**Automatic classification of body parts in X-ray images using neural networks**

**Automatic classification of body parts in X-Ray images using neural networks**

Bachelor thesis

Study programme: Informatics and Cybernetics in Healthcare

Field of study: Biomedical Informatics

Author of bachelor thesis: Patrik Pejchal

Bachelor thesis supervisor: Ing. Michal Reimer

**DECLARATION**

I declare that I have prepared my bachelor thesis entitled " Automatic classification of body parts in X-ray images using neural networks " independently and have used a complete list of citations of the sources used, which I list in the list attached to the bachelor thesis. I declare that this thesis is an exact copy of the electronic version submitted.

I have no serious reason against the use of this school work within the meaning of Section 60 of Act No.   
121/2000 Coll., on Copyright, on Rights Related to Copyright and on Amendments to Certain Acts (Copyright Act), as amended.

In Kladno 7.5.2024 ...........................................

Patrik Pejchal

**ACKNOWLEDGEMENT**

I would like to thank the supervisor of my bachelor thesis, Ing. Michal Reimer. For his valuable advice, for his help in the processing of this thesis. A big thank you to my family, who stood by me during the processing of this bachelor thesis as a support.

**ABSTRACT**

**Automatic classification of body parts in X-ray images using neural networks**

This bachelor thesis investigates improving the classification of X-ray images using convolutional neural networks (CNNs) such as VGG, ResNet, DenseNet and many others, focusing on addressing class imbalance and data scarcity through class weighting and data augmentation. The goal of these methods was to improve the diagnostic accuracy and robustness of models in medical imaging. The study demonstrated that class weighting effectively manages class imbalance, while augmentation techniques prevent overfitting, thereby increasing model generalizability and reliability. These findings suggest that targeted deep learning strategies can significantly improve the performance of CNNs in medical diagnosis, offering substantial benefits for clinical applications.

**Keywords**

Deep learning, medical imaging, X-ray á classification, CNN, data augmentation, class imbalance.

**ABSTRACT**

**Automatic classification of body parts in X-Ray images using neural networks**

This thesis explores the enhancement of X-ray image classification using Convolutional Neural Networks (CNNs) like VGG, ResNet, DenseNet and many more, focusing on addressing class imbalance and data sparsity through class weighting and data augmentation. These methods aimed to improve the diagnostic accuracy and robustness of models in medical imaging. The study demonstrated that class weighting effectively manages class imbalances, while augmentation techniques prevent overfitting, thus enhancing model generalizability and reliability. These findings suggest that targeted deep learning strategies can significantly improve CNN performance in medical diagnostics, offering substantial benefits for clinical applications.

**Keywords**

Deep Learning, Medical Imaging, X-ray Classification, CNN, Data Augmentation, Class Imbalance.

Table of Contents

[1. Introduction 5](#_Toc166494346)

[2. Overview of the current status 7](#_Toc166494347)

[2.1. History and development of medical imaging 7](#_Toc166494348)

[2.1.1. Early developments in medical imaging 7](#_Toc166494349)

[2.1.2. Advances in imaging technologies 9](#_Toc166494350)

[2.1.3. The digital revolution 10](#_Toc166494351)

[2.2. Challenges in X-ray image analysis 12](#_Toc166494352)

[2.2.1. Image quality and standardisation 12](#_Toc166494353)

[2.2.2. Interpretation and expertise 13](#_Toc166494354)

[2.2.3. Volume of data 14](#_Toc166494355)

[2.3. Deep learning in medical imaging 15](#_Toc166494356)

[2.3.1. Introduction to deep learning 15](#_Toc166494357)

[2.3.2. Successful deep learning projects 17](#_Toc166494358)

[2.3.3. Challenges and constraints 18](#_Toc166494359)

[2.4. Overview of existing deep learning models 20](#_Toc166494360)

[2.4.1. Convolutional neural networks 20](#_Toc166494361)

[2.4.2. Models in X-ray image classification 22](#_Toc166494362)

[2.4.3. Innovations and improvements 24](#_Toc166494363)

[2.4.4. Gaps and opportunities for research 25](#_Toc166494364)

[3. Objectives of the work 27](#_Toc166494365)

[4. Methodology 29](#_Toc166494366)

[4.1. Relevance of the dataset and theoretical relevance 29](#_Toc166494367)

[4.2. Preparation and pre-processing of data 31](#_Toc166494368)

[4.3. Data dissemination strategies 32](#_Toc166494369)

[4.3.1. Justification for augmentation 32](#_Toc166494370)

[4.3.2. Data dissemination strategies and their impact on model performance 33](#_Toc166494371)

[4.4. Model selection 34](#_Toc166494372)

[4.5. Evaluation framework 35](#_Toc166494373)

[4.5.1. Evaluation parameters 35](#_Toc166494374)

[4.5.2. Partitioning of the dataset 36](#_Toc166494375)

[5. Implementation 37](#_Toc166494376)

[5.1. Implementation of data preparation and dissemination 37](#_Toc166494377)

[5.1.1. Implementation of pre-processing 37](#_Toc166494378)

[5.1.2. Augmentation techniques used 38](#_Toc166494379)

[5.2. Processing and partitioning of datasets 40](#_Toc166494380)

[5.2.1. Practical overview and dataset setup 40](#_Toc166494381)

[5.2.2. The process of division 41](#_Toc166494382)

[5.3. Model training process 41](#_Toc166494383)

[5.3.1. Setting up the training environment 42](#_Toc166494384)

[5.3.2. Performing model training 42](#_Toc166494385)

[5.3.3. Coping with training challenges 44](#_Toc166494386)

[5.4. Preparing for performance evaluation 44](#_Toc166494387)

[5.4.1. Testing and evaluation of models 44](#_Toc166494388)

[5.4.2. Performance visualisation 45](#_Toc166494389)

[6. Results 47](#_Toc166494390)

[6.1. Overview of experimental results 47](#_Toc166494391)

[6.2. Experiment 1 - Effectiveness of class weights 47](#_Toc166494392)

[6.3. Experiment 2 - Impact of data augmentation techniques 49](#_Toc166494393)

[6.4. Experiment 3 - Comparative analysis of training on original vs. original + augmented data 51](#_Toc166494394)

[7. Discussion 54](#_Toc166494395)

[8. Conclusion 55](#_Toc166494396)

[List of literature used 56](#_Toc166494397)

[List of images used 60](#_Toc166494398)

# Introduction

Most diseases rely heavily on medical diagnosis, and the future of early disease detection and treatment development lies in rapidly developing areas of healthcare. Of all the methods used in imaging, X-ray imaging is one of the most preferred methods in most cases. This is due to its accessibility and performance, as it offers basic details of the internal structure of the human body. However, the interpretation of X-ray images still remains a very technically challenging complex task prone to human error. It is this fact that has recently led to the use of some computational technologies that use artificial intelligence (AI) and machine learning, in particular deep learning, to help improve the accuracy and efficiency of X-ray image analysis.

Deep learning, a subset of machine learning, has demonstrated extraordinary success in medical image analysis through the use of large datasets to train models that ultimately detect, classify, and localize pathological features with accuracy that is comparable to or sometimes even exceeds that of human experts [1]. In the field of visual information processing, convolutional neural networks (CNNs) have become one of the deep learning models and are selected for many tasks in medical diagnosis, such as classification, segmentation, and image enhancement. This demonstrates that hierarchical symptom representations learned from large-scale unstructured image data hold promise for the field of medical imaging, where the extraction of subtle cues can be important.

Despite these advances, there are several challenges in applying CNN to the classification of X-ray images. One of the many challenges is the imbalance of the classes. This is one of the standard cases for medical datasets, in which some conditions are extremely underrepresented compared to others. Such imbalance can cause model bias that could lead to good performance for frequent conditions but very poor performance for rare conditions. Another very important challenge is actually the fact that there is a scarcity of annotated medical images, and most routine deep learning models require large amounts of data if top results are to be achieved.

Such problems require very robust models that can generalize well to large-scale medical data, which are often scarce and heterogeneous. This work therefore addresses the challenge of evaluating the effectiveness of deep learning methods for X-ray image classification tasks, with a particular focus on how the problem of class imbalance, or a highly imbalanced dataset, could be reduced to make such a dataset more useful for training.

In this paper, CNNs are made more robust and accurate by using the following strategies: starting with the use of class weights to compensate for data imbalance, and ending with data augmentation techniques to artificially increase the trained dataset in order to inject more variability into the models. The research has three main objectives: first, to assess the impact of class weights on evaluating and managing class imbalance; second, to assess the impact of different data augmentation techniques on improving the performance of the models; and third, to compare the performance of CNNs between those trained on the original datasets and those trained on the original and augmented datasets. The goal of the given research is to find and investigate the best X-ray image classification techniques and configurations using state-of-the-art CNN architectures: the VGG, ResNet and DenseNet, among others.

# Overview of the current situation

## History and development of medical imaging

This chapter describes the transformative journey of diagnostic imaging technology from the accidental discovery of X-rays in 1895 to the digital revolution at the end of the twentieth century. It explores groundbreaking advances such as the development of computed tomography (CT), magnetic resonance imaging (MRI), and the advent of digital imaging systems. Each section highlights key innovations and their profound impact on medical diagnosis, illustrating how these technologies have revolutionized patient care and expanded the possibilities of medical science.

### Early developments in medical imaging

One of the most important discoveries was made in 1895 by Wilhelm Röntgen (Fig. 1) - the discovery of X-rays - which marked a revolutionary leap for science and subsequently for medicine. While working with cathode rays, Röntgen noticed that barium platinocyanide fluoresced on the screen, even though the instrument was shielded from light, and accidentally discovered X-rays. He noticed that with these rays it was possible not only to penetrate solid bodies, including human tissue, but also to image internal structures by methods that did not interfere with the body. He paid attention to the fact that X-rays make it possible to form an image of the internal structure of a person, and mentioned the possibility of their use in medical diagnostics. [2]

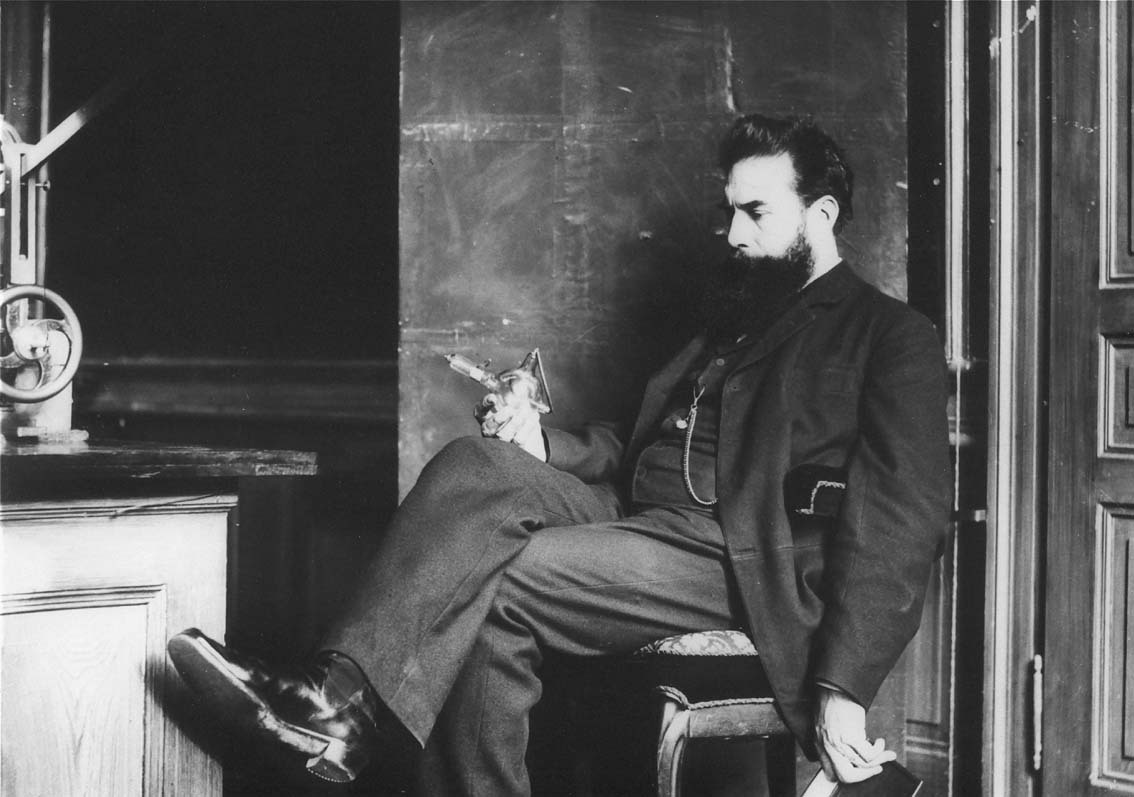


Image 1: Wilheim Conrad Röntgen [3]

Shortly after the discovery of a new type of X-rays, Röntgen produced the first X-ray photographs. One of them clearly showed a hand (Fig. 2); it belonged to his wife, showing the bones and ring on one of the fingers [4]. This iconic image was later to show not only the potential of X-rays for visualizing the internal structures of the human body, but was to arouse great interest and enthusiasm in the scientific community and indeed in the general public. The ability to "see" inside the body without surgery was seen as a breakthrough development in medical science [2].



Image 2: First X-ray image created [3]

Initially, it was used to diagnose bone fractures and locate foreign bodies, for example in gunshot wounds. Shortly after their discovery, a few weeks later, X-rays were already being used to aid medical diagnosis, demonstrating their immediate impact on medicine in aiding treatment and diagnosis. However, this enthusiasm quickly waned due to the dangers of misunderstanding and unidentified radiation exposure, which posed serious obstacles and limitations to the first applications of this technology. [4]

### Advances in imaging technologies

The switch from traditional X-ray machines to computed tomography (CT) in the early 1970s marked a quantum leap in diagnostic imaging. CT, perfected by Godfrey Hounsfield and Allan Cormack, offered a detailed cross-sectional view of the body that was vastly superior to the superimposed images that X-rays could produce [5]. This technology provided a more accurate image of the internal organs of the body and proved to be quite helpful in identifying the structure of soft tissues. In particular, it affected diagnostic accuracy and thus improved it significantly.

In addition to CT, magnetic resonance imaging (MRI) has emerged as another breakthrough technology in medical imaging (Fig. 3). It was founded by Raymond Damadian and greatly improved by Paul Lauterbur. It helped in finding high-resolution images of the internal structures of the body using magnetic fields and radio waves. Unlike X-rays and computed tomography, magnetic resonance imaging does not use ionizing radiation. This was a major breakthrough in patient safety. This made MRI particularly useful in neurology, oncology and in the diagnosis of musculoskeletal disorders, as it provided clearer contrast between different types of tissue, even though there were no differences in their density. [5]

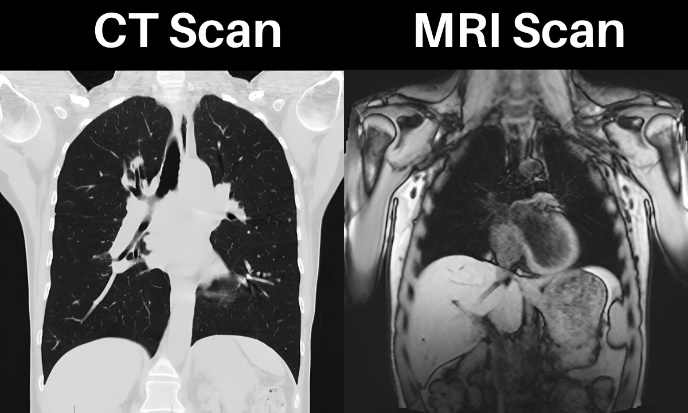


Image 3: Example of chest CT and MRI image [6]

Ultrasound technology (Fig. 4) has evolved from its simplest form to more complex digital systems that provide both clear and reliable images [7]. Recent technologies, such as the development of portable ultrasound devices, have provided mankind with even more extensive use of ultrasound imaging in various fields.

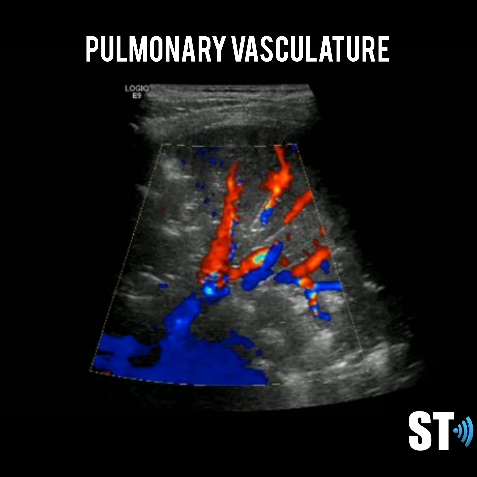


Image 4: Example of lung ultrasound [8]

Such facilities help develop immediate diagnostic capabilities even in remote or resource-limited settings and expand the scope of diagnostic practice, ultimately helping to improve patient care. These imaging modalities-CT, MRI, advanced ultrasonography-each outlined a spectrum of gigantic technological strides that needed to be made and have had a tremendous impact on diagnostic medicine. In most cases, each method complements the other, leaving powerful tools in the hands of the physician to make even more effective diagnoses and implement appropriate treatments for disease.

### The digital revolution

The last two decades of the 1990s and the 1990s itself witnessed some of the most significant technological changes in the field of medical imaging, as computing and digital storage brought about a new trend that led to the development of digital imaging. Initially, this change was met with a great deal of distrust from medical professionals due to the costs involved and also because the technology was new. However, the new advantages of increased image quality and the possibility of easy image manipulation were not to be overlooked, and so there was general acceptance. Key milestones within radiology were the introduction of computed radiography and PACS (Picture Archiving and Communication Systems), an integrated system of different imaging modalities into a single system aimed at streamlining workflow in healthcare facilities. [9]

The digital media revolution has greatly impacted the archiving and retrieval of medical images. Replacing the use of bulky physical materials with digital formats has, among other things, increased efficiency in image handling [10]. Digital databases have enabled easier manipulation and faster retrieval than traditional archives, revolutionizing the medical imaging workflow. The development has led to a rapid increase in access to patient records, thereby reducing the time schedule for diagnosis and treatment.

PACS systems have definitively changed the healthcare landscape by integrating different imaging techniques, improving workflows and even medical record interoperability. Integration has made great improvements in the way healthcare is delivered and patients are cared for, enabling rapid diagnosis and reducing delays associated with the physical transfer of images. The ability to remotely access images has further improved communication between healthcare professionals and has been critical to the globalisation of telemedicine. In this era of artificial intelligence and telemedicine, the importance of digital imaging has increased manifold. [11]

Telemedicine supports remote consultation and diagnosis, pushing access to medical expertise beyond geographical boundaries. The integration of artificial intelligence with digital imaging has further integrated revolutionary changes in image analysis to enhance diagnostic accuracy and predictive analytics. Applications of AI in medical imaging, such as automated detection and diagnosis, continue to advance and show a positive impact on patient outcomes by aiding in timely and accurate treatment decisions. [10]

Together, these findings point to the transformative impact of digital imaging on medicine, as they mark a shift from the field's original emphasis on image creation and acquisition to a more practical, much more focused focus on the processing, organization, and interpretation of medical images.

## Challenges in X-ray image analysis

This chapter explores the multifaceted challenges and innovations in achieving high-quality radiographs that are key to reliable diagnosis. It discusses the factors affecting image quality, including the technologies used, the expertise of technicians, and the environmental conditions in imaging facilities. In addition, the chapter examines the difficulties in standardizing x-ray protocols across different healthcare facilities due to varying equipment and the lack of universal imaging standards. These complexities highlight the importance of continued education, equipment upgrades, and rigorous standardization to enhance diagnostic accuracy and patient care.

### Image quality and standardisation

Achieving high quality X-ray images depends on a number of factors, including the technological capabilities and condition of the equipment, differences between manufacturers, and the expertise of the technicians who operate the systems [12]. It is essential that technicians are properly trained and experienced to optimize the use of X-ray equipment and to ensure proper image capture and interpretation. In addition, environmental conditions in imaging facilities, such as room setup and ongoing maintenance procedures, significantly affect the clarity and detail of X-ray images, highlighting the importance of maintaining a well-adjusted imaging environment [13].

Standardizing X-ray imaging protocols across different healthcare settings, such as small clinics to large hospital systems, poses significant challenges. This is primarily due to the lack of universal image quality standards, which leads to variability that can affect diagnostic procedures and potentially lead to misinterpretation of radiographs, thereby affecting patient care and diagnostic reliability.

Standardization efforts are further complicated by the variety of x-ray equipment used in different institutions. The wide variety of models and technologies, each with unique specifications and imaging capabilities, makes it difficult to establish a uniform standard. Upgrading older equipment to bring it up to modern standards often requires a significant financial investment and a concerted effort to ensure compatibility between different equipment and generations of technology, which is essential for consistent and accurate diagnostic imaging. [13]

Together, these factors point to the complexity of maintaining imaging standards in healthcare. Addressing these challenges through comprehensive training, effective management of the imaging environment, and rigorous standardization of protocols is essential to increase the quality and consistency of radiological diagnostics across healthcare settings.

### Interpretation and expertise

In the current healthcare situation, the role of radiology is indispensable. Without the specific expertise of radiologists in interpreting complex radiographs, the details that are key to diagnosis would remain obscure from the perspective of the average diagnostician. On the other hand, this involves complexities associated with a deep understanding of anatomy, disease and clinical context. Finally, with all these skills, radiologists are vulnerable to human errors such as fatigue, cognitive distortions, and the effects of a large amount of work that can make them commit errors in making a diagnosis. This fact highlights the need for continuous education and quite possibly the adoption of modern technology to help reduce the error rate. [14]

These challenges are compounded by a global shortage of trained radiologists, especially in regions with limited health resources. Waiting times for necessary medical interventions and delays in diagnosis then further contribute to limiting patient care. In most countries, including Mongolia [15] and Malaysia [16] a reduced ratio of radiologists to population further strains the health care system, with high pressure on existing radiologists and a high potential for increasing diagnostic errors. This situation therefore highlights the need for strategic training and deployment plans for radiologists that use more effective diagnostic modalities, such as teleradiology, aimed at mitigating the impact that such a shortage may cause.

The implications of these compounding problems are significant and extend to the quality of care for patients around the world. It would therefore result in poor health outcomes and hence poor access to the services of highly skilled radiologists and high error rates compounded by disparities in health care. This justifies an international partnership to address this deficit and to strengthen and even further develop the capacity of radiology services around the world to ensure that new radiology technologies and techniques reach a wide range of health care settings. Other solutions that need to be put in place include better training programmes, support for continuing professional development, teleradiology and artificial intelligence to make radiology services more accurate and efficient, which will ultimately help global patient care.

### Volume of data

The main reasons for the increase in datasets include technological developments in imaging systems, the growing trend in imaging procedures, and the wider application of screening programs.

In this context, attention should be paid to the fact that the use of modern technologies and tools in healthcare has increased the amount of image data produced to a level never seen before. Modern multi-slice CT scanners and high-resolution MRI machines produce datasets that are significantly larger than those produced by their older equivalents. [17]

This large volume of image data thus puts a lot of pressure on healthcare systems and in particular on the infrastructure for storing, retrieving and maintaining the data. This increased data load could accumulate inefficiencies in workflow, data access problems and a higher risk of data breaches. Such staggering volumes put a strain on healthcare facilities to handle, yet it is only logical that data management systems need to be very strong. Technologies such as cloud storage, data compression techniques, advanced database systems developed for large data volumes, and other scalable solutions are being designed and applied to address these needs as demand grows. [18]

In addition, the continued operation of the system requires sophisticated analytics and artificial intelligence (AI) to manage and analyse these large datasets. Artificial intelligence and machine learning help in automating image analysis, enabling fast and accurate diagnostics, leading to better decision making [17]. There have been significant improvements in some specific applications of AI related to working with and interpreting large image datasets - for example, pattern recognition and anomaly detection within images. These tools serve to automate data management while helping radiologists serve patients more accurately and quickly.

All of this underscores the urgent need for innovative solutions to manage the growing data demands in medical imaging so that health systems can continue to provide quality, efficient and safe patient care.

## Deep learning in medical imaging

This chapter explores the significant impact of deep learning on medical diagnosis, highlighting its ability to increase the accuracy and efficiency of disease detection and analysis across a variety of specialties. It discusses the challenges of data dependency and ethical considerations, highlighting the need for careful integration and regulatory compliance in order to fully exploit the potential of AI in healthcare.

### Introduction to Deep Learning

Deep learning is part of a subset of machine learning (Fig. 5), but with the unparalleled ability to comprehensively process vast amounts of data through many layers of abstraction, thereby mimicking the functions of the human brain. While traditional machine learning has often relied on manual feature extraction, deep learning has revolutionized by automation through layered neural network architectures. These networks consist of neurons in layers that communicate through interconnected nodes, simulating the way information is processed in biological brains. [19]

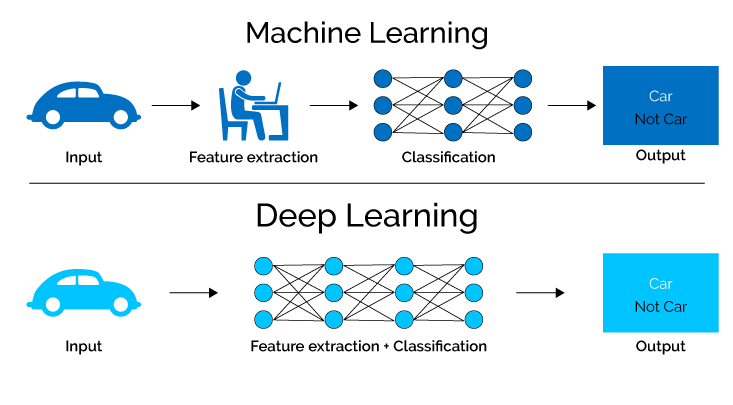


Image 5: Difference between deep and machine learning [20]

Deep learning capabilities far exceed those of conventional machine learning, especially when it comes to processing large volumes of data [21]. With the rise of data-intensive applications such as image recognition and natural language processing, deep learning outperforms most existing approaches in computational efficiency and accuracy. Deep learning enables finer and more detailed understanding and interpretation of data and can therefore outperform previous algorithms in speed and sophistication.

In addition, large datasets are a source of training that allows deep learning models to be built and improved autonomously. This is best demonstrated by convolutional neural networks (CNNs), which perform exceptionally well in tasks such as image classification and object detection. CNNs, like many other deep learning architectures, automatically learn to extract essential features from the data, a process that improves as more training data becomes available [22].

This self-learning capability makes deep learning particularly suitable for any application where new data is constantly being added, allowing the model to evolve and adapt to this new information over time. In short, deep learning is arguably the most innovative advancement in the field of artificial intelligence that can learn and even make decisions based on large data sets. What sets it apart are the intensive neural networks that are set up to automate the feature extraction and decision-making processes, making a turn from conventional learning methods to something much more dynamic and adaptive.

### Successful deep learning projects

When it comes to deep learning techniques, the use of convolutional neural networks (CNNs) has revolutionized medical diagnosis in various specialties, be it ophthalmology, oncology or cardiology or others.

Diagnosis of diabetic retinopathy: Detection of diabetic retinopathy, one of the leading causes of blindness in people with diabetes, has been greatly improved by deep learning models. Thanks to it, CNNs can detect subtle patterns in retinal images that indicate the disease, but which a human observer may not recognize. For example, a study by Alexander Rakhlin, 2018 [23] using deep learning achieved up to 99% sensitivity and 71% specificity, with an area under the curve (AUC) value of 0.97, which, among other things, surpassed the diagnostic accuracy achieved by some trained optometrists. This reflects the promise of AI for deep learning to increase efficiency and coverage in screening programs.

Cancer detection and classification: deep learning technology can play a role in efforts to improve quality of life by helping to detect and classify cancer. Deep learning has shown many potentials towards advancing the detection and classification of cancers, such as breast cancer, from a variety of images. More importantly, it is also valuable to highlight that these models specifically provided faster and more accurate measurement speed compared to traditional methods through complex training approaches. Indeed, deep learning will be able to achieve impressive levels of detection by leveraging large training datasets, whether fully clinically annotated or with only image-level labels. For example, in one of the studies by Li Shen and co-authors, 2019 [24] on the INbreast database on the well-known Full-Field Digital Mammography (FFDM) image repository, the best single deep learning model showed an impressive AUC of 0.95. For the four models, the AUC was 0.98 and the sensitivity and specificity were 86.7% and 96.1%, respectively, in the group of results with the combination. This highlights the potential role that deep learning could play in enabling earlier and more accurate detection of malignant tumors, which is critical not only for proper planning and treatment, but also for increasing survival rates through better diagnosis.

Cardiac imaging analysis: Deep learning has revolutionized cardiac care by enabling better assessment of the function and structure of the heart in a much more efficient and effective way than traditional methods. From a human perspective, automation of such analyses in cardiac imaging gives the assurance of accuracy and chances of early detection of any abnormality in the heart for precise treatment intervention. One of the great attributes of deep learning in this procedure - the process of segmenting images of the heart, which is an essential tool for physicians to diagnose the condition of the heart very accurately - is to streamline workflows for patients. [25]

The integration of deep learning has dramatically changed medical imaging for clinical practice. These technological breakthroughs are contributing to even greater automation of the terribly laborious and repetitive tasks associated with image analysis, taking away invaluable time and attention from professionals who must instead devote it to more complicated case analyses. All of these improvements improve the efficiency of the clinical workflow somewhat, but more importantly, they reduce the likelihood of diagnostic errors. In other words, deep learning models provide accurate insights that will increase diagnostic accuracy and aid in better patient management, thereby significantly improving patient care outcomes across medical specialties.

### Challenges and constraints

Deep learning has revolutionized the field by dramatically improving the ability to diagnose diseases and analyze medical images, not to mention the potential to predict a patient's prognosis. However, these models remain dependent on large, high-quality annotated datasets to train them, and therefore still pose a significant challenge. Collecting and annotating such data will require a great deal of effort, skill, and resources-something that is typically complicated by issues of privacy and the variety of data needed to avoid bias. Models trained on data without diversity may fail to produce adequate results across different demographic groups, which can lead to disparities in health care outcomes. [26]

This, in turn, makes the problems of interpretability and credibility of deep learning models much more challenging. In other words, healthcare professionals will only implement AI-generated decisions into clinical practice if they have a high level of understanding and trust in AI - perhaps even more than in the decisions themselves. Opaque, i.e. non-transparent, models of decision-making processes are still a significant barrier to greater acceptance of deep learning models. This is a classic requirement in the development of explanatory deep learning models (EDLMs), as they aim to make the AI decision-making process much more transparent, understandable and clinically relevant [27]. These, in turn, aim to provide clinicians with not only accurate but also interpretable knowledge, which is quite crucial for gaining trust and enabling clinical adoption.

Ethical considerations - patient privacy and algorithmic bias - must be at the forefront. Ensuring the ethical use of AI in healthcare would mean adhering to the strictest data protection laws, conducting regular bias checks and working with ethicists to guide them through the moral maze that AI brings. These measures are essential to preserve the integrity of patient information and create a fair playing field for healthcare. [26]

In addition, regulatory and integration issues need to be addressed to ensure successful implementation of the technologies in clinical settings. Important challenges include compliance with medical standards and regulatory approvals, which can act as barriers to smooth integration into everyday clinical use. This is compounded by ensuring technical integration so that AI systems connect and coexist well with existing healthcare IT systems to ensure that adoption does not impact workflow. This will require collaboration between AI developers, healthcare professionals and regulators to ensure that the deep learning tools developed add value to healthcare rather than complicate matters.

All of these factors together suggest that the field of deep learning deployment in healthcare is vast and somewhat complex, with many potential promises, but also many potential pitfalls that need to be carefully circumvented in order to realize the full potential of AI in healthcare.

## Overview of existing deep learning models

This chapter discusses the key role of convolutional neural networks (CNNs) in deep learning, highlighting in particular their effectiveness in processing and analysing visual data. It explores how the CNN architecture - consisting of convolutional, clustered, and fully connected layers - enables complex image analysis tasks, revolutionizing fields such as face recognition, autonomous vehicle navigation, and medical diagnostics. The chapter also discusses the broader implications and challenges of using these models in real-world applications, highlighting their transformative impact on how machines perceive and interact with the visual world.

### Convolutional neural networks

Convolutional neural networks (CNNs) are the most successful and widely used techniques in deep learning precisely because of their analytical processing of visual data. The architecture includes convolutional layers, clustering layers, and fully connected layers (Fig. 6). Each of these layers plays a critical role in the network's ability to perform complex image analysis tasks. The convolutional layers apply several filters to the input images and create feature maps that would convey important information, such as edges or textures. Association layers reduce the spatial dimensions of these symptom maps. Thus, the number of parameters and computations required is greatly reduced, which increases the computational efficiency of the network and helps it not to be overfitted. Fully connected layers combine these elements to produce a final output for either classification or other forms of prediction. [28]



Image 6: CNN model architecture [29]

The power of CNNs in processing visual data is based on their design, which mimics some of the behavior of the center of the human eye. CNN neurons respond to overlapping areas in the visual field, defining local receptive fields that help capture spatial and temporal dependencies. This design therefore allows the network to maintain sensitivity to given features regardless of their location in the input image, leading to translational invariance, which is very important in tasks where the orientation and position of objects have considerable variance. [30]

CNNs also use common weights and biases (Fig. 7). This reduces model complexity and the number of parameters, increasing its computational accessibility, as well as further reducing the risk of overfitting [28]. CNNs have dramatically changed the potential of systems in practical application areas such as face recognition, autonomous vehicle navigation, and medical image diagnostics. In autonomous driving systems, CNNs are used to improve interpretation over road signs, obstacles, etc. and thus enhance the current safety and reliability of the self-driving system [31]. CNNs in medical imaging contribute to early cancer detection and categorization of different cancers based on image scans, thereby improving accurate diagnosis and ultimately leading to effective therapy planning [32].

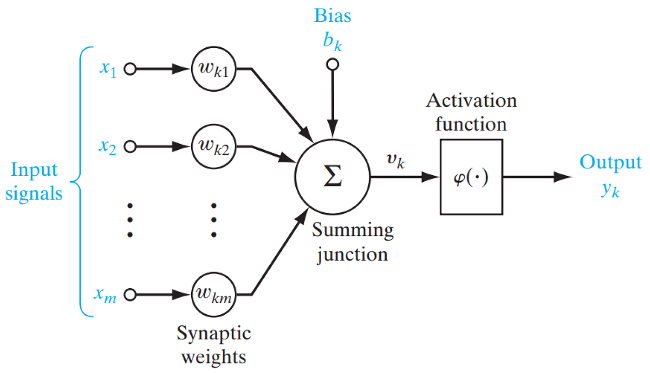


Image 7: Weights and biases in neurons [33]

These applications are a challenge that shows that a deep learning model can recognize and interpret some very complex patterns in visual data sometimes even more accurately than the level of human ability. What has opened the way to breakthroughs in performance and efficiency largely through the general development and application of CNNs in image recognition and related fields; the general development and application of CNNs in image recognition have fundamentally changed the way machines perceive and interact with the visual world. These technological advances support a wide range of applications that lead to significant impact not only in industry but also in everyday life; this means the ability of artificial intelligence to greatly enhance human-machine interaction.

### Models in X-ray image classification

Deep learning models including convolutional neural networks (CNNs), residual networks (ResNets) and U-Nets have become indispensable tools in X-ray image classification, revolutionizing medical diagnosis.

CNNs are the foundation for image classification tasks, using layers of convolutional filters to automatically detect basic features such as edges, shapes, and textures that are key to understanding complex medical images such as X-ray images. This architecture, enriched with activation features such as ReLU and layer pooling, efficiently processes pixel data while reducing dimensionality and computational burden, thereby increasing the accuracy of feature extraction. [34]

ResNet (Fig. 8) solves the problem of training deep neural networks by introducing skip connections that alleviate the vanishing gradient problem, allowing much deeper networks to be successfully trained. By enabling feature reuse and preserving the flow of information through successive layers, ResNet excels at maintaining accuracy and capturing complex patterns in medical images. Their adaptability to complex diagnostic tasks makes them invaluable in X-ray image classification, where accuracy and depth of analysis are paramount. [35]

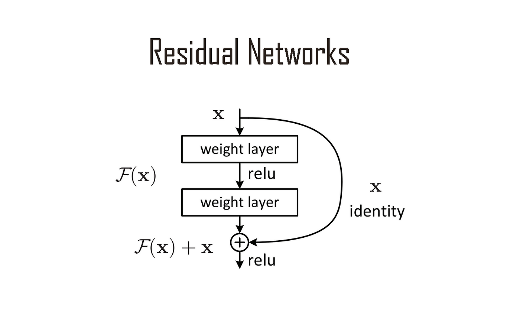


Image 8: ResNet model [36]

U-Nets (Fig. 9), originally designed for biomedical image segmentation, offer exceptional capabilities in segmenting images into individual parts, facilitating accurate localization of abnormalities in X-ray images. The unique U-shaped architecture, consisting of a tapering path for context capture and an expanding path for precise localization, enables U-Nets to achieve high accuracy even with limited annotated data, making them indispensable in medical imaging tasks. [37]

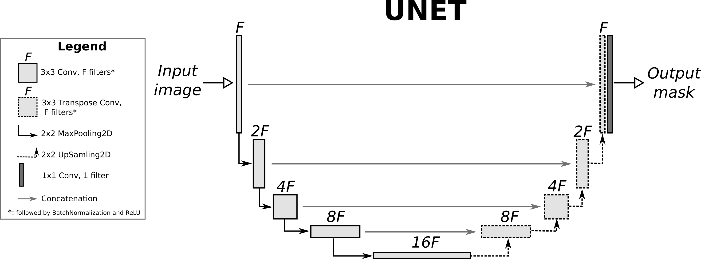


Image 9: U-Net model [38]

These models are particularly well suited for X-ray image classification because they efficiently process large amounts of data typical of medical images and learn from the complex patterns present in medical diagnostics. They reduce the dependency on human input and improve the ability to automate the detection and classification of various medical conditions, which greatly aids in diagnosis and treatment planning.

### Innovations and improvements

These recent advances in deep learning have increased the accuracy and speed of diagnosing diseases from medical images using sophisticated models and hardware improvements to the way diagnosis is done.

Increasing accuracy in disease detection: this is the main reason why ECA-DenseNet (Efficient Channel Attention-DenseNet) has been proposed in recent neural network architectures to refine the distinction between benign and malignant tumors. For example, in the work by AYDIN, Halise Nur, and Oktay YILDIZ, 2023 [39] the classification accuracy of the ECA-DenseNet model was shown to be very promising, reaching 95.07%, effectively demonstrating the possibility of efficiently extracting dense connections from significant features by omitting out-of-place information. This higher level of detail is therefore quite crucial for accurate diagnosis and more precise treatment planning, thereby reducing the risk of misdiagnosis with all its likely consequences.

Increased processing speed through hardware optimization:The integration of AI chips into the technology stack is designed to increase the computational efficiency of medical imaging by optimally tuning hardware-accelerated deep learning operations. These chips now enable parallel processing and efficient computation, greatly reducing the time required to perform complex tasks such as image analysis. For example, in the case of an AI chip with dedicated convolutional accelerators, image processing can be several times faster. [40]

Detecting subtle diagnostic features: these new deep learning models can capture the subtle diagnostic features needed to detect diseases such as COVID-19 in their early stages. In addition, multiscale and multimodal learning approaches increase the models' sensitivity to finer details that are critically important for detecting early-stage diseases and minor anomalies. Advanced DenseNet-type models with convolutional attention modules have been particularly effective in tackling the problem of intensive level segmentation of brain tumors, which would be useful for accurate analysis of complex medical images [41]. This capability is essential for early and timely diagnosis and consequently can improve patient outcomes through early and targeted treatment.

### Gaps and opportunities for research

Medical imaging is an area where data preparation using augmented methods has a large impact on the effectiveness of models during training. However, working with noisy data remains the most widespread research gap to date. In some medical images, noise can dominate the data, obscuring the distinctive patterns needed for diagnosis. Unbalanced datasets can lead to biased models and consequently poor performance in underrepresented classes. Further, comprehensive studies have been carried out to analyze the same with the integration of various advanced preprocessing and augmentation techniques to systematically analyze their impact in different model architectures. [42]

There are huge opportunities for further development in this research area. In other words, the quality of training datasets and the efficiency of their use could be significantly improved by innovative improvements in noise reduction algorithms and the development of new ways of generating synthetic data [42]. Advanced noise reduction cleaning could improve the input data to better highlight the essential features and improve the model training results. Some interesting data augmentation methods, such as complex geometric transforms and photometric augmentations, combined with the use of generative adversarial networks (GANs) could bring the needed variability to the datasets. In particular, GANs (Fig. 10) offer promising prospects for generating synthetic medical images that can simulate a wide range of pathological conditions and thus address the problem of class imbalance. [43]

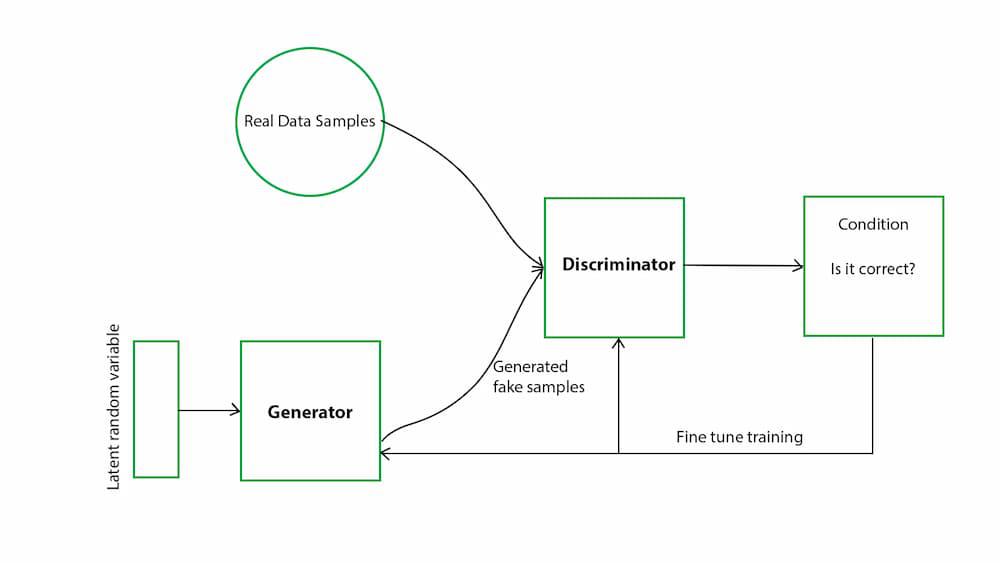


Image 10:GAN architecture [44]

Such extension techniques will not only help in enriching the dataset, but also provide a more robust basis for training deep learning models, thereby improving diagnostic accuracy and reliability. Together, these all contribute to improving the preprocessing and data augmentation necessary to improve the capabilities of medical imaging technologies. All of these research gaps need to be appropriately filled in order for the field of medical imaging to more successfully leverage deep learning to benefit patients through more accurate and timely diagnosis.

# Goals of the work

In general, the main goal of this work is to exploit the potential offered by neural networks, intensive learning models, in the classification of X-ray images that reflect anatomical structures. Therefore, this study attempts to address some important issues related to the classification of X-ray images, such as the imbalance of classes and the intrinsic limitations existing in X-ray image datasets. The goal of this research is to develop a robust diagnostic tool, specifically in the healthcare domain, that can significantly improve accuracy and reliability in these environments. The following sections outline the basic approaches and methodologies:

Advanced techniques to improve model performance:

* *Class imbalance reconciliation:* research techniques, such as augmenting data and class weights during training, to address class imbalances, reduce bias in dominant classes, and increase performance in underrepresented classes.
* *Improve data quality:* improve data consistency and quality through standardized preprocessing steps such as image resizing and normalization, thereby improving the ability to interpret and learn models from images.

Testing and comparing different CNN architectures:

* *Evaluation of performance metrics:* evaluate the overall performance metrics (Precision, Recall, F1-score) of different CNN models and gain insights to identify the best performing models for X-ray image classification.
* *Transfer learning potential:* Explore the possibility of using pre-trained models (e.g., VGG, ResNet, etc.) as symptom extractors and assess how they can be fine-tuned for medical imaging data.

Adapting CNN models to multi-label classification:

* *Architecture Modifications:* modifying CNN architectures to efficiently handle multiple labels using appropriate loss functions such as binary cross entropy, and activation functions such as sigmoid.
* *Fine-tuning strategies:* fine-tuning models using hyperparameter techniques to maximize performance, adjusting parameters such as learning rate and batch size based on the characteristics of the dataset.

Implementation and evaluation of data augmentation techniques:

* *Impact of enhancement categories:* testing the impact of different enhancement techniques (geometric transformations, color adjustments, noise addition) on the ability of models to generalize and perform robustly under different imaging conditions.
* *Generalization capability and sampling efficiency:* Focus on how augmentation techniques can help reduce overfitting and improve performance, especially when learning from datasets with a low number of labels.

The results of this work are expected to make a valuable contribution to the field of medical diagnostic tools as an improved tool for the analysis of X-ray images and eventually lead to the development of more reliable and accurate diagnostic systems in medical facilities.

# Methodology

## Relevance of the dataset and theoretical relevance

The quality of the dataset significantly affects the effectiveness of deep learning models in medical imaging. For the purpose of this research, a dataset containing annotated X-ray images of various body parts was obtained from Kaggle [45]. This dataset includes key aspects such as the resolution of the images, the variety of body parts represented, and the clarity with which these anatomical structures are depicted, each of which plays a key role in the training and performance of the models.

High-resolution images are essential for medical diagnosis because they allow the model to discern the precise features needed to detect subtle pathological signs. However, the high dimensionality of high-resolution datasets can increase the computational demands and risk of overfitting. The dataset in this study contains images (Fig. 11) from 22 different classes representing a wide variety of medical conditions. This variability not only challenges the model but also increases its robustness. The presence of 1,738 annotated images provides a substantial amount of data that is crucial for training sophisticated deep learning models capable of complex pattern recognition and decision making tasks. In general, a larger dataset translates into better model performance, especially in critical domains such as medical diagnosis.

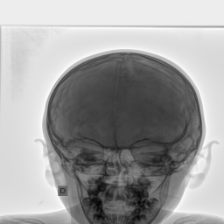
Image 11: Example of X-ray image classes from the dataset used: 1- Skull (Class 18), 2- Hand (Class 9), 3- Foot (Class 6), 4- Knee (Class 11) Source: author

**4**

**3**

**2**

**1**



The properties of the dataset directly affect the robustness of the model, which is defined as the ability of the model to maintain high performance on new, unobserved data. The diversity and volume of the dataset are helpful in mitigating common problems such as overfitting. In addition, the variety of conditions and high-quality images within the dataset allows deep learning models to efficiently learn the subtle features of medical images. This learning is essential to develop models that are accurate, reliable, and effective for diagnostic prediction.

The richness of the dataset provides ample opportunity for advanced model training, but also presents challenges, such as the significant preprocessing required to standardize the images for consistent model input. In addition, the diverse nature of the dataset requires sophisticated data augmentation strategies to simulate real-world variability, thereby increasing the ability of the model to perform under different conditions.

## Preparation and pre-processing of data

Efficient preprocessing is crucial for preparing medical image data for deep learning models.

Image resizing is necessary to ensure consistent image sizes across the dataset, allowing the neural network to process the data efficiently. Typically, images are scaled to a standard size, in this case, for example, 224 × 224 or 299 × 299 pixels. This standardization simplifies the computational processes by reducing the variability of the input data structure and ensuring a uniform input data size for the model. Such uniformity is very important as it helps the neural network in learning and efficient feature extraction without the additional complexity resulting from different image sizes.

Normalizing the pixel values to a common scale, either from 0 to 1 or from -1 to 1, is the next key step. This scaling stabilizes the performance of deep learning models by helping them converge faster during training and reducing sensitivity to the range of input values. Normalization facilitates a more consistent learning rate for different features, thus increasing the overall efficiency and effectiveness of model training.

Conversion to grayscale is particularly important in many types of medical imaging where color does not add diagnostic value. Converting images to grayscale reduces the computational requirements by reducing the dimensionality of the data. This allows the model to focus more on structural differences rather than color differences, which is particularly important in tasks such as X-ray and MRI image analysis where structural details are more significant than color.

Together, these preprocessing strategies address both the need for standardization and the challenges posed by the inherent variability and complexity of medical image data. They are designed to thoroughly prepare the data and ensure that deep learning models are trained on data that is as consistent and informative as possible.

## Data dissemination strategies

This chapter discusses the importance of data augmentation in machine learning, particularly in the field of medical imaging. It explains how various data augmentation techniques, such as geometric transformations, color adjustments, and noise introduction, substantially improve the robustness and generalization of the model. By artificially augmenting the training dataset, these methods help prevent overfitting and ensure that models can effectively handle a wide range of real-world scenarios.

### Justification for augmentation

Data augmentation is essential in machine learning, especially in medical imaging, where data variability and dataset size significantly affect model performance. Artificially augmenting the dataset through various transformations increases the robustness and generalizability of the model. This process not only reduces the risk of overfitting, but also ensures that the model can efficiently handle a wide range of real-world scenarios. The types of extensions used can vary considerably, each serving a specific purpose in improving model learning:

*Geometric transformations*: these include rotations, rotations, scaling and translations. Such transformations are particularly valuable in medical imaging, where the position of objects can vary significantly. By allowing the model to recognize features in images regardless of their orientation and scale, geometric transformations help ensure that diagnostic accuracy is not compromised by the physical presentation of the image.

*Colour adjustments*: although not always applied in medical imaging, colour adjustments play a key role in specific areas such as dermatological images or colour scans. Adjustments such as brightness, contrast, and histogram equalization can be essential to enhance image clarity, making it easier for the model to detect relevant features in images.

*Introduction of noise*: The addition of synthetic noise, such as Gaussian noise or salt-and-pepper noise, simulates common artifacts that can occur in medical images. This type of enhancement trains the model to focus on relevant features despite the presence of noise, thereby increasing its diagnostic accuracy in less than ideal conditions.

Data augmentation proves particularly beneficial when the dataset as a whole or specific classes within it have limited samples. By increasing the effective size of the trained dataset through these augmentation techniques, the model learns from a wider and more diverse set of presented data. This reduces the tendency of the model to memorize the trained data and increases its ability to generalize to new, unobserved data.

### Data dissemination strategies and their impact on model performance

Data augmentation plays a key role in improving pattern learning and recognition capabilities in medical imaging. By training with augmented data, models gain better generalization capabilities, making them less likely to overfit to the nuances of the training set, and more likely to perform accurately on new, unfamiliar datasets. This is particularly important in medical applications where consistent performance across different clinical environments is essential. In addition, the extension introduces variability that trains the model to ignore irrelevant variations in the input data, such as noise or non-standardized image orientation, thereby increasing the robustness of the model. Such robustness is crucial for clinical deployments where images may not always be presented under ideal conditions.

Specific examples and benefits:

*Rotation and rotation*: these extensions help the model learn to recognize anatomical structures from different angles, increasing the accuracy and flexibility of the model in clinical diagnosis.

*Noise*: Introducing noise during training prepares the model to process low-quality clinical images more efficiently, ensuring reliability even under suboptimal conditions.

When selecting specific enhancement techniques, the choice is often justified on the basis of their demonstrated efficacy in previous studies and their relevance to the types of medical imaging data used in the research. For example:

*Geometric and noise extensions*: these are selected for their ability to simulate real differences in medical imaging, such as patient movement or differences in imaging technology. These types of extensions help ensure that the model can handle a wide range of imaging conditions.

*Colour adjustments*: these are primarily selected for datasets involving visible light images where colour information can provide significant diagnostic value. Adjustments such as brightness, contrast, and histogram equalization are particularly valuable in improving the detectability of subtle features in images.

Together, these augmentation strategies not only prepare the models to cope with different imaging conditions, but also contribute significantly to the overall efficiency and reliability of medical diagnosis in real-world applications.

## Choosing a model

The selection of appropriate deep learning models for medical imaging is a key decision that directly affects the efficiency, effectiveness, and practical usability of a diagnostic tool. The selected models must excel in architectural capabilities, computational efficiency, and have proven results in similar applications.

Models are selected primarily based on their ability to efficiently process complex image data, with an emphasis on features such as architectural depth and specialized layers, such as convolutional layers, which are key to extracting the detailed features needed for accurate diagnostics. Models must also exhibit high computational efficiency to ensure fast processing in clinical environments where rapid decision making is critical.

Key models used:

*- 224x224 Optimal size*:

* VGG-19: It features a deep architecture that is effective in detailed feature extraction.
* ResNet variants (ResNet-50, ResNet-101, ResNet-152): include residual learning to support the deeper network training that is necessary for comprehensive image analysis.
* DenseNet variants (DenseNet-121, DenseNet-169, DenseNet-201): use densely connected layers for efficient parameter utilization and robust function transfer to optimize computational resources.
* EfficientNet variants (EfficientNet-B0, EfficientNet-B1, EfficientNet-B2): systematically scaled to efficiently process different image resolutions.

*- Optimal size 299 × 299*:

* InceptionResNetV2: Merges the Inception architecture with residual connections, balancing advanced learning capabilities with computational efficiency, ideal for higher resolution images.
* InceptionV3: Known for its operational efficiency and accuracy at higher resolutions, it uses sophisticated convolution and dimension reduction strategies to optimize computational demands.

These models were carefully selected to increase the accuracy and efficiency of medical image analysis, relying on their proven ability to process complex image data and innovative architectural features that ensure adaptability and robustness in a variety of medical imaging scenarios.

## Evaluation framework

This chapter outlines the importance of selecting appropriate evaluation metrics, such as accuracy, precision, feedback, and F1 scores, which are critical for measuring performance and ensuring the clinical reliability of medical imaging models. It also describes a structured testing strategy that splits the data into training, validation, and test sets, allowing for thorough model tuning and validation to assess how well the model performs in real-world diagnostic scenarios.

### Evaluation parameters

The selection of appropriate evaluation metrics is critical to accurately estimate the performance of developed methods in medical imaging, ensuring that they meet clinical standards for accuracy and reproducibility. Accurate metrics guide model improvement throughout the learning process and are essential for clinical validation.

Accuracy measures the total number of correct predictions made by the model across all classes, which is essential for monitoring and improving the model's learning curve as it evolves. Precision, or positive predictive value, indicates the proportion of correct positive identifications made by the model, which is critical in medical settings where the cost of false positives is high. Recall, also known as sensitivity, evaluates the ability of the model to capture all relevant instances of a condition, ensuring that no critical information is missed in the diagnosis. The F1 score, a harmonic mean of precision and recall, offers a balanced measure of the model's precision and robustness, which is particularly useful in settings with unbalanced data sets.

Together, these metrics provide a comprehensive framework for assessing model performance, detailing results for each class during testing and offering insight into the model's ability to effectively manage a variety of health conditions.

### Partitioning of the dataset

To ensure a robust assessment process that reflects real-world medical diagnostic scenarios, the datasets are strategically divided into three key sections.

The training set is used for initial tuning of the model parameters - adjusting weights and biases over several iterations to minimize the loss function. This set forms the largest subset and provides the model with a substantial base for learning the various patterns and functions necessary for accurate diagnosis.

The validation set plays a key role in tuning the hyperparameters of the model and implementing early stopping mechanisms to avoid overfitting. It acts as an important checkpoint that enables model refinement in a controlled environment that prevents leakage of information about test data. This helps ensure that improvements can be generalized and are not optimized only for the trained data.

Finally, the test set serves as a definitive criterion to assess how the model will perform in practical scenarios. It is only used after the model has been fully trained and validated. This set, which contains new, untested data, is crucial for the final assessment of the model's capabilities. The performance of the model is evaluated using precision, feedback, and F1-scores for each class, offering a comprehensive view of its diagnostic capabilities across a variety of medical conditions.

# Implementation

## Implementation of data preparation and dissemination

This chapter discusses the basic steps involved in preparing medical image data for effective training of machine learning models, which include image conversion, resizing, and normalization to ensure uniformity. It also includes the implementation of various data augmentation techniques to increase the robustness and generalization of the model. These processes include geometric transformations, noise and blur adjustments, and color adjustments, ensuring that the models can accurately interpret medical images in a variety of conditions that occur in the real world.

### Implementation of pre-processing

Proper preparation of medical imaging data is essential to train accurate diagnostic models. This involves several key steps to standardize the images and ensure they are in a form from which computer models can learn effectively.

A screenshot of a cell phone

Description automatically generated

Once converted to PNG format, the images are arranged in a simpler folder system. This organization simplifies the search process during model training, increasing efficiency.

Each image is scaled to a standard size (e.g., 224 × 224 or 299 × 299 pixels) that is needed for consistency in machine learning models. The normalization is used to ensure uniform brightness and contrast, which helps the models in recognizing structural differences rather than illumination differences.

In cases where color information is unnecessary, such as X-ray or MRI images, the images are converted to grayscale. This reduces data complexity and file size, speeds up the training process, and focuses the model on structural information.

Medical images are initially retrieved from the DICOM format, which contains extensive metadata. Extracting only the image data removes unnecessary details to focus the models on visual patterns. These images are then converted to the easier to use PNG format.

Image 12 Preprocessing steps source: author

These preprocessing steps (Fig. 12) are automated by a custom software script. Using this script ensures that each image in the dataset is treated in the same way, which is critical for the reliability of the models. This consistency helps to eliminate potential biases or errors that could arise from manually processing a larger dataset.

### Augmentation techniques used

Data augmentation is a technique used to create a more diverse data set by artificially modifying images. This process helps models learn to interpret medical images under a variety of conditions that they may encounter in a real-world environment.

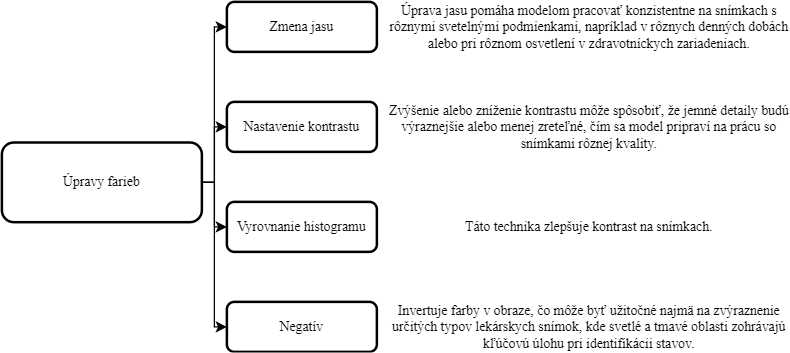
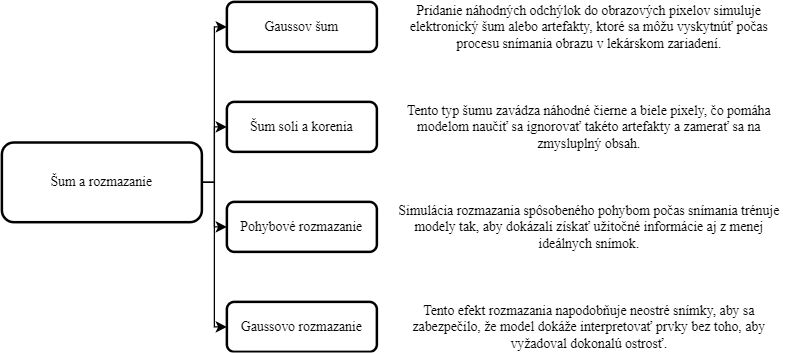
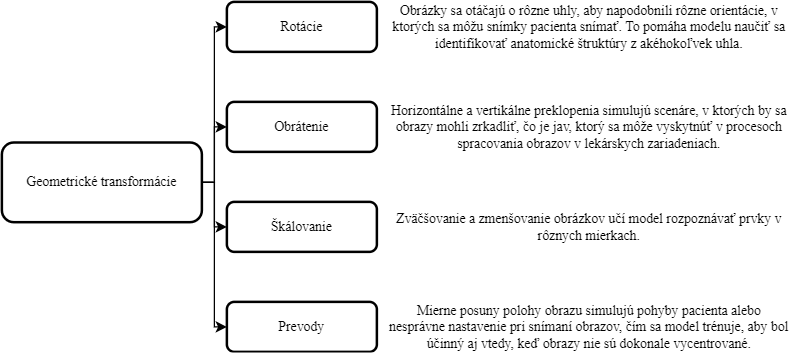


Image 13 Data augmentation techniques source: author

Three different categories of data augmentation techniques were used (Fig. 13)

The extension process uses a robust script that randomly applies these transformations to the images in the dataset. This randomness ensures that the model does not overfit to specific attributes of the training set and can generalize well to new, unseen images. The script systematically stores each augmented image with a unique identifier, thus maintaining an organized and traceable dataset that ensures reproducibility and systematic evaluation.

## Processing and splitting of data sets

This chapter outlines the final preparations needed to ensure that the dataset is ready to train the model, including thorough checks and the use of CSV files to effectively manage the data. It explains the partitioning of the dataset into training, validation, and test sets using a customized approach to address unique constraints such as limited image availability for certain classes. Stratified sampling is used to ensure a balanced distribution across all segments of the dataset, thereby improving model performance and fairness in training.

### Practical overview and data set setup

After pre-processing and extension, final preparations must be made to ensure that the dataset is fully ready for experimentation, which is critical to facilitate seamless model training. This phase includes final checks and setup of the dataset.

First, final checks and configuration will be carried out. All images, now pre-processed and enhanced, are carefully checked to ensure they meet the required quality and format standards for training. This includes verifying that image resolutions and formats are consistent and properly aligned with the model input specifications.

Further, CSV file management plays a key role in the efficient processing of the dataset. CSV files are used to catalog details such as image paths and their corresponding labels. This system facilitates systematic access and manipulation of the dataset throughout the training process. The use of CSV files ensures that each image is accurately labeled and easily retrievable, greatly streamlining the training workflow.

### The process of division

In order to train and validate models efficiently, it is essential to partition the dataset into training, validation and test sets. This structured partitioning is crucial for evaluating models in controlled and unbiased conditions. Due to the unique limitations of this dataset, such as the limited availability of images for certain classes, the common division into 80% training, 10% validation, and 10% testing is not appropriate. Instead, a tailored approach is used to ensure fairness and efficiency of training:

Partitioning the dataset involves allocating a certain number of samples for each class to the test and validation sets, ensuring that even underrepresented classes are adequately evaluated and contribute meaningfully to the model learning process. Although a standard percentage distribution is not used, the majority of the data is still designed for the training set to provide a reliable basis for learning. This is critical for developing strong predictive capabilities.

To achieve a balanced distribution, stratified sampling is used to ensure that each set - training, validation and test - contains a representative proportion of each class. This method is particularly important when dealing with unbalanced datasets, as it prevents the model from developing a bias towards more frequently represented classes.

Automated scripts are used to accurately assign images to each segment of the dataset according to specific numbers defined for each class.

## Model training process

This chapter describes the setup and execution of model training for medical imaging applications using robust hardware and deep learning frameworks such as TensorFlow and Keras to efficiently handle computationally intensive tasks. It details various CNN architectures and training strategies, including the use of pre-trained layers and techniques such as layer freezing and layer skipping to improve generalization and reduce overfitting.

### Setting up the training environment

The training uses powerful hardware including the Nvidia GeForce RTX 3060 Laptop GPU and Intel i7 processor housed in the ASUS TUF Dash F15 (2022). This build ensures efficient handling of computationally intensive deep learning tasks. TensorFlow and Keras are used due to their extensive support for deep learning applications, providing flexible and powerful tools for model development.

### Conducting model training

Different CNN architectures adapted for multi-label classification and augmented with multiple layers (Fig. 14) are used based on their proven capabilities in image recognition tasks.

A black and white rectangle

Description automatically generated

This layer is used to flatten the input data into a one-dimensional array for input to the next layer. It does not affect the batch size, but is necessary because it allows the output of convolutional layers to be processed by densely connected layers.

This layer is used to prevent overfitting and randomly drop units (along with their connections) from the neural network during training. This process helps make the model less sensitive to specific neuron weights, leading to a more general model that performs better on new, unseen data.

This layer is crucial for learning from features extracted from CNN layers. Dense layers are fully connected layers of a neural network where each input is connected to each output using weights (usually followed by distortions). Dense layers add an important level of complexity and learning capability to the model. They are where the neural network has the ability to learn deeply complex patterns in the data.

Depending on the architecture (e.g., DenseNet, ResNet, EfficientNet, and others), the initial convolutional layers are often pre-trained on large datasets such as ImageNet.

Image 14 Architecture of CNN model layers source: author

The initial layers of each pre-trained model are frozen to preserve the general properties that these layers have learned from extensive pre-training. These features are broadly applicable across a variety of images and tasks, providing a strong and relevant foundation that does not require retraining. Freezing these layers speeds up the training process because only the top layers are updated, significantly reducing the computational overhead. This strategy also improves the model's ability to generalize because it prevents overfitting the model to the nuances of a new dataset.

Training is performed over multiple epochs using both training and validation datasets. This approach allows for continuous monitoring and adjustment based on model performance on the validation data, which is critical for tuning the model to achieve the best results.

### Coping with training challenges

Efficient management of computational resources is crucial given the intensive nature of training deep neural networks. Strategies such as batch processing and memory management are optimized to maximize throughput and training efficiency. In specific experiments, class weights are computed and used to address imbalances in class representation. This adjustment helps ensure that the model is not biased towards the predominant classes. Training is continuously monitored for performance metrics such as accuracy and loss, allowing for real-time adjustments to optimize model performance. Graphical visualization of training and validation results helps in assessing model performance and identifying trends such as overfitting or underfitting.

## Preparing for performance evaluation

This chapter details the process of evaluating medical diagnostic models after training, with an emphasis on preserving and accurately assessing model configurations through systematic storage and retrieval techniques. Performance is rigorously assessed using key metrics that facilitate detailed classification reports and visual representations of the data. This comprehensive evaluation ensures models are accurately tested under consistent conditions, highlighting their potential for effective deployment in medical diagnostics.

### Testing and evaluation of models

Once the thorough training phase is complete, the models are immediately subjected to a performance evaluation using a prepared set of test data. This evaluation is crucial to assess the ability of the models to effectively generalize new, unobserved data - a key factor for their potential deployment in medical diagnostics.

The evaluation process starts by fitting each model together with the corresponding MultiLabelBinarizer (MLB). This procedure ensures that all classification labels are preserved and the models can be reliably recovered with their learned configurations for accurate evaluation. Models are saved using the format *model.save('path\_to\_model.h5')*, which provides the learned weights and architecture. The MLB is saved using *joblib.dump(mlb, 'path\_to\_mlb.pkl')*, a step critical for maintaining label encoding consistency between the training and testing phases to ensure labels are interpreted correctly during evaluations.

During testing, the models are reloaded from their stored states to ensure that the evaluation accurately reflects their conditional performance after training. The models are reloaded from their *.h5* files, which contain the entire model architecture along with the trained weights, while the stored MLB is also reloaded, facilitating correct interpretation of the labels during testing.

In addition, the test data is processed using the same methods as the training data to ensure consistency. This step is essential for a valid evaluation because it reflects the conditions under which the models were trained, thus providing a reliable measure of how well the models can perform under practical diagnostic conditions. This integrated approach to model storage and evaluation ensures that each step-from model training to testing-is conducted under consistent and controlled conditions, increasing the reliability and accuracy of performance evaluation.

### Performance visualisation

This phase of the process focuses on the analysis of the model's performance indicators and the visual interpretation of these results, providing a clear and comprehensive assessment of the model's effectiveness.

The evaluation starts by generating a detailed classification report using the classification\_report function of scikit-learn. This report compiles metrics for each class based on the predictions made by the model on the test dataset, including precision, feedback, and F1 scores. These metrics are very important as they provide a detailed view of the model's performance across different classifications.

*Precision* measures the accuracy of positive predictions and indicates how many of the predicted positive cases were actually positive.

*Feedback* (Recall) identifies the ability of the model to find all relevant cases within a class, a key metric for ensuring that no significant state is overlooked.

*The F1 score* provides a balance between precision and feedback, which is essential to assess the overall performance of the model. This balance is particularly important in the medical field, where both overlooking a condition (low recall) and misdiagnosing a condition (low precision) can have serious consequences.

These metrics are selected not only to provide a comprehensive view of the model's capabilities, but also to ensure that any decision regarding its deployment is well-informed. The performance metrics are also visually represented through bar charts that compare different classes. This visualization not only quantitatively highlights the performance of the model, but also serves as a visual summary that helps to effectively evaluate the strengths and weaknesses of the model.

# Results

## Overview of experimental results

This chapter presents the results of a series of experiments conducted to evaluate the effectiveness of deep learning techniques in improving the classification of X-ray images. Based on the methodologies detailed in Chapter 4, these experiments were designed to address critical challenges in the field of medical imaging, in particular, class imbalance and limitations arising from the typically small size of annotated medical datasets.

The research was structured into three key investigations:

* Using class weights to improve the objectivity and accuracy of the model across classes, particularly focusing on those that are underrepresented in the training datasets. This approach tests the hypothesis that balancing the influence of different classes leads to more accurate diagnostics.
* The impact of different data augmentation techniques on model performance, assessing how different modifications to the trained data affect the CNN's ability to generalize to new, unseen images. The goal of this experiment is to increase the robustness of the model, which is a key factor in real-world medical applications.
* A comparative analysis of models trained with original datasets versus models trained with both original and augmented data to determine the overall benefits of incorporating augmented data into the training process. The aim of this comparison is to verify the effectiveness of the augmentation in overcoming the limitations of the dataset.

Each of these experiments was aimed at improving the robustness and diagnostic accuracy of CNNs used in medical imaging. The findings from these experiments are critical in suggesting how deep learning could be more effectively deployed to improve the accuracy and reliability of diagnostic processes in medical facilities.

## Experiment 1 - Effectiveness of class weights

The main objective of this experiment was to determine whether the implementation of class weights could effectively solve the problem of class imbalance within sets of X-ray images. The hypothesis was that class weights would improve the accuracy of the model by increasing its ability to correctly classify underrepresented classes.

Class weights were computed and used to address class imbalances in the X-ray image datasets. These weights provided targeted training on less frequent classes as they were directly related to the frequency of classes in the trained data. The model's loss function was then adjusted during training to include the calculated weights, which motivated the model to prioritize accuracy in underrepresented classes. By addressing the imbalance, this strategy aims to improve the overall diagnostic capabilities of the model.

The effectiveness of the class weights was evaluated in three different deep learning architectures in classifying X-ray images selected from a set of thirteen tested models. The selected models - EfficientNetB2, DenseNet201 and InceptionResNetV2

EfficientNetB2, known for its high baseline accuracy, demonstrated an overall accuracy improvement of 2.74% after the implementation of class weights. In particular, accuracy in class 13 (Lumbar Spine) saw an improvement from 90.91% to 95.24% , indicating the potential of class weights to further fine-tune high-performance models. This improvement highlights the usefulness of class scales in improving the accuracy of classes that already perform well but could still benefit from fine-tuning to achieve optimal results.

DenseNet201, another high-performance model with a different architecture from EfficientNet, showed a 6.65% increase in average accuracy. In particular, for class 4 (Clavicle), its accuracy increased dramatically from 33.33% to 88.89% , demonstrating the impact of class weights on classes that initially had poor recognition rates. This significant improvement illustrates how class weights can significantly increase the sensitivity of the model to features in classes that are less well represented or whose identification is more difficult due to imbalance.

InceptionResNetV2 was chosen for its initially lower accuracy, which provided an opportunity to test whether class weights can significantly benefit models that require more significant improvement. This model showed the most significant improvement, with an 8.32% increase in overall accuracy. Remarkably, class 18 (Skull), which had an initial accuracy of 0% , improved to 75% after the class weights were applied. This improvement not only demonstrates the dramatic impact that class weights can have on improving model performance for underrepresented classes, but also suggests the potential for transforming poorly performing models into more reliable tools for medical diagnosis.

Image 15- Comparison of the performance of different neural network architectures in Experiment 1: This graph shows the classification accuracy of three neural network models - DenseNet201, EfficientNetB2 and InceptionResNetV2 - under two conditions: using the original settings (red line) and after applying adjustments to the class weights (purple line). source: author

The experimental results (Fig. 15) suggest that class weights positively affected the performance of deep learning models, especially for models that initially had difficulty with specific classes. However, this effect was not uniform; different models and classes experienced varying degrees of improvement.

The practical implications for medical diagnosis, although not the primary focus of this research, suggest that carefully adjusted class weights could potentially lead to more accurate detection of rare conditions. This is very important given the high demands on medical imaging diagnostics.

## Experiment 2 - Impact of data augmentation techniques

The main goal of this experiment was to assess how different data augmentation techniques - specifically geometric transformations, color adjustments, and noise addition - affect the performance of CNN models. The goal was to determine which types of augmentation most effectively increase the robustness and generalization capabilities of the models.

This experiment, which focuses on the basic data augmentation strategies detailed in Chapter 4, involved training CNN models exclusively on augmented datasets. The augmentations implemented - geometric transformations, color adjustments, and noise addition - were chosen to challenge model adaptation and learning capabilities and provide insight into the effectiveness of each type of augmentation.

Image 16- Effectiveness of different data augmentation techniques in Experiment 2: Bar graph shows the effect of three types of data augmentation techniques-geometric, color, and noise-on model performance. source: author

The results (Fig. 16) showed that color adjustments had the most significant effect on the performance of the models, suggesting that these features are crucial for the correct classification of X-ray images. The addition of noise was also beneficial, although to a lesser extent, suggesting its role in helping the models distinguish signal from noise. In contrast, geometric transformations had the least impact, suggesting potential alignment issues with clinical outliers that are naturally present in the data.

The positive impact of data augmentation on the CNN models was particularly evident due to the increased variability and volume of the trained data. The use of these techniques directly addresses the problems associated with data scarcity and heterogeneity, which are essential for the development of reliable diagnostic tools.

Color treatments have been shown to be the most effective, probably because they enhance critical features in grayscale images that carry important diagnostic information on radiographs. This type of enhancement appears to fine-tune the sensitivity of the models to subtle variations indicative of medical conditions.

Geometric transformations, on the other hand, yielded the least significant improvements. Due to the standardization of the location of body parts in X-ray images, drastic geometric changes can distort the spatial features that models rely on for accurate classification. This highlights the importance of context in the choice of enhancement techniques - what benefits natural image processing may not translate favorably to medical imaging, where the accuracy of the location of anatomical structures is critical.

The effectiveness of adding noise suggests that models can learn robustness to artifacts that are commonly encountered in clinical settings. However, too much noise can be detrimental and overshadow important features, highlighting the need for a balanced approach to augmentation.

Ultimately, the experiment highlights the potential of data augmentation not only as a tool to address data limitations, but also as a means to substantially improve the robustness and accuracy of CNNs in medical imaging.

## Experiment 3 - Comparative analysis of training on original vs. original + augmented data

The purpose of this experiment is to determine whether incorporating augmented data into the original datasets can improve the performance and generalization of CNN models in contexts with imbalanced classes and limited data availability. The goal is to assess how augmented data affects the diagnostic accuracy and reliability of models in medical imaging.

This experiment uses the augmentation techniques explored in Experiment 2 - including geometric transformations, color adjustments, and noise addition - and compares the performance of the CNN models on the two datasets. The first dataset, consisting entirely of original images, was prepared with a specific split: 5 samples per class for testing, 2 for validation, and the remainder for training. The second dataset combined these original images with their augmented equivalents and was divided using a standard split: 80% for training, 10% for validation, and 10% for testing.

The CNN models used in this analysis were selected based on their performance in Experiment 1. Two models that showed promising results were selected to investigate possible further improvements using augmented training, and one model that performed poorly was included to test whether augmentation could significantly improve its accuracy. This approach provides a comprehensive assessment of how augmentation affects models with different initial performance levels, and offers insight into the potential of augmented training to effectively address the problems of data sparsity and class imbalance.

A comparative analysis of CNN models trained on the original and augmented datasets showed remarkable findings.

DenseNet201, which previously showed an average accuracy of 66.58%, saw a remarkable increase to 90.95% when trained on the combined original and extended dataset. This significant improvement suggests that the augmentation techniques provided the model with a wider range of features to learn, increasing its generalization capability and thus significantly improving its accuracy.

EfficientNetB2 also benefited from extended training, with its average accuracy increasing from 73.40% to 96.05%. This model, which was already one of the better performing ones, showed even greater potential with the enriched dataset, suggesting that the extension may further improve the diagnostic accuracy of the model.

InceptionResNetV2 showed the most significant improvement; from the least accurate model with an average accuracy of 53.02%, it climbed to 89.87% accuracy with extended training. This highlights the potential of data augmentation in improving the performance of models that initially struggle to classify images correctly.

Image 17 - Comparative analysis of model performance with and without data augmentation in Experiment 3: This graph compares the effectiveness of using the original trained data alone versus combining it with data augmentation for three neural network architectures - DenseNet201, EfficientNetB2 and InceptionResNetV2. source: author

In all models, training on the extended dataset yielded clear improvements in performance metrics. The observed gains support the hypothesis that data augmentation can effectively address the limitations caused by imbalanced and insufficient training data, which is often a problem in medical imaging. This improvement is very important given the high level of risk in medical diagnostics, where every percentage increase in accuracy can significantly affect patient outcomes.

The results of this experiment (Fig. 17) show that investing additional time in training CNN models on augmented data leads to a significant improvement in the accuracy of body part classification based on X-ray images. This suggests that the augmented approach is valuable despite the extended training time, especially in settings with imbalanced or limited datasets. Advances in model performance suggest that such models could be effectively used in clinical practice to aid accurate categorization of radiographs, thereby supporting healthcare professionals in diagnostic processes.

# Discussion

This work demonstrated the critical role of specific deep learning strategies in improving the classification of X-ray images through the application of convolutional neural networks (CNNs). By methodically improving the training processes and adjusting the model parameters, the research has significantly advanced the practical and theoretical understanding of medical image analysis.

Experiment 6.2 demonstrated how the use of class weights can effectively address class imbalance, thereby significantly improving model accuracy. This experiment showed that models can achieve better accuracy when the training process is adapted to the frequency of occurrence of classes.

Experiment 6.3 investigated the effects of various data augmentation techniques, with colour adjustments proving particularly influential. This adaptation is crucial for X-ray diagnostics, where accurate interpretation of greyscale images can be crucial.

Experiment 6.4 highlighted the benefits of incorporating augmented datasets into the training regime. It showed that models trained on a combination of original and augmented data not only performed better, but also exhibited better generalization abilities over the unobserved data.

These findings are of fundamental importance to the field of medical diagnostics. By integrating sophisticated data processing techniques and employing advanced neural network training strategies, the accuracy and reliability of diagnostic tools can be significantly improved. These improvements could lead to more effective treatment plans for patients and potentially save lives by providing more accurate diagnostic information.

The trajectory outlined in this research suggests a promising future in which enhanced CNNs are an integral part of medical imaging diagnostics. Continued exploration of optimization of these models may further refine their efficiency and effectiveness, making them indispensable tools in clinical settings.

# Conclusion

The research conducted in this work confirms the significant impact of targeted computational techniques on the performance of CNNs used for medical image classification. The strategic use of class weights and data augmentation not only solved critical problems such as data scarcity and class imbalance, but also paved the way for the implementation of these models in real diagnostic processes.

Although the experiments conducted provide substantial evidence of the benefits of these methods, they also highlight the limitations inherent in any scientific study, such as the specificity of the datasets and the potential need for wider validation in different medical contexts.

Future research could focus on extending the applicability of these findings by integrating the datasets more broadly and exploring other augmentation techniques that could more accurately mimic clinical differences. There is also a significant opportunity to leverage newer, more powerful computational architectures that could further improve the diagnostic capabilities of CNNs. In addition, the development of user-friendly graphical user interfaces (GUIs) or the integration of these models into a mobile or web application could significantly increase their utility in clinical settings. By making the models available via an app, healthcare professionals could use these advanced diagnostic tools directly in hospitals and clinics, facilitating real-time decision-making and improving patient care. Such app development would not only improve the accessibility of these models, but also ensure their practicality and effectiveness in real-world medical settings.

Overall, this work not only contributes to the academic field, but also sets a practical framework for future advances that could revolutionize medical diagnostics. The integration of deep learning into medical imaging offers a promising avenue to improve healthcare delivery, underscoring the value and potential of this research.

# List of literature used

[1] LITJENS, Geert, Thijs KOOI, Babak Ehteshami BEJNORDI, Arnaud Arindra Adiyoso SETIO, Francesco CIOMPI, Mohsen GHAFOORIAN, Jeroen A. W. M. VAN DER LAAK, Bram VAN GINNEKEN and Clara I. SÁNCHEZ. A survey on deep learning in medical image analysis. *Medical Image Analysis* [online]. 2017, **42**, 60-88. ISSN 1361-8415. Available from: doi:10.1016/j.media.2017.07.005

[2] PANCHBHAI, Arati S. Wilhelm Conrad Röntgen and the discovery of X-rays: Revisited after centennial. *Journal of Indian Academy of Oral Medicine and Radiology* [online]. 2015, **27**(1), 90. ISSN 0972-1363. Available from: doi:10.4103/0972-1363.167119

[3] *Über Röntgen - Röntgen-Gedächtnisstätte* [online]. [vid. 2024-04-25]. Available from: https://wilhelmconradroentgen.de/de/ueber-roentgen

[4] FRANKEL, R I. Centennial of roentgen's discovery of x-rays. *western journal of medicine*. 1996, **164**(6), 497-501. issn 0093-0415.

[5] HENDEE, William R. Physics and applications of medical imaging. *Reviews of Modern Physics* [online]. 1999, **71**(2), S444-S450. Available from: doi:10.1103/RevModPhys.71.S444

[6] RAKSHA, D. S. CT Scans VS MRI Scans: What are the Differences Between Them. *Kiran Lab* [online]. 14 July 2023 [accessed 2024-04-25]. Available from: https://kiranpetct.com/ct-scans-vs-mri-scans-what-are-the-differences-between-them/

[7] HARVEY, Christopher J., James M. PILCHER, Robert J. ECKERSLEY, Martin J. K. BLOMLEY, and David O. COSGROVE. Advances in Ultrasound. *Clinical Radiology* [online]. 2002, **57**(3), 157-177. ISSN 0009-9260. Available from: doi:10.1053/crad.2001.0918

[8] SONOGRAPHICTENDENCIES. Chest and Lung Ultrasound. *Sonographic Tendencies* [online]. 18 March 2020 [accessed 2024-04-25]. Available from: https://sonographictendencies.com/2020/03/18/chest-and-lung-ultrasound/

[9] ISHIGAKI, Takeo, Mitsuru IKEDA, Kazuhiro SHIMAMOTO, Hideki HIROTA and Naoki MAKINO. Digital radiology and PACS. *Nagoya journal of medical science*. 1993, **56**, 53-67.

[10] BHADORIA, Sonali and C.G. DETHE. Study of Medical Image Retrieval. In: *2010 International Conference on Data Storage and Data Engineering*: *2010 International Conference on Data Storage and Data Engineering* [online]. 2010, pp. 192-196 [vid. 2024-04-15]. Available from: doi:10.1109/DSDE.2010.55

[11] HAINS, Isla M., Andrew GEORGIOU and Johanna I. WESTBROOK. The impact of PACS on clinician work practices in the intensive care unit: a systematic review of the literature. *Journal of the American Medical Informatics Association : JAMIA* [online]. 2012, **19**(4), 506. Available from: doi:10.1136/amiajnl-2011-000422

[12] BLINOV, N. N., E. B. KOZLOVSKII and O. V. ROMANOV. A New Stage for Standardization of Medical X-Ray Equipment. *Biomedical Engineering* [online]. 2014, **47**(5), 241-242. ISSN 1573-8256. Available from: doi:10.1007/s10527-014-9380-y

[13] VENKATARAMAN, Viswanathan, Travis BROWNING, Ivan PEDROSA, Suhny ABBARA, David FETZER, Seth TOOMAY and Ronald M. PESHOCK. Implementing Shared, Standardized Imaging Protocols to Improve Cross-Enterprise Workflow and Quality. *Journal of Digital Imaging* [online]. 2019, **32**(5), 880. Available from: doi:10.1007/s10278-019-00185-4

[14] BRUNO, Michael A., Eric A. WALKER and Hani H. ABUJUDEH. Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. *RadioGraphics* [online]. 2015, **35**(6), 1668-1676. ISSN 0271-5333. Available from: doi:10.1148/rg.2015150023

[15] DAGVASUMBEREL, Gonchigsuren, Bayarbaatar BOLD and Munkhbaatar DAGVASUMBEREL. The Growing Problem of Radiologist Shortage: Mongolia's Perspective. *Korean Journal of Radiology* [online]. 2023, **24**(10), 938-940. ISSN 1229-6929. Available from: doi:10.3348/kjr.2023.0787

[16] RAMLI, Norlisah Mohd and Norzaini Rose MOHD ZAIN. The Growing Problem of Radiologist Shortage: Malaysia's Perspective. *Korean Journal of Radiology* [online]. 2023, **24**(10), 936-937. ISSN 1229-6929. Available from: doi:10.3348/kjr.2023.0742

[17] *Deep Learning Applications in Medical Image Analysis | IEEE Journals & Magazine | IEEE Xplore* [online]. [vid. 2024-04-15]. Available from: https://ieeexplore.ieee.org/document/8241753

[18] LANGER, Steve G. Challenges for Data Storage in Medical Imaging Research. *Journal of Digital Imaging* [online]. 2011, **24**(2), 203-207. ISSN 0897-1889. Available from: doi:10.1007/s10278-010-9311-8

[19] JAN, Bilal, Haleem FARMAN, Murad KHAN, Muhammad IMRAN, Ihtesham Ul ISLAM, Awais AHMAD, Shaukat ALI and Gwanggil JEON. Deep learning in big data Analytics: A comparative study. *Computers & Electrical Engineering* [online]. 2019, **75**, 275-287. ISSN 0045-7906. Available from: doi:10.1016/j.compeleceng.2017.12.009

[20] SPERLING, Ed. Deep Learning Spreads. *Semiconductor Engineering* [online]. 31 January 2018 [accessed 2024-04-25]. Available from: https://semiengineering.com/deep-learning-spreads/

[21] ZHANG, Qingchen, Laurence T. YANG, Zhikui CHEN and Peng LI. A survey on deep learning for big data. *Information Fusion* [online]. 2018, **42**, 146-157. ISSN 1566-2535. Available from: doi:10.1016/j.inffus.2017.10.006

[22] LI, Zewen, Fan LIU, Wenjie YANG, Shouheng PENG and Jun ZHOU. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems* [online]. 2022, **33**(12), 6999-7019. ISSN 2162-2388. Available from: doi:10.1109/TNNLS.2021.3084827

[23] RAKHLIN, Alexander. *Diabetic Retinopathy detection through integration of Deep Learning classification framework* [online]. B.m.: bioRxiv. 19 June 2018 [accessed 2024-04-15]. Available from: doi:10.1101/225508

[24] SHEN, Li, Laurie R. MARGOLIES, Joseph H. ROTHSTEIN, Eugene FLUDER, Russell MCBRIDE, and Weiva SIEH. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Scientific Reports* [online]. 2019, **9**(1), 12495. ISSN 2045-2322. Available from: doi:10.1038/s41598-019-48995-4

[25] CHEN, Chen, Chen QIN, Huaqi QIU, Giacomo TARRONI, Jinming DUAN, Wenjia BAI and Daniel RUECKERT. Deep Learning for Cardiac Image Segmentation: A Review. *Frontiers in Cardiovascular Medicine* [online]. 2020, **7** [vid. 2024-04-15]. ISSN 2297-055X. Available from: doi:10.3389/fcvm.2020.00025

[26] Explainable Deep Learning Models for Healthcare Decision Support. *International Journal of Advances in Computer Science and Technology* [online]. 2023, **12**(10), 63-69. ISSN 23202602. Available from: doi:10.30534/ijacst/2023/0112102023

[27] Article Detail. *International Journal of Advanced Research* [online]. [vid. 2024-04-15]. Available from: https://www.journalijar.com/article/

[28] KHAN, Asifullah, Anabia SOHAIL, Umme ZAHOORA and Aqsa Saeed QURESHI. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review* [online]. 2020, **53**(8), 5455-5516. ISSN 1573-7462. Available from: doi:10.1007/s10462-020-09825-6

[29] Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network. *upGrad blog* [online]. [vid. 2024-04-27]. Available from: https://www.upgrad.com/blog/basic-cnn-architecture/

[30] *Review of Deep Learning Algorithms and Architectures | IEEE Journals & Magazine | IEEE Xplore* [online]. [vid. 2024-04-16]. Available from: https://ieeexplore.ieee.org/document/8694781

[31] ALFAHDAWI, Mohammed Gharkan, Khattab M Ali ALHEETI and Salah Sleibi AL-RAWI. Object Recognition System for Autonomous Vehicles Based on PCA and 1D-CNN. In: *2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM)*: *2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM)* [online]. 2021, pp. 219-225 [vid. 2024-04-16]. Available from: doi:10.1109/ICCITM53167.2021.9677676

[32] SHIN, Hoo-Chang, Holger R. ROTH, Mingchen GAO, Le LU, Ziyue XU, Isabella NOGUES, Jianhua YAO, Daniel MOLLURA, and Ronald M. SUMMERS. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging* [online]. 2016, **35**(5), 1285-1298. ISSN 1558-254X. Available from: doi:10.1109/TMI.2016.2528162

[33] SAXENA, Akarsh. Building a Simple Neural Network from Scratch. *Medium* [online]. 2 June 2020 [accessed 2024-04-27]. Available from: https://towardsdatascience.com/building-a-simple-neural-network-from-scratch-a5c6b2eb0c34

[34] CHEN, Leiyu, Shaobo LI, Qiang BAI, Jing YANG, Sanlong JIANG and Yanming MIAO. Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sensing* [online]. 2021, **13**(22), 4712. ISSN 2072-4292. Available from: doi:10.3390/rs13224712

[35] ABDELHAMID MAHRI UNIVERSITY, ALGERIA, Sara SABBA, Meroua SMARA, ABDELHAMID MAHRI UNIVERSITY, ALGERIA, Mehdi BENHACINE, ABDELHAMID MAHRI UNIVERSITY, ALGERIA, Loubna TERRA, ABDELHAMID MAHRI UNIVERSITY, ALGERIA, and Zine EDDINE TERRA, ABDELHAMID MAHRI UNIVERSITY, ALGERIA. Residual Neural Network in Genomics. *Computer Science Journal of Moldova* [online]. 2022, **30**(3(90)), 308-334. ISSN 15614042, 25874330. Available from: doi:10.56415/csjm.v30.17

[36] BRITAL, Anas. Residual Networks With Examples. *Medium* [online]. 28 October 2021 [accessed 2024-04-27]. Available from: https://medium.com/@AnasBrital98/residual-networks-with-examples-80b47cacecf4

[37] *U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications | IEEE Journals & Magazine | IEEE Xplore* [online]. [vid. 2024-04-16]. Available from: https://ieeexplore.ieee.org/document/9446143

[38] *42: A blog on A.I.* [online]. [vid. 2024-04-27]. Available from: https://nchlis.github.io/2019\_10\_30/page.html

[39] AYDIN, Halise Nur and Oktay YILDIZ. Improved ECA-DenseNet Framework for Brain MRI Image Classification. In: *2023 31st Signal Processing and Communications Applications Conference (SIU)*: *2023 31st Signal Processing and Communications Applications Conference (SIU)* [online]. 2023, pp. 1-4 [vid. 2024-04-16]. ISSN 2165-0608. Available from: doi:10.1109/SIU59756.2023.10223886

[40] CHEN, Zhimei. Hardware Accelerated Optimization of Deep Learning Model on Artificial Intelligence Chip. *Frontiers in Computing and Intelligent Systems* [online]. 2023, **6**(2), 11-14. ISSN 2832-6024. Available from: doi:10.54097/fcis.v6i2.03

[41] CHEN, Bin, Jiajun WANG and Zheru CHI. Improved DenseNet with Convolutional Attention Module for Brain Tumor Segmentation. In: *Proceedings of the Third International Symposium on Image Computing and Digital Medicine* [online]. New York, NY, USA: Association for Computing Machinery, 2019, pp. 22-26 [vid. 2024-04-16]. ISICDM 2019. ISBN 978-1-4503-7262-6. Available from: doi:10.1145/3364836.3364841

[42] CHLAP, Phillip, Hang MIN, Nym VANDENBERG, Jason DOWLING, Lois HOLLOWAY and Annette HAWORTH. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology* [online]. 2021, **65**(5), 545-563. ISSN 1754-9485. Available from: doi:10.1111/1754-9485.13261

[43] FRID-ADAR, Maayan, Eyal KLANG, Michal AMITAI, Jacob GOLDBERGER and Hayit GREENSPAN. Synthetic data augmentation using GAN for improved liver lesion classification. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* [online]. 2018, pp. 289-293 [vid. 2024-04-16]. ISSN 1945-8452. Available from: doi:10.1109/ISBI.2018.8363576

[44] Generative Adversarial Network (GAN). *GeeksforGeeks* [online]. 15 January 2019 [accessed 2024-04-27]. Available from: https://www.geeksforgeeks.org/generative-adversarial-network-gan/

[45] *The UNIFESP X-Ray Body Part Classification Dataset* [online]. [vid. 2024-04-22]. Available from: https://www.kaggle.com/datasets/felipekitamura/unifesp-xray-bodypart-classification

# List of images used

[Figure 1: Wilheim Conrad Röntgen [3] 7](#_Toc166360575)

[Figure 2: The first X-ray image created [3] 8](#_Toc166360576)

[Figure 3: Example of chest CT and MRI images [6] 9](#_Toc166360577)

[Figure 4: Example of lung ultrasound [8] 10](#_Toc166360578)

[Figure 5: The difference between deep and machine learning [20] 16](#_Toc166360579)

[Figure 6: CNN model architecture [29] 20](#_Toc166360580)

[Figure 7: Weights and biases in neurons [33] 21](#_Toc166360581)

[Figure 8: ResNet model [36] 23](#_Toc166360582)

[Figure 9: U-Net model [38] 23](#_Toc166360583)

[Figure 10: GAN architecture [44] 25](#_Toc166360584)

[Figure 11 Example of the classes of radiographs from the dataset used: 1- Skull (Class 18), 2- Hand (Class 9), 3- Foot (Class 6), 4- Knee (Class 11) Source: author 30](#_Toc166360585)

[Figure 12 Preprocessing steps source: author 38](#_Toc166360586)

[Figure 13 Data augmentation techniques source: author 39](#_Toc166360587)

[Figure 14 Architecture of CNN model layers source: author 43](#_Toc166360588)

[Figure 15- Comparison of the performance of different neural network architectures in Experiment 1: This graph shows the classification accuracy of three neural network models - DenseNet201, EfficientNetB2 and InceptionResNetV2 - under two conditions: using the original settings (red line) and after applying adjustments to the class weights (purple line). source: author 49](#_Toc166360589)

[Figure 16- Effectiveness of different data augmentation techniques in Experiment 2: The bar chart shows the effect of three types of data augmentation techniques-geometric, color, and noise-on model performance. source: author 50](#_Toc166360590)

[Figure 17 - Comparative analysis of model performance with and without data augmentation in Experiment 3: This graph compares the effectiveness of using the original trained data alone versus combining it with data augmentation for three neural network architectures - DenseNet201, EfficientNetB2 and InceptionResNetV2. source: author 52](#_Toc166360591)