergetic imaging: technique and clinical applications. *Br J Radiol*. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6592074/, doi:10.1259/bjr.20180546.

[21] Davide Cester, Matthias Eberhard, Hatem Alkadhi, and André Euler. Virtual monoenergetic images from dual-energy ct. *Quantitative Imaging in Medicine and Surgery*, 12(1):726–741, 2022. URL: https://qims.amegroups.com/article/view/77005/html, doi:10.21037/qims-21-477.

[22] Tommaso D'Angelo, Moritz H. Albrecht, Danilo Caudo, Silvio Mazziotti, Thomas J. Vogl, Julian L. Wichamann, Simon Martin, Ibrahim Yel, Giorgio Ascenti, Vitali Koch, Giuseppe Cicero, Alfredo Blandino, and Christian Booz. Virtual non-calcium dual-energy ct: clinical applications. *Eur Radiol Exp*, 2021. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8413416/, doi:10.1186/s41747-021-00228-y.

[23] Sean Kitson. Transforming medical diagnostics: The emergence and impact of virtual non-contrast imaging. URL: https://openmedscience.com/transforming-medical-diagnostics-the-emergence-and-impact-of-virtual-non-contrast-imaging/.

[24] Marius E. Mayerhoefer, Andrzej Materka, Georg Langs, Ida Häggström, Piotr Szczypiński, Peter Gibbs, and Gary Cook. Introduction to radiomics. *Journal of Nuclear Medicine*, 61(4):488–495, 2020-04-01. URL: http://jnm.snmjournals.org/lookup/doi/10.2967/jnumed.118.222893, doi:10.2967/jnumed.118.222893.

[25] A.A. Litvin, D.A. Burkin, A.A. Kropinov, and F.N. Paramzin. Radiomics and digital image texture analysis in oncology (review). *Sovremennye tehnologii v medicine*, 13(2), 2021. URL: http://www.stm-journal.ru/en/numbers/2021/2/1712, doi:10.17691/stm2021.13.2.11.

[26] Giuseppe Corrias, Giulio Micheletti, Luigi Barberini, Jasjit S Suri, and Luca Saba. Texture analysis imaging "what a clinical radiologist needs to know". *European Journal of Radiology*, 146, 2022. URL: https://linkinghub.elsevier.com/retrieve/pii/S0720048X21005362, doi:10.1016/j.ejrad.2021.110055.

[27] Feature selection using statistical tests, 2022. URL: https://www.analyticsvidhya.com/blog/2021/06/feature-selection-using-statistical-tests/.

[28] Abdulrahman Alkandari and Soha Jaber Aljaber. Principle component analysis algorithm (pca) for image recognition. In *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)*, pages 76–80. IEEE, 2015. URL: http://ieeexplore.ieee.org/document/7224596/, doi:10.1109/ICCTIM.2015.7224596.

[29] Dong Yuan, Jian Huang, Xu Yang, and Jiarui Cui. Improved random forest classification approach based on hybrid clustering selection. In *2020 Chinese Automation Congress (CAC)*, pages 1559–1563. IEEE, 2020-11-6. URL: https://ieeexplore.ieee.org/document/9326711/, doi:10.1109/CAC51589.2020.9326711.

[30] Megha. P. Arakeri and G. Ram Mohana Reddy. Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images. *Signal, Image and Video Processing*, 9(2):409–425, 2015. URL: http://link.springer.com/10.1007/s11760-013-0456-z, doi:10.1007/s11760-013-0456-z.

[31] Kishor Kumar Reddy and Narasimha Prasad. A novel approach for detecting the bone cancer and its stage based on mean intensity and tumor size. 2015. URL: https://api.semanticscholar.org/CorpusID:53579331.

[32] Sinthia Balaji and K. Sujatha. A novel approach to detect bone cancer using k-means clustering algorithm and edge detection method. 11:8002–8007, 01 2016.

[33] A. Asuntha, A. Parveen Banu, K. Ainthaviarasi, S. Bharath Kumar, and A. Srinivasan. Feature extraction to detect bone cancer using image processing. *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, 2017.

[34] Madhuri Avula, Narasimha Prasad Lakkakula, and Murali Prasad Raja. Bone cancer detection from mri scan imagery using mean pixel intensity. In *2014 8th Asia Modelling Symposium*, pages 141–146. IEEE, 2014. URL: http://ieeexplore.ieee.org/document/7079289/, doi:10.1109/AMS.2014.36.

[35] Ranjitha M M, Taranath N L, Arpitha C N, and C.K. Subbaraya. Bone cancer detection using k-means segmentation and knn classification. In *2019 1st International Conference on Advances in Information Technology (ICAIT)*,

pages 76–80. IEEE, 2019. URL: https://ieeexplore.ieee.org/document/8987328/, doi:10.1109/ICAIT47043.2019.8987328.

[36] Ashish Sharma, Dhirendra P. Yadav, Hitendra Garg, Mukesh Kumar, Bhisham Sharma, Deepika Koundal, and Huiling Chen. Bone cancer detection using feature extraction based machine learning model. *Computational and Mathematical Methods in Medicine*, 2021:1–13, 2021-12-20. URL: https://www.hindawi.com/journals/cmmm/2021/7433186/, doi:10.1155/2021/7433186.

[37] Daniel Alsoof, Christopher L. McDonald, Wesley M. Durand, Bassel G. Diebo, Eren O. Kuris, and Alan H. Daniels. Radiomics in spine surgery. *International Journal of Spine Surgery*, 17(S1):S57–S64, 2023-06-28. URL: http://ijssurgery.com/lookup/doi/10.14444/8501, doi:10.14444/8501.

[38] Zhang Yingdi, Shi Zelin, Wang Haun, Cui Shaoqian, Zhang Lei, Liu Jiachen, Shan Xiuqi, Liu Yuneng, and Fang Lei. Lumvbcanseg: A lumbar vertebral body cancellous bone segmentation dataset. URL: https://doi.org/10.5281/zenodo.8181250.

[39] Medical imaging interaction toolkit documentation. URL: https://docs.mitk.org/nightly/index.html.

[40] Michal Nohel, Roman Jakubicek, Lenka Blazkova, Vlastimil Valek, Marek Dostal, Petr Ourednicek, and Jiri Chmelik. Comparison of spine segmentation algorithms on clinical data from spectral ct of patients with multiple myeloma. In *MEDICON'23 and CMBEBIH'23*, pages 309–317. Springer Nature Switzerland, Cham, 2024. URL: https://link.springer.com/10.1007/978-3-031-49062-0_34, doi:10.1007/978-3-031-49062-0_34.

[41] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. doi:10.1038/s41592-020-01008-z.

[42] German Cancer Research Center Division of Medical Image Computing. Nnunet. URL: https://github.com/MIC-DKFZ/nnUNet.

[43] Tomáš Vičar and Jiří Chmelík. Nnu-net v2 metacentrum example. URL: https://github.com/tomasvicar/nnunet_metacentrum_example.

[44] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945. URL: https:

//esajournals.onlinelibrary.wiley.com/doi/10.2307/1932409, doi:10.2307/1932409.

[45] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863. URL: http://ieeexplore.ieee.org/document/232073/, doi:10.1109/34.232073.

[46] Elliot Mcclenaghan. Mann-whitney u test. URL: http://www.technologynetworks.com/informatics/articles/mann-whitney-u-test-assumptions-and-example-363425.

[47] Pyradiomics documentation. URL: https://pyradiomics.readthedocs.io/en/latest/.

[48] A step-by-step explanation of principal component analysis (pca). URL: https://builtin.com/data-science/step-step-explanation-principal-component-analysis.

[49] Feature importance with random forests. URL: https://www.geeksforgeeks.org/feature-importance-with-random-forests/.

[50] Follow up. URL: https://about-cancer.cancerresearchuk.org/about-cancer/myeloma/treatment/follow-up.

Symbols and abbreviations

BMPC	bone marrrow plasma-cell percentage
CT	computed tomography
DICOM	Digital Imaging and Communications in Medicine
GLCM	gray level co-occurence matrix
HOG	histogram of oriented gradients
HU	Hounsfield unit
IBSI	Image Biomarker Standardisation Iniciative
IMWG	International Myeloma Working Group
MGUS	monoclonal gammopathy of undetermined significance
MITK	Medical Imaging Interaction Toolkit
MM	Multiple myeloma
MMBD	Multiple myeloma bone disease
MRI	magnetic resonance imaging
NIFTI	Neuroimaging Informatics Technology Iniciative
OCL	osteoclast
PCA	principal component analysis
RF	renal failure
ROI	region of interest
SVM	support vector machine
VMI	virtual monoenergetic image
VOI	volume of interest

A List of electronic attachments

The electronic appendix contains Python scripts used for data preprocessing and analysis. The learned nnU-Net model is due to its size available in the Zenodo repository `https://doi.org/10.5281/zenodo.11245116`.

- `/` ..... root directory
  - `bachelor_python_codes` ..... python codes used for preprocessing and analysis
  - `nnU-Net.txt` ..... commands used for nnU-Net planning and preprocessing
  - `package_versions.txt` ..... versions of used packages