Machine learning models for detection of targeted features in pediatric medical X-ray images

Hržić, Franko

Doctoral thesis / Disertacija

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Engineering / Sveučilište u Rijeci, Tehnički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:190:429587

Rights / Prava: Attribution 4.0 International/Imenovanje 4.0 međunarodna

Download date / Datum preuzimanja: **2024-04-16**

Repository / Repozitorij:

Repository of the University of Rijeka, Faculty of Engineering

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Franko Hržić

MACHINE LEARNING MODELS FOR DETECTION OF TARGETED FEATURES IN PEDIATRIC MEDICAL X-RAY IMAGES

DOCTORAL DISSERTATION

Rijeka, 2022.

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Franko Hržić

MACHINE LEARNING MODELS FOR DETECTION OF TARGETED FEATURES IN PEDIATRIC MEDICAL X-RAY IMAGES

DOCTORAL DISSERTATION

Thesis Supervisor: prof. dr. sc. Ivan Štajduhar
Thesis Co-supervisor: Univ. FA Priv.-Doz. Dr.med.univ.
Dr.scient.med. Sebastian Tschauner

Rijeka, 2022.

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET

Franko Hržić

MODELI STROJNOGA UČENJA ZA DETEKTIRANJE CILJANIH ZNAČAJKI U PEDIJATRIJSKIM MEDICINSKIM RENDGENSKIM SLIKAMA

DOKTORSKA DISERTACIJA

Mentor: prof. dr. sc. Ivan Štajduhar
Komentor: Univ. FA Priv.-Doz. Dr.med.univ. Dr.scient.med. Sebastian Tschauner

Rijeka, 2022.

Thesis Supervisor: prof. dr. sc. Ivan Štajduhar

Thesis Co-supervisor: Univ. FA Priv.-Doz. Dr.med.univ. Dr.scient.med. Sebastian Tschauner

This doctoral dissertation was discussed on _____ at the University of Rijeka, Croatia, Faculty of Engineering in front of the following Evaluation Committee:

1. _____

2. _____

3. _____

Acknowledgments

First of all, I would like to express my gratitude to my supervisor, prof. Ph.D. Ivan Štajduhar, for the help, patience, and guidelines during my doctoral studies. Thanks to his motivation and advice, I have started exploring machine learning in medicine, which has become my primary scientific research interest.

Speaking of medicine, I must mention two great physicians who are meritorious for my knowledge of X-ray images: my co-supervisor, Univ. FA Priv.-Doz. Doctor of Medicine, Univ. Dr. Scient. Med. Sebastian Tschauner and Univ. FA Priv.-Doz. Doctor of Medicine, Univ. Dr. Scient. Med. Erich Sorantin. Their enthusiasm and will to expand artificial intelligence methods to radiology was contagious and highly motivating during my doctoral studies. I can strongly claim that this dissertation would not be possible without them, and I can only hope we will never run out of ideas for new research and machine learning applications in medicine.

I want to thank all members of the Department of computer engineering for their support, especially those who contributed the most to making this dissertation possible: Jonatan Lerga, Sandi Ljubić, Diego Sušanj, Ana Vranković Lacković, and Lovorka Malinić. Thank you for all the coffees and numerous talks.

Last but not least, I owe my deepest gratitude to my family for their support, understanding, and motivation when the burden and pressure of PhD were the heaviest. Also, I want to thank my friends: Dajana Barbir, Ana Rijavec, Željka Gligora, Sandra Borjanić, Sanja Mikašinović, and Fran Ledić, for countless memories and happy moments during my PHD studies that will follow me for the rest of my life.

Abstract

In the context of bone-related injuries, radiography is the most commonly used technique for noninvasive diagnosis. Standard procedure involves a technician who obtains a high-quality X-ray image and a radiologist who sets a diagnosis based on the obtained image. To make diagnostic procedures easier, faster and more accurate, computer-aided diagnostics (CADx) systems are being developed to support the radiologists in their decision-making process. In recent years, machine learning (ML) has become the focus of CADx systems development and research because it is capable of seamlessly capturing highly complex distributions. This is especially important in pediatric radiology, where the variations in the data are often very demanding for modeling.

This doctoral dissertation first investigates preprocessing of pediatric X-ray images and the detection of targeted features related to wrist fractures on wrist radiographs of children. Inspecting the dataset of pediatric X-ray images provided by the Division of Pediatric Radiology, Department of Radiology, Medical University of Graz, Austria, the issue of X-ray image alignment and orientation arose. Therefore, the alignment and orientation of X-ray images became the focus of the first developed ML model.

Following image alignment and orientation, targeted features extraction is needed, such that would help radiologists detect wrist fractures on pediatric X-ray images. Two separate ML models for pediatric wrist fracture detection were developed. The first developed model was based on local-entropy bone segmentation, while the second model utilized a YOLOv4 convolutional neural network to cope with the shortcomings of the first developed model.

Besides fracture detection, it is helpful to estimate the age of the fracture. Therefore, another deep learning-based system was developed for fracture age estimation. The developed multimodal system based on fusing multiple X-ray projections (of the same case) with a patient's age and gender information provides uncertainty in its decision. By estimating uncertainty in its decisions, the system becomes more trustworthy for the experts who use it.

Finally, to enhance the visibility of the fractures and tissue obstructed by the cast during the wrist fracture healing monitoring, a CycleGAN-based system was trained for cast suppression in X-ray images. Also, in order to appropriately evaluate the developed system, rigorous quantitative and qualitative evaluation was proposed.

All things considered, merging all developed models into one system creates a backbone of a CADx system capable of providing crucial information about pediatric wrist fractures that

could improve diagnostics and make the whole process less labored for radiologists.

Sažetak

U kontekstu ozljeda povezanih s kostima, radiografija je najčešće korištena tehnika neinvazivne dijagnostike. Standardni postupak uključuje tehničara koji izrađuje kvalitetan radiogram i radiologa koji na temelju dobivene slike postavlja dijagnozu. Kako bi se dijagnostički postupci učinili lakšima, bržima i preciznijima, razvijaju se sustavi računalne dijagnostike (CADx) koji pružaju podršku radiolozima u procesu donošenja odluka. Posljednjih godina strojno učenje (ML) postalo je fokus razvoja i istraživanja CADx sustava jer je sposobno s lakoćom prepoznati vrlo složene distribucije podataka. Ovo je posebno važno u pedijatrijskoj radiologiji, gdje su varijacije u podacima često vrlo zahtjevne za modeliranje.

U ovoj doktorskoj disertaciji prvo se istražuje predobrada pedijatrijskih radiograma i otkrivanje ciljanih značajki povezanih s prijelomima zapešća na dječjim radiogramima zapešća. Pregledom skupa podataka pedijatrijskih slika prikupljenih na Odjelu za pedijatrijsku radiologiju, Odsjeku za radiologiju Medicinskog sveučilišta u Grazu, Austrija, pojavilo se pitanje poravnanja i orijentacije radiograma. Stoga su poravnanje i orijentacija radiograma postali fokus razvijanja prvog ML modela.

Nakon poravnavanja i orijentacije slike, potrebno je izdvojiti ciljane značajke koje mogu pomoći radiolozima u otkrivanju prijeloma zapešća na pedijatrijskim medicinskim radiogramima. Razvijena su dva odvojena ML modela za otkrivanje prijeloma zapešća u djece. Prvi razvijeni model temelji se na segmentaciji kostiju korištenjem lokalne entropije. Drugi model koristi YOLOv4 konvolucijsku neuronsku mrežu kako bi otklonio nedostatke prvog razvijenog modela.

Osim otkrivanja prijeloma, korisno je procijeniti i starost prijeloma. Stoga je razvijen još jedan sustav temeljen na dubokom učenju za procjenu starosti prijeloma. Razvijeni multimodalni sustav temelji se na spajanju višestrukih projekcija radiograma (istog slučaja) s informacijama o dobi i spolu pacijenta te pruža neizvjesnost u svojoj odluci. Procjenom neizvjesnosti svojih odluka, sustav postaje pouzdanijim za stručnjake koji ga koriste.

Konačno, kako bi se poboljšala vidljivost prijeloma i tkiva prekrivenog gipsom tijekom praćenja cijeljenja prijeloma zapešća, razvijen je sustav temeljen na arhitekturi CycleGAN za supresiju gipsa na radiogramima. Također, kako bi se razvijeni sustav prikladno vrednovao, predloženo je metoda rigoroznog kvantitativnog i kvalitativnog vrednovanja.

Uzimajući sve u obzir, spajanje svih razvijenih modela u jedan sustav stvara okosnicu CADx sustava sposobnog pružiti ključne informacije o pedijatrijskim prijelomima zapešća koje bi mogle poboljšati dijagnostiku i učiniti cijeli proces manje napornim za radiologe.

Ključne riječi: strojno učenje; pretprocesiranje pedijatrijskih radiograma; detekcija fraktura zapešća na radiogramima; generativne suparničke mreže; računalom-potpomognuto dijagnosticiranje; interpretabilnost neuronskih mreža

Contents

II   Included Publications   55

Part I

Introduction

Chapter 1

Introduction and Problem Statement

1.1 Computer-Aided Diagnostics in Medical Imaging

Computer-aided diagnostics (CADx) refers to systems that assist physicians and medical experts in analyzing and evaluating medical images [12]. The systems are based on the extraction of various valuable features that support physicians in their decision making. The extracted features are called radiomics, and they can represent different shapes, sizes of individual regions of interest, pixel intensity histograms, texture irregularities, etc. – in other words, anything that can help a physician to set up a diagnosis [58].

Based on the radiomics definition, it can be easily noticed that CADx systems are strongly related to the scientific field of computer vision. In the last decade, the most popular algorithms in computer vision were based on machine learning and artificial intelligence [75]. Therefore, CADx systems have also started relying on those algorithms and methods. Some examples of CADx systems based on machine learning models are as follows: distinction of patient mental conditions [11], classification of colon cancer stages [74], detection of lung nodules and lung cancer [67, 39], retinal vessel segmentation [82] or brain cancer detection [40]. Furthermore, advances in machine learning led to image augmentation techniques that can increase the quality of medical images. A few of such examples are the reconstruction of CT images from X-ray images [83], increasing the quality of medical images by creating super-resolution images [84], or simply generating new artificial magnetic resonance images of a brain [45].

This dissertation will focus on preprocessing and artifact removal from pediatric medical X-ray images, as well as extraction of targeted features related to wrist fractures from pediatric medical X-ray images that can help radiologists and medical experts set up diagnoses.

1.2 Preprocessing of Pediatric Medical X-ray Images

In order to be trained, many machine learning algorithms (especially vastly popular neural networks) require a large amount of high-quality data. This requirement resulted in the generation

of publicly available datasets such as ImageNet (for general purpose data [10]) or CheXpert (for medical data [35].)

To obtain high-quality data, it is desired to have them in a predefined format with correct labels [68, 7]. Aligning the X-ray images in the desired direction is an additional step in post-processing commonly omitted due to the high workload issues of radiologists. In cases where a body part is too long for the X-ray detector, the X-ray image is misaligned on purpose. However, there are cases where the misalignment was made unintentionally – by mistake. Intentionally or mistakenly, it resulted in images appearing rotated on the workstation and saved that way [55, 51]. Furthermore, there are no guidelines on correctly orienting X-ray images, concluding that different hospitals have different X-ray image orientation and alignment policies. This issue makes merging the different hospital databases tedious work or, in extreme cases, impossible. Also, several state-of-the-art machine learning models have trained on adequately oriented and aligned medical images where authors considered it a mandatory preprocessing step [52, 20].

In this dissertation, a novel method based on machine learning algorithms named X-ray Alignment and Orientation Method (XAOM) is proposed to solve a given problem of X-ray images alignment and orientation. XAOM consists of two stages. The first stage is focused on X-ray image alignment based on several experimentally compared algorithms for X-ray plate (X-ray image content) alignment estimation. The second stage's goal is to determine the correct orientation (to a given standard) of the aligned X-ray image provided by the first stage. For the second stage of XAOM, several Convolutional Neural Networks (CNNs) have been tested and evaluated. In the end, the proposed method is experimentally compared with another existing method on the dataset consisting of pediatric X-ray images of 21 different body-part regions. Besides correctly aligning and orienting, X-ray image preprocessing can include various image enhancements, from increasing quality and denoising to enhancing targeted features and artifact removal [56, 21]. Because a part of this dissertation is focused on developing a CADx system for wrist fracture detection on pediatric medical X-ray images, it was necessary to investigate if such a preprocessing method can be developed to help the radiologists in the task mentioned above.

Injured extremities (also wrists) need to be immobilized by a cast to allow their proper healing. Often, the superimposition of the cast can be treated as an artifact over the injured area, making it harder to spot if the injury is properly healing. To cope with this problem, in this dissertation, a generative adversarial network (CycleGAN) model is developed to suppress cast on pediatric wrist X-ray images making fractures proper healing easier to examine by the radiologists [85].

Several model topologies were tested qualitatively (by statistical methods) and quantitatively (by the experts' opinion). The proposed qualitative and quantitative tests were discussed in detail because the utilized model is the pioneer for cast suppression on X-ray images. Hence, the argumentation and motivation behind conducted tests (and evaluations) were a necessity. Also, the developed method can work in the opposite direction – it can add realistically looking

cast to an X-ray image. The generated data can be used to make models detecting other targeted features more robust.

1.3 Targeted Feature Detection in Pediatric Medical X-ray Images

After preprocessing X-ray images and obtaining high-quality data, the focus is set on developing models for pediatric wrist fracture detection. Namely, such models are the basis of the CADx system that could be developed one day to help radiologists in their everyday trauma-diagnosing routine. The detection of wrist fractures from pediatric medical X-ray images was chosen as a task because wrist fractures are one of the most common injuries/traumas among children [37, 28]. A large number of wrist trauma occurrences means that radiologists spend a lot of their time solving cases related to wrist fracture detection (including ulna and radius bone traumas adjacent to the wrist). Therefore, providing help in extracting targeted features such as fracture region, bone contours, or fracture age would greatly help the radiologists set up diagnosis, and possibly lead to faster examination times requiring less effort.

In this dissertation, wrist fracture detection is investigated using two different automation approaches. The first approach is based on local entropy and bone segmentation. This approach relies on detecting humanly described features based on bone irregularities around fractures. To be precise, after segmenting the contours of ulna and radius bones, the proposed approach detects anomalies in the curvature of the bones estimated by polynomial regression. In the second approach, the deep learning model is trained to overcome the shortcomings of the first approach. The second approach, based on a state-of-the-art object detection algorithm, is compared with the experts on a specially designed test. Furthermore, it is also investigated if medical experts benefit from utilizing the developed fracture detection deep learning model in real-use cases.

Besides fracture detection, a deep learning-based system is designed to estimate fracture age to provide additional information about fractures. Knowing the approximate age of bone fractures is relevant information for medical and forensic issues, especially in domestic violence cases. Furthermore, deep learning-based models often lack the interpretability of their predictions. Guided by these insights, the proposed deep learning model based system is the first-ever attempt at fracture age estimation. The system provides an estimation of fracture age, as well as uncertainty estimation of its decision, making the outcome more interpretable by the medical experts (and therefore more helpful in supporting their decisions).

To summarize, the targeted features researched in this dissertation revolve around wrist fracture detection and include bone segmentation with local entropy, fracture detection based on approximation of bone curvature, detection of fractures based on deep learning models, and fracture age estimation (focused on the models' interpretability). All outputs of the developed models can be used as extracted information/features to help radiologists in their decision-making.

Chapter 2

Machine Learning Models for X-ray Images Preprocessing

2.1 Problem Statement and Data overview

As mentioned in section 1.2, the focus of X-ray image preprocessing in this dissertation is on developing a method based on machine learning models for X-ray image content alignment and orientation and another method for artifact (cast) removal. While the X-ray alignment and orientation method (XAOM) aims to be general in its application to the radiograms of different body regions, the artifact (cast) removal is focused only on pediatric X-ray images of the wrist region. Namely, cast removal can be beneficial for fracture detection, which is another research focus of this dissertation.

The dataset utilized in this dissertation is provided by the Medical University of Graz, Department of Radiology, Division of Pediatric Radiology. The dataset counts over $232,000$ pediatric X-ray images of 21 different body regions [59]. A random sample of each of the 21 different body regions is shown in Figure 2.1. As can be seen in the figure, the images came in different projections (lateral, anteroposterior, or oblique) sides (left or right) with or without artifacts (metal or cast). The distribution of the images per body region is shown in Figure 2.2. As can be seen, the dataset is unevenly distributed, which can be an issue for some machine learning algorithms [38]. The reason behind the skewness of the data is simply the uneven occurrence of injuries and diseases related to different body regions. For instance, it is more common for children to injure their arms or leg than the spine. To cope with the problem of data skewness, only a part of the data (sampled from the complete dataset) is used for model training. It is also necessary to mention that from $232,000$ images, some images were erroneously labeled or corrupted, so the actual distribution of images may deviate by a small amount from the numbers presented in Figure 2.2.

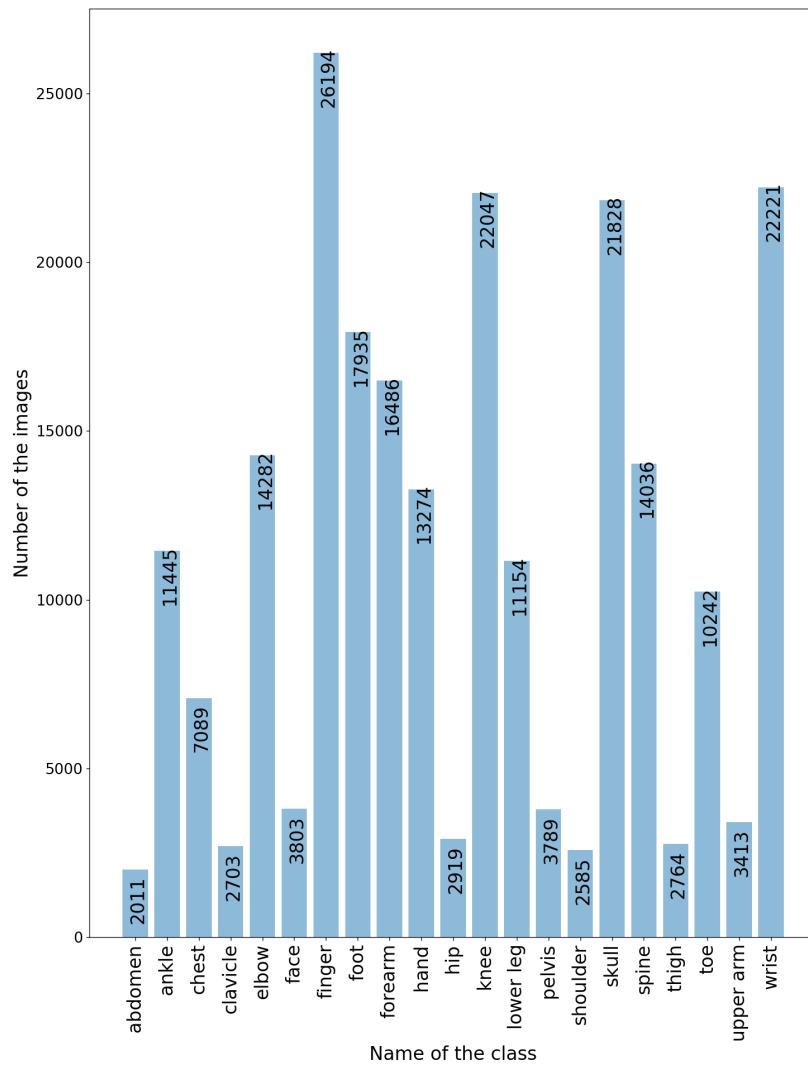Figure 2.1 Sample X-ray images for each of the 21 different body regions [30].

Figure 2.2 Histogram representing dataset distribution per body region [30].

2.2 X-ray Image Alignment and Orientation

Modern radiography mainly employs flat-panel detectors and needle imaging phosphorus rectangular plates of different sizes that offer high digital quantum efficiencies (DQE) and high resolution [8, 60]. However, X-ray image alignment is an additional manual step in post-processing that radiologists often overlook due to workload issues. Yet, many CADx systems require data to be aligned and oriented according to previously determined policy [55, 77, 52, 20, 71]. To cope with this issue, in this dissertation, an X-ray image alignment and orientation method is proposed [30].

The idea for developing XAOM came from the first wrist-fracture detection approach that will be presented later in this dissertation. The method strongly required X-ray images to be aligned properly [32]. After seeing the need for X-ray image alignment and orientation, it was decided to create a general-purpose method that is capable to align any body region depicted on an X-ray image. XAOM is a two-stage method. In the first stage, the plate (content) of the X-ray image is aligned so that it is parallel to the vertical and horizontal axes of the image. Namely, choosing an X-ray plate as the image's content was necessary to make a robust method. In the case of selecting tissue or bone as a content that will serve as a base for image alignment, some body regions will be erroneously oriented (e.g., flipped upside-down) due to the dominant presence of bone or tissue in them (e.g., abdomen or skull). The dataset utilized for the alignment stage consisted of $1,200$ randomly sampled X-ray images. Namely, which body part was on the image was not important because the whole X-ray plate was selected as image content. Furthermore, the five-fold cross-validation was utilized to cope with a relatively small dataset problem during the evaluation phase [16]. The expert determined the ground truth of the X-ray plate's slope by manually labeling the correct alignment of the X-ray plate slope on every image.

The output of the first stage of XAOM (aligned X-ray image) is used as the input for the second stage of XAOM. In the second stage, the orientation of the image is determined by utilizing convolutional neural networks (CNN) [1]. As mentioned before, different hospitals can heave different orientation policies/rules; however, once the images are oriented according to defined rules, they can be easily adapted to any other rules or orientation policies. Therefore, for each of the 21 different body regions, the four orientations N-north, S-south, W-west, and E-east were defined. Although it might seem that the second stage of XAOM is unnecessary, the first stage can still produce erroneously oriented images (but correctly aligned). For instance, the X-ray image can be perfectly aligned but flipped upside-down. Because the second stage of XAOM relies on neural networks, another dataset was sampled that consists of $21,021$ X-ray images in the training set ($1,001$ image per body region) and $4,221$ X-ray images in the test set (201 image per body region). Furthermore, the trivial augmentation was done by rotating every X-ray image after the first stage of the proposed method for $90°$ clockwise, resulting in a total of 84,084 training images. This augmentation was done to equally represent each of the four orientations during the training. The validation subset was created from the training subset by

randomly sampling $16,817$ images ($\sim 1/5$ of the total images in the training set). It was done so to remove any possibility that augmented and original data end up in different subsets. The correct orientation of every image was done by manually examining the X-ray images.

The following subsections will give a more detailed overview of XAOM's stages.

2.2.1   XAOM stage one: X-ray Image Alignment Algorithm

The image alignment (stage one of XAOM) can be divided into three parts:

1. X-ray plate (content of the X-ray image) estimation and enhancement,

2. Estimation of *rotation angle* α from the enhanced content of the X-ray image. The *roation angle* α is an angle for which the content of the X-ray image must be rotated in order to align it with the vertical and horizontal axes of the image,

3. Reduction of the X-ray image to its content (X-ray plate).

In the first part of image alignment, the whole X-ray image is padded with 100 black pixels to each border. Namely, this step is necessary to enhance the difference between the border of the X-ray plate (content) and the image border. In some cases, the small plates indicating the side of the image (letters R and L) can be placed nearby the border of the image. Those small plates can cause issues in X-ray plate estimation. Therefore, to enhance the plate, binarization of the image followed by an operation of morphological closing is performed. First, all pixels with non-zero intensity values are set to 255 and considered as a part of the X-ray image content. Because the utilized binarization process can cause artifacts – regions with pixels intensity value zero, while it is supposed to be 255 (white pixel), morphological closing was utilized to reduce those regions. It was experimentally determined that morphological closing is best performed with 25 iterations of 7×7 ellipse kernel [54]. The leftover regions after morphological closing were filled up by connecting "disconnected" white pixels in every row and column of the X-ray image. The disconnected white pixels are considered two white pixels in the same row or column with at least one black pixel between them. Last step is to estimate a rectangular X-ray plate from the generated white area of pixels. To do so, a minimum bounding rectangle algorithm based on the rotating calipers algorithm was employed [15]. The final result of this – first part of image alignment is an estimated X-ray plate.

In the second part of the image alignment, the slope – *rotation angle* α of the estimated X-ray plate is determined. The *rotation angle* α of the X-ray plate is calculated from the vector that represents the general direction of the estimated rectangle (which represents the X-ray plate). In order to calculate the vector, three different approaches were tested and evaluated:

- **Hough transform based approach** is based on the Hough transform feature extractor, which can be used for line detection [13]. It relies on transforming the points belonging

to the edge of the estimated X-ray plate to another space called Hough space. In Hough space, it is much easier to find lines consisting of the most considerable number of points in the given dataset (in our case, pixels belonging to the edge of the estimated X-ray plate). Also, the approach utilizes various modalities to make it more robust. One of the most common modalities is merging two adjacent smaller lines into one and setting a requirement on the minimum number of pixels needed to form a line. All parameters required by chosen modality were experimentally determined using a grid search method over the parameter space. Once the longest line is estimated, its slope can be used as the slope of the whole X-ray plate (*roation angle* α).

- **Fast line detector based approach** is based on Fast line detector algorithm implemented in OpenCV programming library [49]. The Fast line detector algorithm first detects edges by performing a Canny filter and then extracts line segments as follows: it connects adjacent edge pixels by making straight lines between them. The pixels are all connected until a new added line segment is not co-linear with the previously generated line segment and until the high curvature of the line is preserved. The segments are further joined into the lines based on the line descriptor vectors generated by the mean–standard deviation line descriptor (MSLD) method [79]. The algorithm is far simpler than the Hough transform-based approach in terms of hyperparameters. The Fast line detector requires only two parameters: distance d between two pixels to consider them adjacent and the minimum number of pixels n required for a line segment generated by MSLD to be considered as a line (and not only a segment).

- **PCA based approach** is an approach that is based on the first principal component estimated by the Principal Component Analysis (PCA) method [72]. PCA is a statistical method that extracts linearly uncorrelated variables from a set of linearly correlated variables so that the first linearly uncorrelated variable (the first principal component) preserves the most variability in the provided linearly correlated variables. Each subsequent component contains less and less variability. In the case of X-ray plate direction estimation, the first principal component will preserve the most variability in the given white rectangle, which represents the X-ray plate. This concludes that the first principal component of the given rectangular estimation of the X-ray plate will be the vector representing the X-ray plate's direction. Compared with other approaches, the PCA approach, in this particular case, has no hyperparameters to set.

The experimental analysis has shown that PCA was the best performing line estimation approach and, as such, is used in XAOM for image alignment. Once the PCA approach determines the angle of the X-ray plate, the image is rotated for the calculated angle. In the final-third part of the algorithm, all black pixels around the X-ray plate (the image's content) are deleted, so the final image contains only useful content. This step also saves the memory necessary for image storage. The example of utilizing all three steps is depicted in Figure 2.3.
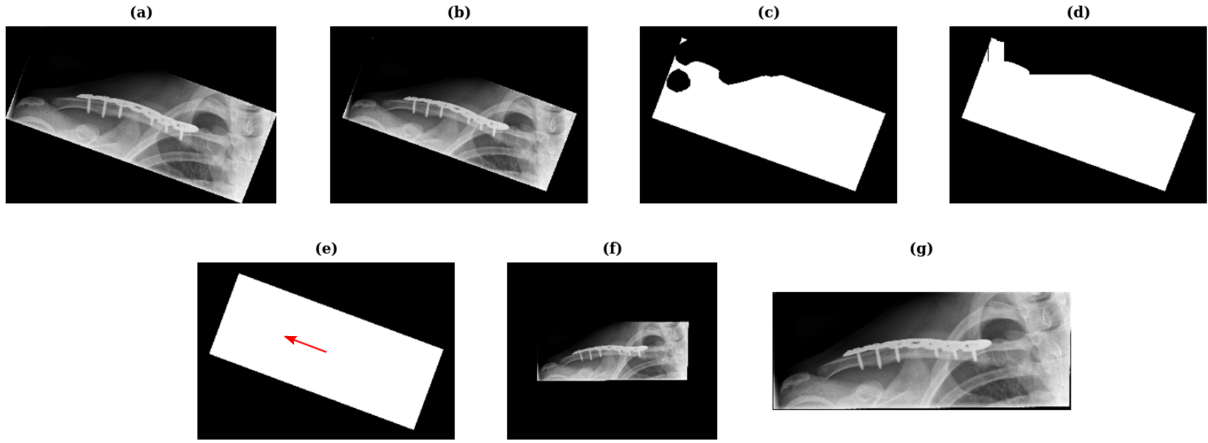
Figure 2.3 An example of the first stage of XAOM: (a) Depicts a randomly selected X-ray image of a right clavicle, (b) Shows the same image padded by 100 black pixels enhancing the image content, (c) Depicts a binarized version of the image after morphological closing operation, (d) Previews the image after connecting the neighboring pixels, (e) Shows estimated X-ray plate by the rotating-calipers algorithm with a red arrow representing the vector of X-ray plate orientation estimated by the PCA, (f) Shows the image rotated by the calculated angle, and finally in (g) is the output of the first stage: an aligned X-ray image with only the useful content preserved. [30].

2.2.2 XAOM stage two: X-ray Image Orientation Algorithm

The second stage of XAOM is focused on predicting the orientation of the X-ray image that was previously aligned in stage one. Namely, the X-ray image is sometimes flipped upside-down or mirrored around the image's vertical axis after stage one. Furthermore, as mentioned in section 1.2, different hospitals can have different orientation policies, which means that, e.g., a wrist with orientation set to the north in one hospital can have a wrist pointing to the top of the image. In contrast, another hospital can consider the north orientation of the wrist when the wrist points to the bottom of the image. Nevertheless, once the north label is determined for an image, it can be easily adjusted to another hospital's orientation rule by rotating the image clockwise for $90°$ until the defined orientation rules are satisfied.

To find the correct orientation, the content of the X-ray image must be examined. Namely, edges and contours do not provide enough information to make one model that will determine the defined orientation of different body parts. Nonetheless, it is probably possible to create a separate model for each body part that determines defined orientation based on the edges and contours. However, this approach is not feasible due to the requirement of training too many models. Therefore, one of the most prominent options to build one model that will successfully determine the orientation of an X-ray image is by utilizing a CNN [1]. To find the best CNN topology for this problem, several popular topologies were tested. Every tested CNN was divided into two parts: the base of the model, which is extracting useful features from the input

X-ray image, and the head of the model, a fully connected neural network that combines the extracted features and produces the final output of the neural network. The output is one of the four labels representing the orientation of the image. Tested CNNs are as follows:

- **LeNet5** is a pioneer CNN. It was first introduced in 1998 in a challenge of classifying hand-written digits [47]. It can be considered the simplest neural network in terms of complexity, layers, and trainable parameters. Since this task is not overly complicated, this neural network was a good starting point in the research of available topologies.

- **AlexNet** is considered as one of the most influential convolutional neural networks [44]. It raised the popularity of CNNs by winning first place in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 competition. In terms of parameters, with eight layers in total (five convolutional and three fully connected), it is more complex than the LeNet5 neural network and, therefore, it is capable of detecting more complex patterns, which can lead to higher model accuracy.

- **VGGNet** was a group of CNNs developed by the Visual Geometry Group Lab of Oxford to compete in the ILSVRC 2014 competition [69, 66]. Although it was a runner-up after the ResNet, it became vastly popular due to its simplicity and utilization of convolutional layers only. VGG topology consists of 16 and 19 weight layers, and nowadays, it is often used as feature extractors in bigger neural network systems.

- **ResNet** CNN continued where VGGNet stopped by accumulating even more convolutional layers and therefore went "deeper" in terms of feature extraction [27]. The ResNet employed skip connections to overcome the issue of vanishing gradients (which is the result of stacking too many layers). This novelty enabled ResNet to have 34, 50, 101, or even 152 weighted layers in the topology, significantly improving its accuracy in given tasks.

The grid search method selected the best hyperparameters for each tested model on the validation dataset. In table 2.1 the best hyperparameter combination of each model is presented. In the given table, softmax and ReLU are activation functions in the head of the model, and the number before the activation function marks the number of neurons in a fully connected layer. Among all tested optimizers, the best performing optimizer for all models was Adam [43]. In Table 2.1 best performing hyperparameters for the optimizer of each neural network are presented. The evaluation has yielded the VGG16 neural network topology as the best one [30]. Last but not least, the proposed method was compared with the existing method for X-ray image orientation proposed by Nose et al. [61]. Although the proposed method based on the calculation of the center of gravity from the image was initially designed only for chest images in antero-posterior projection, it was proven in this dissertation that their method could also be adapted for other body parts. The comparison has proven the significantly better result of XAOM over the existing method on every body region [30].

Table 2.1 Selected parameters for each neural network model used in stage two of XAOM. [30]

Base of the model	Head of the model	Batch size	Hyperparameters
LeNet5	4(softmax)	32	Adam optimizer with learning rate $\alpha = 5 \cdot 10^{-4}$, and coefficients $\beta_1 = 0.99, \beta_2 = 0.999, \epsilon = 1 \cdot 10^{-6}$
AlexNet	4(softmax)	32	Adam optimizer with learning rate $\alpha = 5 \cdot 10^{-5}$,, and coefficients $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \cdot 10^{-6}$
VGG16	4(softmax)	32	Adam optimizer with learning rate $\alpha = 5 \cdot 10^{-5}$,, and coefficients $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \cdot 10^{-6}$
VGG19	4(softmax)	32	Adam optimizer with learning rate $\alpha = 5 \cdot 10^{-5}$,, and coefficients $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \cdot 10^{-6}$
ResNet50	2048(ReLU) $\rightarrow$ 2048(ReLU) $\rightarrow$ 4(softmax)	32	Adam optimizer with learning rate $\alpha = 5 \cdot 10^{-6}$, and coefficients $\beta_1 = 0.9, \beta_2 = 0.9, \epsilon = 1 \cdot 10^{-6}$
ResNet50	4096(ReLU) $\rightarrow$ 4096(ReLU) $\rightarrow$ 4(softmax)	16	Adam optimizer with learning rate $\alpha = 5 \cdot 10^{-6}$, and coefficients $\beta_1 = 0.9, \beta_2 = 0.9, \epsilon = 1 \cdot 10^{-6}$
ResNet50	4096(ReLU) $\rightarrow$ 4096(ReLU) $\rightarrow$ 4(softmax)	32	Adam optimizer with learning rate $\alpha = 5 \cdot 10^{-6}$, and coefficients $\beta_1 = 0.9, \beta_2 = 0.9, \epsilon = 1 \cdot 10^{-6}$

2.3 Cast Suppression in Pediatric Wrist X-ray Images

Another way to increase the quality of X-ray image content is to enhance some targeted features of the image. Usually, this means to de-noise an X-ray image or enhance its brightness, contrast, or edges [34]. In the case of pediatric wrist fracture detection, the typical treatment is to apply a cast over the injured area, which results in the superimposition of the cast over the tissue, bones, and fractures. Ultimately, this makes fractures harder to observe (or see changes in the tissue). In this dissertation, a deep learning model for cast suppression on pediatric wrist X-ray images is proposed and tested.

Cast cannot be treated as noise; it is more like an artifact/style added to the image. Furthermore, the "perfect" pair of two wrist images: one with the cast and one without the cast applied, does not exist, making developing a cast removal model a challenging task. The model must understand what the cast is in the image to achieve cast suppression while preserving bone, tissue, and fractures. With a dataset of $9,672$ pediatric wrist X-ray images ($4,836$ with the cast and the same amount of images without the cast), the decision was made to utilize the generative adversarial network (GAN) based approach called CycleGAN [22, 85]. The dataset was split into training ($7,600$ images), validation (636 images), and test ($1,436$ images) sets.

The CycleGAN architecture enables transfer of images from one domain (X_{domain}) to another (targeted) domain (Y_{domain}) and vice-versa. In cast removal, X-ray images containing the cast were addressed as X_{domain} images, while the X-ray images without the cast were annotated as Y_{domain} images (depicted in Figure 2.4a). To achieve cast removal, the CycleGAN architecture demands the training of four deep neural networks in a fashion of a zero-sum game [9]. During the training, as the input of the CycleGAN system, a pair of images (one from the X_{domain} and one from the Y_{domain}) are randomly selected. The four deep neural networks (two generators and two discriminators) building CycleGAN are as follows:
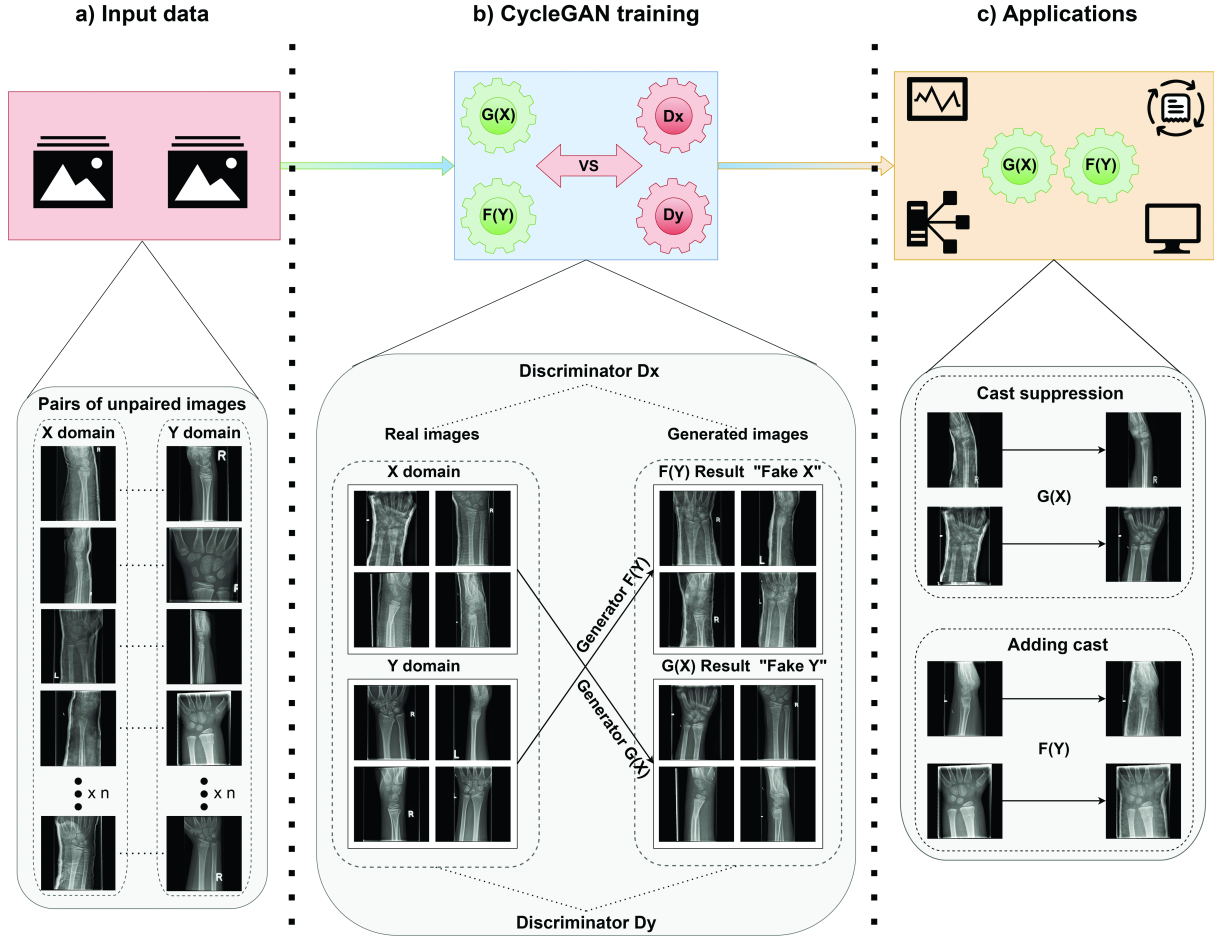
Figure 2.4 CycleGAN for cast suppresion [33].

- **Generator** $G(X)$ is a convolutional neural network that is removing the cast from the Y_{domain} images (images with the cast),

- **Generator** $F(Y)$ is a convolutional neural network that applies the cast to the image from the Y_{domain} (castless) images,

- **Discriminator** D_x is a convolutional neural network with the goal set on distinguishing between real images from X_{domain} images (real images with the cast) and generated images by the Generator $F(Y)$ (generated images with the cast),

- **Discriminator** D_y is a convolutional neural network with the goal set on distinguishing between real images from Y_{domain} (real castless images) and generated images by the Generator $G(X)$ (generated images without the cast).

To better understand the CycleGAN architecture, in Figure 2.4b the system with described four neural networks is depicted. The complete system is trained by utilizing a complex loss function that motivates generators to produce realistic images and discriminators to distinguish real from fake images better. To learn what the cast is, it is necessary to define cycle consistency loss ($\mathcal{L}$) which is given by equation 2.1:

$$\mathcal{L}_{cyc} = Y_{domain} \approx G(F(Y_{domain}) = (X_{domain} \approx F(G(X_{domain})) \tag{2.1}$$

The cycle consistency loss states that an image from one domain transferred with the generator to another domain and then transferred back to the original domain by another generator must be retrieved without any loss of information. For instance, the image with the cast (X_{domain}) is given to the generator $G(X)$ which is removing the cast and making a fake Y_{domain} image. Then the fake Y_{domain} image is transferred back to the (X_{domain}) with the generator $F(Y)$; the image which is the result of this "cycle" must be the same as the original (X_{domain}) image that was given as the input. This loss motivates the CycleGAN system to obtain two crucial qualities: learning what the cast is and the ability to preserve important details in the image (in this case, tissue and fractures).

Another problem tackled during this research was the evaluation of the results. Since this is the first-ever attempt at cast removal, developing a rigorous evaluation system of models' performance was necessary. Five topologies based on ResNet and U-Net architectures were tested for generators [27, 65], while discriminator neural networks due to the efficiency were PatchGANs [50]. The evaluation involved quantitative and qualitative evaluation.

Quantitative evaluation utilized correlation [57], histogram intersection [36], chi-squared distance [17], and Hellinger distance [46] measures given by equations 2.2, 2.3 2.4, and 2.5 respectively. In the equation, H_1 and H_2 respectively represent histograms of generated X-ray images without the cast, and the average histogram calculated from the histograms of the original images without the cast in the test subset of images.

$$d_\rho(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}}. \tag{2.2}$$

$$d_\cap(H_1, H_2) = \sum_I min(H_1(I), H_2(I)). \tag{2.3}$$

$$d_{\chi^2}(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}. \tag{2.4}$$

$$d_H(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{\bar{H}_1 \bar{H}_2 N^2}} \sum_I \sqrt{H_1(I) \cdot H_2(I)}}. \tag{2.5}$$

The qualitative evaluation involved three radiologists. The radiologists needed to rank five tested generator topologies based on the quality of generated castless images. Each radiologist got 20 randomly selected images that had cast removed by each of the tested topologies. Based on their rankings, the best topology was determined.

Both: qualitative and quantitative evaluation yielded U-net with an input image size of 512×512 pixels as the best performing generator topology. Although the aim of the Cycle-GAN was cast suppression, it is important to note that it is also possible to apply cast to the

castless images, which can help increase the number of data for some other tasks, or simply it can be used for educational purposes. As a final product, an application for cast removal based on the interpolation between the original image with the cast and generated castless image was developed (Figure 2.4c).

Chapter 3

Machine Learning Models for Targeted Features Extraction

3.1 Machine Learning Models for Pediatric Wrist Fracture Detection

Wrist fractures, as mentioned in section 1.3, are one of the most common fractures among children. This resulted in a relatively large number of medical cases, which made collecting data easy. As presented in Figure 2.2, the amount of cases related to wrist fractures is also significant ($22, 221$ cases) in the dataset utilized in this dissertation. The more challenging part is the annotation of fractures in the X-ray images. The process of fractures annotation in the utilized dataset took three years in a double-blind labeling fashion.

Nevertheless, easy access to a significant number of X-ray images resulted in the rise of research focused on fracture detection and segmentation [63, 4, 19, 71, 81, 41, 51, 62]. Although none of the mentioned research deals with pediatric wrist fractures, they provide valuable insights into fracture detection models. Therefore, the models for adult fracture detection can be examined as starting points for developing a model for pediatric wrist fracture detection. Pediatric wrist fractures could be more demanding than adults due to the varieties in bone growth, the presence of growth plates and the range of possible injury patterns in children. Most of the mentioned research is based on deep learning models with limited explainability of its decisions/outputs. Therefore, the first approach presented in this dissertation aims to detect fracture regions with a fully explainable approach based on entropy and bone contours. The second approach utilizes the YOLOv4 deep learning model to estimate fracture regions by placing bounding boxes around the fractures [5]. Last but not least, the developed model is compared with the radiologists to get the most realistic estimation of the developed model's performance.
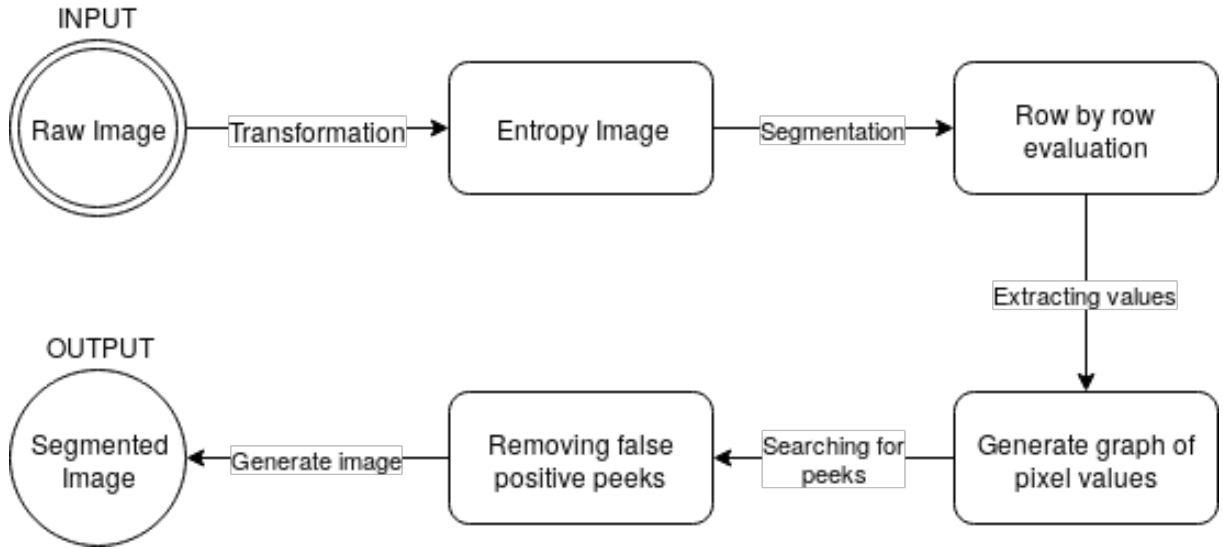
Figure 3.1 Diagram of the developed algorithm based on local entropy [32].

3.1.1 Local-Entropy Based Model

The first developed approach for fracture detection was a fully explainable approach based on bone segmentation and detection of irregularities in segmented bone shape. Namely, the easiest way to explain the common fracture to a non-expert is through bone irregularities. The developed algorithm is fairly complex and is built from several modules shown in Figure 3.1. The raw image is a pediatric X-ray image of the wrist that was previously aligned by a method similar to XAOM. More precisely, the developed alignment method was a foundation for XAOM. XAOM can be observed as a general case of the developed alignment method explicitly tailored for wrist alignment. X-ray image alignment was crucial for the bone segmentation part of the developed algorithm. One example of aligned X-ray image is depicted in Figure 3.2a.

To detect fractures, it was required to segment bone contours, as well as the bone region – the bone itself. To enhance the bone contours, a local-entropy-based approach was adjusted and tested against other popular edge detection algorithms such as Canny filter, Sobel operator, and Laplacian edge detector [3, 73]. The local-entropy approach calculates Shannon entropy on image patches. Each image patch is generated by a 2D sliding window where the window includes an area of 9×9 pixels [2]. To elaborate, sliding windows can be observed as a 2D convolution over the signal where the Shannon entropy calculation substitutes the multiplication with the convolution kernel. Shannon entropy is given in equation 3.1 where $P(x_i)$ represents a distribution of pixels' intensities inside the window, and b is a base of logarithm set to $b = 2$. Namely, in the windows where only tissue is present or the whole bone, entropy will be 0 or close to 0 since there is low variation between the pixels' intensities values. In case the observed window contains part of the bone edge, the pixel intensities values will differ more and consequently raise the entropy of a window. The entropy calculated for a window is a value of a pixel with the position in the center of the window in the new image called entropy image.

Furthermore, the local entropy was multiplied by the standard deviation $\sigma(X)$ of the pixels inside the observed window to emphasize the differences in entropies. Therefore the final equation for entropy calculation is given in equation 3.2. After calculating the entropy of every window in the X-ray image, the resulting image with enhanced bone contours is given in Figure 3.2b.

$$H(X) = \sum_{i=1}^{n} P(x_i)I(x_i) = -\sum_{i=1}^{n} P(x_i)log_b P(x_i) \tag{3.1}$$

$$V(X) = \sum_{i=1}^{n} P(x_i)I(x_i) \cdot \sigma(X) = -\sum_{i=1}^{n} P(x_i)log_b P(x_i) \cdot \sigma(X) \tag{3.2}$$

The next step is to preserve only pixels that are part of the bones and suppress the tissue even more. The tissue suppression is done by "Row by row evaluation": the entropy image is scanned top-to-bottom, and the following two steps are performed for each row. Firstly, the distribution of the intensities of the pixels in the row is plotted, and local maxima are detected. Secondly, detected local maxima are filtered by merging neighboring local maxima to a single local maximum. In addition, local maxima with pixel intensity zero are eliminated. The described process resulted in the final segmentation of bone pixels, which can be seen in Figure 3.2c. After comparing the local-entropy-based approach with other edge detection algorithms, it achieved the lowest segmentation entropy, which was taken as a measure of image "pureness" [24].

After the described suppression process, the image contains only pixels related to bones and potentially some artifacts. Therefore, the bone regions can be extracted from the image. The bone region must be cropped on top and both the left and right sides (the bottom side is where the bones "begins," so it does not need cropping). A similar technique performed in "Row by row evaluation" was developed to detect cropping lines. The top cropping line was found as a line with a maximum number of white pixels. Namely, the end of the radius and ulna bones commonly contains many white pixels. Left and right vertical cropping lines are determined as the average of all left and most right maximum values in every row. Namely, the most left and the most right maximum values in each row represent the bones' outer left and outer right contours. Once the coordinates are found, they are expanded by 20%, so the cropped image contains some additional area around the bone (depicted in Figure 3.2d).

Once the cropped image is obtained, bone extraction is performed on the lines representing bone (Figure 3.2e). This was done by generating a graph from pixel values. For each white pixel in the bottom line of the image, the algorithm tries to build a line towards the top of the image by connecting it to the neighboring white pixels. The beginning pixel (white pixel in the bottom line) is connected to the neighboring pixel with the lowest Euclidean distance. The algorithm takes the last connected pixel as a new beginning pixel and proceeds to seek its neighbor pixel. This way, multiple lines representing bone contours are created. Because of blurry bone edges, two lines might belong to the same bone contour. In that case, those lines are approximated by the average line between them. Once the bone lines are estimated, as shown in Figure 3.2f, an algorithm can be designed to detect fractures in them.

Fractures in the bone lines can be seen as irregularities in their flows. It was empirically proven that the third-degree polynomial can be fitted to points of each bone line. The difference between the estimated polynomial and the estimated bone line is negligible if the bone is not fractured. On the other hand, in the case of a fractured bone, at the spot of fracture, the significant deviation of the estimated bone line from the "healthy" bone estimated by the third-degree polynomial is present. Example of deviation for one bone line can be seen in Figure 3.2g. Mean squared error (MSE) was utilized to measure the deviation (Figure 3.2h). The fracture was marked if the MSE was greater than 3 pixels. The threshold of 3 pixels was empirically found as the best-performing threshold on the training dataset. Although the utilized method achieved satisfactory results on the test set of 860 images, the proposed method can only detect fractures where the bone contours are not regular. This means that any fracture which does not impair bone contours will be undetected. For this reason, a more robust method based on a deep CNN was created. To cope with interpretability issues of the model's decision, a goal was set that the CNN must produce bounding boxes around fractured bone regions. Namely, a deep CNN can learn many different filters on its own to detect different bone fractures.
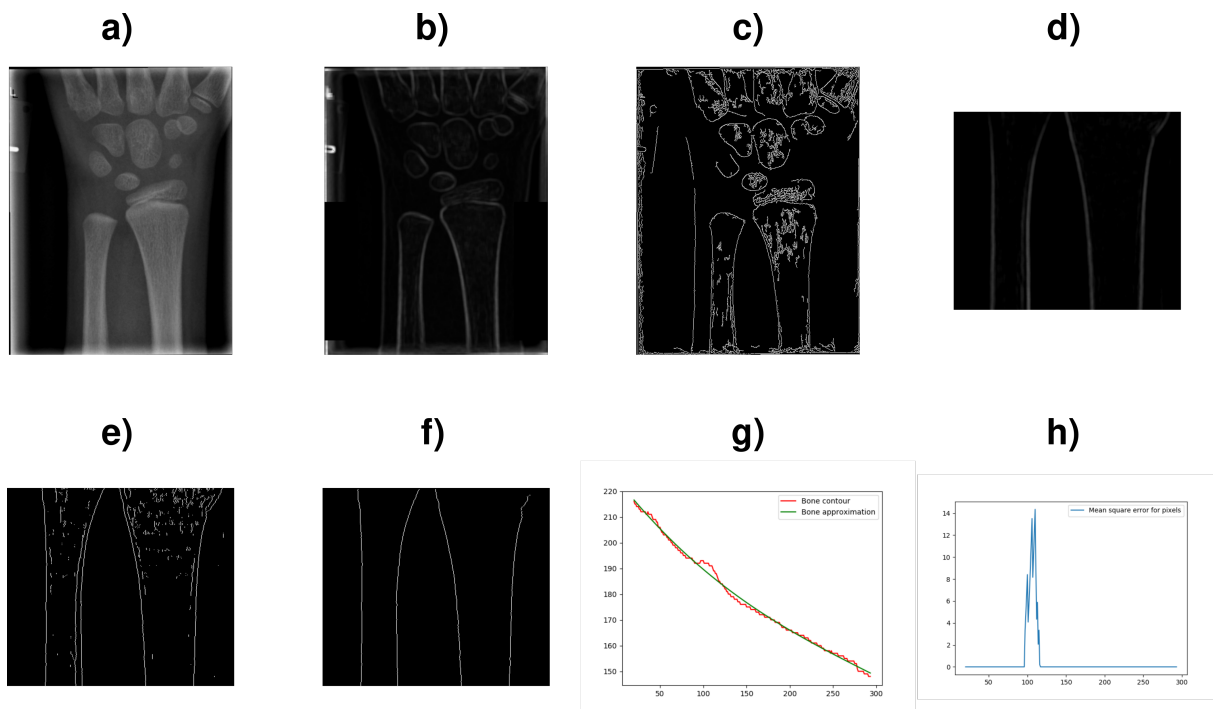

Figure 3.2 Workflow of the proposed algorithm based on local entropy. a) Input X-ray image, b) Image generated by local-entropy, c) Result image of "row by row" evaluation of pixels intensities d) Cropped bones region e) Peak values for bone extraction f) Lines representing bones g) Deviation between estimated bone line and healthy bone line approximated by a third-degree polynomial g) MSE error denoting the fracture [32].

3.1.2 Deep Learning Based Model

To make a more robust method that will include not only bone contours but also different patterns to detect fractures on pediatric wrist X-ray images, a deep learning approach based on CNN is one of the possible solutions that has been proven to achieve sufficiently good results in similar tasks. After carefully evaluating existing solutions for adult fracture detection, the hypothesis "Object detection approach is more informative than the segmentation or classification approaches" was set. To test this hypothesis, the current state-of-the-art method called You Only Look Once version 4 (YOLOv4 [5]) was trained and compared against the current state-of-the-art method for the adult wrist fracture segmentation proposed by Lindsey et al. [51].

It is a known fact that deep learning models require a considerable amount of labeled data [10]. Therefore, the first step was to collect images and correctly label them. The utilized dataset consisted of $19,700$ 8-bit labeled pediatric wrist X-ray images collected from $10,150$ unique studies of $5,997$ unique pediatric patients, acquired between 2008 and 2018 by the Division of Pediatric Radiology, Department of Radiology, Medical University of Graz, Austria [59]. The collection was randomly split into three subsets: a training dataset with $15,600$ images, a validation set with $1,950$ images, and a test set with another $1,950$ images. The remaining 200 images were left as a separate test set for model evaluation against the radiologists.

The utilized YOLOv4 model consists of four different parts:

- **Input data** part, as the name points out, handles preprocessing and augmentation of the input data. All images were first preprocessed with XAOM [30] and then augmented with *bag of freebies* and *bag of specials*, proposed in the original YOLOv4 paper [5].

- **Backbone network** part serves the role of a feature extractor, similar to the VGG neural network in XAOM's orientation part. The main difference is that the backbone of YOLOv4 was chosen as the best-performing backbone network presented in the original paper called CSPDarknet53 [78].

- **YOLOv4 neck** part merges and combines features from different convolutional layers. For this part, PANet [53] was chosen as a neural network topology according to its best performance in the original YOLOv4 paper [5].

- **YOLOv4 head** part provides the final output of the YOLOv4 model – bounding boxes around detected objects. The YOLOv4 head is the same as the head used for the YOLOv3 version due to its good performance [64].

Since the YOLOv4 outputs are bounding boxes around detected objects, setting the initial size of bounding boxes to be close to the real size of objects in images can enhance prediction accuracy. These bounding boxes are called anchors and are estimated by running k-means clustering over the training dataset with a number of clusters set to nine [25]. The number of

Table 3.1 Input sizes and anchor shapes for four tested YOLOv4 models [31].

Model name	Input image size	Anchors
YOLO 512	512×512	$(12, 16), (19, 36), (40, 28), (36, 75), (76, 55),$ $(72, 146), (142, 110), (192, 243), (459, 401)$
YOLO 512 Anchors	512×512	$(23, 22), (44, 27), (35, 40), (73, 31), (55, 41),$ $(61, 62), (84, 47), (101, 67), (92, 115)$
YOLO 608	608×608	$(12, 16), (19, 36), (40, 28), (36, 75), (76, 55),$ $(72, 146), (142, 110), (192, 243), (459, 401)$
YOLO 608 Anchors	608×608	$(40, 41), (70, 51), (109, 58), (82, 82), (157, 70),$ $(122, 94), (114, 155), (176, 111), (206, 176)$

clusters is selected on the principle that for each position (cell) inside the image, YOLOv4 estimates up to nine different objects during the prediction phase. Table 3.1 presents the parameters for four different YOLOv4 variants. Since the input image size and anchor sizes influence the output accuracy of the YOLOv4 method the most, the main focus was put on their estimation. Besides the mentioned hyperparameters, other hyperparameters were the same for all tested models and were chosen based on the literature overview and search over the hyperparameter space. The utilized hyperparameters are as follows: the optimizer was stochastic gradient descent with learning rate $\alpha = 0.0013$ multiplied by factor 0.1 after 80 epochs, momentum was set to 0.949 with decay of 0.0005 (burn-in for $1,000$ steps), batch size was set to 64 with 32 subdivisions and dropout regularization (dropout rate was set to 0.1) [6, 76]. All tested models converged between $\approx 100 - 150$ epochs.

3.1.3 Models Evaluation

The evaluation was performed in three stages to evaluate trained models properly. The first stage compares developed YOLOv4 models with state of the art model proposed by Lindsey et al. [51]. Namely, their developed model based on U-net neural network topology by all defined standards was the most robust developed model for adult fracture detection. The standards included performance evaluation, comparison with radiologists, variety of the dataset, and amount of data in the utilized dataset. These standards were defined and elaborated in the paper, which is part of this dissertation as a fundament for developing a good CADx system for fracture detection [31]. The YOLOv4 models were compared with the U-net state-of-the-art method on three different levels:

- First models' evaluation was a simple binary evaluation in which models had to correctly predict if a given X-ray image contains a wrist fracture or not. In this evaluation, models do not need to predict the location of the fracture.

- Image-based evaluation compared models based on the number of fractures they have detected. This evaluation was crucial for models' training because it gave a rough estimation of models' performance. Since the U-net model segments fractures, an algorithm for fracture number estimation was developed. In the case of the YOLOv4 models, the number of fractures detected was the same as the number of bounding boxes in the image.

- Lastly, the third fracture-based evaluation was the most rigorous one. In this evaluation, every fracture and the model's prediction for that fracture was manually examined, giving the most realistic measure of the model's performance. The correct fracture prediction was determined based on overlapping regions between the model's prediction and the ground truth bounding box of the fracture. To count the prediction of fracture location as a correct one, the overlapping between the predicted fracture bounding box and the fracture's ground truth bounding box had to be over 50%.

Models' evaluation was performed on $1,950$ images. All YOLOv4 models outperformed the U-net model with statical significance [31]. Between the YOLOv4 models, the difference was not statistically significant. However, YOLO 512 Anchors model achieved the overall best result. Hence, the YOLO 512 Anchors model was used for comparison with radiologists. The statistical significance was determined by McNemar's test with confidence level set to $p < 0.05$ [42]. In the next level of evaluation, the YOLO 512 Anchors model was compared against five radiologists.

To compare the model against radiologists, the dataset of 200 images were split into two subsets of 100 images each on which the model performed similarly. This way of splitting the dataset was needed to test the hypothesis that radiologists perform better when aided by a machine learning model than when operating unaided – the third level of evaluation. On 100 images, YOLO 512 Anchors model statistically outperformed all five radiologists (four of them with confidence level $p < 0.05$, and one with $p < 0.01$). Once more, to determine statistical significance, McNemar's test was conducted.

On the remaining 100 images (from the 200 images test set), radiologists could see the decisions made by YOLO 512 Anchors model. Since the model had the same performance on both subsets, they were equally demanding in terms of fracture complexity. This observation is important for statistical evaluation. Namely, utilized two-tailed independent samples t-test [86] ($df_1 = 129, df_2 = 124$) proved that two radiologist performed significantly better with the aid of the developed model ($p < 0.05$), one radiologist performed better when aided by the model at a significance level $p < 0.10$, while two radiologists performed better with the aid of the model, however statistically insignificant. To conclude, the developed model outperformed both the current state-of-the-art method and radiologists and increased radiologists' performance in pediatric wrist fracture detection.

Furthermore, after inspecting the related work, guidelines and observations of good practice were put together for other researchers to follow. The guidelines mainly focus on how to conduct

a proper evaluation, collect and build a proper dataset, and benchmark the developed models against the experts. To help the scientific community, as a part of this dissertation, an expanded version of the dataset utilized in this experiment was made publicly available [59]. Namely, the dataset utilized in the presented research was enhanced with different tags in order to provide additional data for different evaluations or studies.

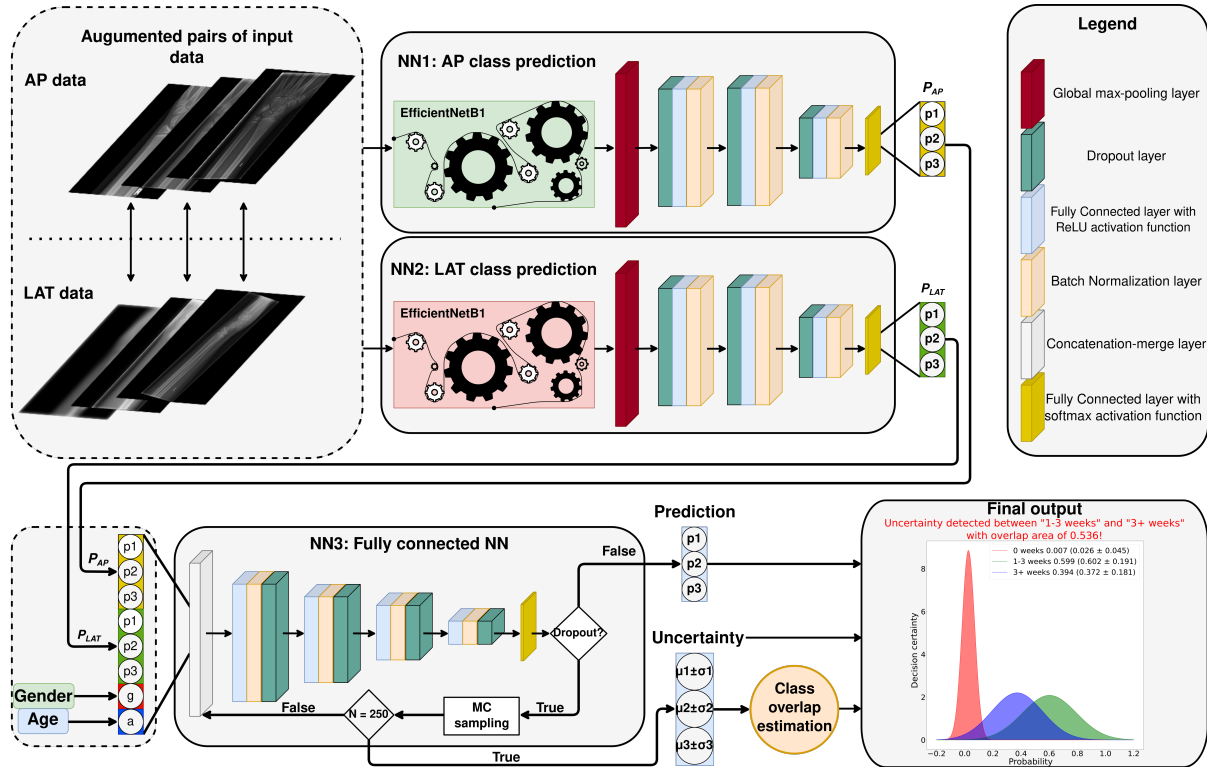## 3.2   Pediatric Wrist Fracture Age Estimation

Figure 3.3 Diagram of the developed system for fracture age estimation [29].

In addition to fracture detection, helpful information is the time when the fracture occurred. Therefore another CADx system based on a deep neural network has been developed. To estimate the age of the fracture, it was necessary to use all available information about the patient and its fractures: lateral and anteroposterior images of their wrist, patient age, and gender. The dataset provided by the Division of Pediatric Radiology, Department of Radiology, Medical University of Graz, Austria, contained $3,570$ studies with all four information necessary for building this CADx system. The issue of the dataset was an uneven distribution of fractures' age: the dataset contained more new fractures (0 days old) than all other "older" fractures combined. To solve the skewed distribution problem, the cases were split into three distinctive groups: "fresh" fractures (0 weeks old), half-a-month old fractures ($1-3$ weeks old), and "old" fractures (fractures older than 3 weeks). With the introduction of data grouping and modifying loss function, the training of neural networks was enabled [38].

26

The inspiration behind the proposed system lay in the solutions of "The RSNA Pediatric Bone Age Machine Learning Challenge" [23]. Namely, the proposed system is the first system dealing with fracture age estimation, which means that the only related work can be found in bone age estimation. Following good practice in systems estimating bone age from X-ray images, the proposed system, depicted in Figure 3.3, consisted of two parts [48, 80].

The first part of the proposed system utilizes two neural networks based on the Efficient-NetB1 neural network topology [70]. One neural network estimates fracture age from lateral projection images, while the other one estimates fracture age from anteroposterior projection images. Both outputs are stacked into a vector and merged with the patient's age and gender to form a new input vector for the second part of the proposed system.

The second part of the proposed system takes the generated vector from the first part and estimates the final age of the fracture by outputting the probability distribution for the possible output classes. The second part of the system is a fully connected neural network with dropout layers. The dropout layers play a key role in the uncertainty estimation of the system outputs. Namely, when the system outputs the probability that the input vector belongs to one class, it is crucial to know how confident the model is in its decision. To obtain this information, Yarin and Zoubin have proven in their paper that a neural network with dropout regularization is mathematically equivalent to a Bayesian approximation of the Gaussian process [18, 14]. In other words, utilizing the Monte Carlo method with dropout active during inference makes it possible to obtain different fracture age predictions for various sampled fully connected neural networks [26]. The neural networks are sampled from the initially trained neural network. Based on the predictions, generating a Gaussian distribution that represents uncertainty for each class is possible. Lastly, the distributions can have overlapping regions. Based on the overlapping, a simple rule-based system can be created to include uncertainty in the final model prediction. The rules of the rule-based system determine the prediction as a regular only if it has high-class probability and model confidence in it is high. High model confidence means that the highest class probability is not overlapped with the second highest class probability by more than 50%.

To conclude, the uncertainty estimation provided additional explainability of the developed system decisions leading to a more reliable system to use. This was important because even for the experts, estimating fracture age is a demanding task where the CADx system can be of great help. As mentioned before, the potential application of fracture age estimation can be found in forensic medicine, especially in domestic violence prevention.

Chapter 4

Conclusion

4.1 Main Scientific Contributions of Dissertation

In this dissertation, multiple novel applications of machine learning models for pediatric wrist X-ray image preprocessing and targeted feature extraction are discussed and presented. Enclosed journal papers show in-depth methodology and results of developed models. The primary connection between all the papers is the motivation to build a CADx system that will support experts in their decision-making concerning pediatric wrist fractures. It was shown that developed machine learning models successfully solved targeted tasks and currently represent state-of-the-art solutions for those tasks.

The main scientific contributions can be summarized in the following points:

- A novel two-stage method was developed based on PCA and VGG16 convolutional neural network for medical X-ray image alignment and orientation of 21 different body regions. The developed X-ray alignment and orientation method (XAOM) is a single model that can easily adapt to any hospital X-ray orientation policies. Furthermore, XAOM was tailored to be easily expandable to other body regions (besides the initial 21 regions), making it an out-of-the-box solution. The obtained experimental results were compared with another developed method based on the X-ray images' center of gravity. The comparison showed that XAOM obtains significantly better results on every tested body region.

- A novel approach based on deep neural networks was proposed for cast suppression from pediatric wrist X-ray images. This approach was the first attempt to suppress cast on pediatric wrist X-ray images and generally the first attempt to remove cast from X-ray images. The approach was based on the CycleGAN method, which utilized U-Net neural network topologies as backbone. Since there was no comparable work, a rigorous evaluation process was proposed. The evaluation employed quantitative and qualitative analysis of model performance. Both confirmed that cast suppression by the proposed model was of high fidelity. Lastly, the application based on the trained CycleGAN model was developed and demonstrated on a real-world use case.

- A novel approach based on machine learning algorithms for pediatric wrist fracture detection from X-ray images was proposed. The approach is based on local-entropy bone segmentation that was proven superior to the other tested methods for bone segmentation. Furthermore, the approach successfully extracts bone regions and estimates the bone contours with a developed graph-theory-based method. Last but not least, a novel approach based on polynomial regression for fracture detection and localization was developed. The proposed approach is fully explainable and provides complete introspection into why the model considers a bone to be fractured. This characteristic is often a flaw of the neural network-based models making it a strong benefit of the developed approach.

- A computer-aided diagnostic system was developed based on the YOLOv4 object detection approach for pediatric wrist fracture detection on medical X-ray images. The developed CADx system was compared with the current state-of-the-art algorithm and medical experts. In both cases, it comes on top: it outperformed the current state-of-the-art model and has proven that it can perform significantly better than the radiologists. Along the same lines, it was proven that radiologists could perform similarly or better when the developed CADx system aids them. The evaluation was performed on the representable dataset that consisted of $19,700$ annotated X-ray images. The evaluation process revealed a problem in models' comparison, which ultimately resulted in writing a set of guidelines for developing a reliable CADx system for pediatric wrist fracture detection and its evaluation.

- Lastly, a novel system for fracture age estimation from pediatric wrist X-ray images was developed to provide more information about the fracture. This was the first-ever attempt at estimating the pediatric wrist fracture age, which means that conducted research serves as a standard for other researchers to follow. The complete system becomes more explainable and trustworthy by employing the Monte-Carlo dropout method. This way, the black-box status of the neural networks and neural networks-based system was overcome. Furthermore, due to different inputs being combined, the developed system can be called multimodal, making them a general-purpose approach that can be applied to other research fields.

All scientific contributions taken into account, this dissertation provides a significant step towards a decision support system for pediatric wrist fracture detection. All presented research: from X-ray images preprocessing with XAOM method, enhancing visibility of fractures by cast suppression, bone segmentation by local-entropy approach, to the final fracture detection and fracture age estimation – assembles a complete system. Therefore, this dissertation can serve as a basis for a CADx system that could help radiologists in detecting pediatric wrist fractures.

4.2 Future work

As a continuation of research conducted in this dissertation, possible expansions are as follows:

- As mentioned, all research done in this paper can be united in one CADx system that can be developed as an online platform. By making the developed system an internet platform, the whole medical community will benefit. Developing such a platform could be extended even further by building educational and training modules for radiologists. This also opens the possibility of enhancing models even more via online learning. In online learning, the models are constantly being trained with newly acquired data. Therefore, one of the niches for future work is creating an online platform based on the developed systems presented in the dissertation.

- Fracture detection based on the YOLOv4 model can be enhanced by merging it with a U-net model. Developing a novel method that will take the best of both worlds: object detection from YOLOv4 and pixel precision from the U-net model should be possible. Therefore, one of the future aims is to develop a model that will combine the two. To combine them together, it would be interesting to test novel self-attention mechanisms that are the main reason for Generative Pre-trained Transformer (GPT) rise.

- This dissertation mainly focused on X-ray image processing, neglecting each study's textual medical history. Extracting information from the medical history via natural language processing tools should increase the accuracy of currently developed systems even more. Therefore, merging currently developed models with the information from the medical history is worth investigating.

- An investigation of the correlation between cast suppression and radiation should be done. Namely, the hypothesis that developed cast suppression CycleGan model can enhance X-ray image quality to a similar level as the X-ray image quality taken with higher radiation dose must be tested. If the hypothesis is correct, it would be possible to lower the radiation dose when making X-ray images with a cast applied which is always beneficial.

- The next step would be expanding the developed fracture detection system to other body regions. In other words, developing a more robust system (data from different hospitals) and detecting fractures of more versatile body regions (other long bones would probably be a good start) is the ultimate goal. Developing such a CADx software presents a challenge from a machine learning point of view, but creating it could be a game-changer in radiology regarding reporting fractures by radiologists.

- The output of currently developed systems is a probability or a class concerning the context of a given task. However, generating a textual report on which radiologists could

agree or edit it would greatly reduce the necessary labor. Therefore, embedding the developed systems' outputs in a textual report – the one that radiologists usually generate – is also a goal to achieve.

Chapter 5

Summary of Papers

A   XAOM: A method for automatic alignment and orientation of radiographs for computer-aided medical diagnosis

Background and objectives: Computer-aided diagnosis relies on machine learning algorithms that require filtered and preprocessed data as the input. Aligning the image in the desired direction is an additional manual step in post-processing, commonly overlooked due to workload issues. Several state-of-the-art approaches for fracture detection and disease-struck region segmentation benefit from correctly oriented images, thus requiring such preprocessing of X-ray images. Furthermore, it is desirable to have archived studies in a standardized format. Radiograph hanging protocols also differ from case to case, which means that images are not always aligned and oriented correctly. As a solution, the paper proposes XAOM, an X-ray Alignment and Orientation Method for images from 21 different body regions.

Methods: Typically, other methods are crafted for this purpose to suit a specific body region and form of usage. In contrast, the method proposed in this paper is comprehensive and easily tuned to align and orient X-ray images of any body region. XAOM consists of two stages. For the first stage of the method, aligning X-ray images, we experimented with the following approaches: Hough transform, Fast line detection algorithm, and Principal Component Analysis method. For the second stage, we have experimented with the adaptations of several well known convolutional neural network topologies for correctly predicting image orientation: LeNet5, AlexNet, VGG16, VGG19, and ResNet50.

Results: In the first stage, the PCA-based approach performed best. The average difference between the angle detected by the algorithm and the angle marked by the experts on the test set containing 200 pediatric X-ray images was $1.65°$, while the median value was $0.11°$. In the second stage, the VGG16-based network topology achieved the best accuracy of 0.993 on a test set containing 4,221 images.

Conclusion: XAOM is highly accurate at aligning and orienting pediatric X-ray images of 21 common body regions according to a set standard. The proposed method is also robust and can

be easily adjusted to the different alignment and rotation criteria.

Availability: The Python source code of the best performing implementation of XAOM is publicly available at https://github.com/fhrzic/XAOM.

Hržić F., Tschauner S., Sorantin E., Štajduhar I. XAOM: A method for automatic alignment and orientation of radiographs for computer-aided medical diagnosis. Comput Biol Med. 2021 May;132:104300. doi: 10.1016/j.compbiomed.2021.104300. Epub 2021 Mar 3. PMID: 33714842.

B   Cast suppression in radiographs by generative adversarial networks

Injured extremities commonly need to be immobilized by casts to allow proper healing. We propose a method to suppress cast superimpositions in pediatric wrist radiographs based on the cycle generative adversarial network (CycleGAN) model. We retrospectively reviewed unpaired pediatric wrist radiographs $n = 9672$ and sampled them into 2 equal groups, with and without cast. The test subset consisted of 718 radiographs with cast. We evaluated different quadratic input sizes (256, 512, and 1024 pixels) for U-Net and ResNet-based CycleGAN architectures in cast suppression, quantitatively and qualitatively. The mean age was 11 ± 3 years in images containing cast $n = 4836$, and 11 ± 4 years in castless samples $n = 4836$. A total of 5956 X-rays had been done in males and 3716 in females. A U-Net 512 CycleGAN performed best $P \leq .001$. CycleGAN models successfully suppressed casts in pediatric wrist radiographs, allowing the development of a related software tool for radiology image viewers.

C   Local-Entropy Based Approach for X-Ray Image Segmentation and Fracture Detection

The paper proposes a segmentation and classification technique for fracture detection in X-ray images. This novel rotation-invariant method introduces the concept of local entropy for de-noising and removing tissue from the analysed X-ray images, followed by an improved procedure for image segmentation and the detection of regions of interest. The proposed local Shannon entropy was calculated for each image pixel using a sliding 2D window. An initial image segmentation was performed on the entropy representation of the original image. Next, a graph theory-based technique was implemented for the purpose of removing false bone contours and improving the edge detection of long bones. Finally, the paper introduces a classification and localisation procedure for fracture detection by tracking the difference between the extracted contour and the estimation of an ideal healthy one. The proposed hybrid method excels at detecting small fractures (which are hard to detect visually by a radiologist) in the ulna and radius bones—common injuries in children. Therefore, it is imperative that a radiologist inspecting the X-ray image receives a warning from the computerised X-ray analysis system, in order to prevent false-negative diagnoses. The proposed method was applied to a data-set containing 860 X-ray images of child radius and ulna bones (642 fracture-free images and 218 images containing fractures). The obtained results showed the efficiency and robustness of the proposed approach, in terms of segmentation quality and classification accuracy and precision (up to 91.16% and 86.22%, respectively).

D   Fracture Recognition in Pediatric Wrist Radiographs: An Object Detection Approach

Wrist fractures are commonly diagnosed using X-ray imaging, supplemented by magnetic resonance imaging and computed tomography when required. Radiologists can sometimes overlook the fractures because they are difficult to spot. In contrast, some fractures can be easily spotted and only slow down the radiologists because of the reporting systems. We propose a machine learning model based on the YOLOv4 method that can help solve these issues. The rigorous testing on three levels showed that the YOLOv4-based model obtained significantly better results in comparison to the state-of-the-art method based on the U-Net model. In the comparison against five radiologists, YOLO 512 Anchor model-AI (the best performing YOLOv4-based model) was significantly better than the four radiologists (AI $AUC - ROC = 0.965$, Radiologist average $AUC - ROC = 0.831 \pm 0.075$). Furthermore, we have shown that three out of five radiologists significantly improved their performance when aided by the AI model. Finally, we compared our work with other related work and discussed what to consider when building an ML-based predictive model for wrist fracture detection. All our findings are based on a complex dataset of 19,700 pediatric X-ray images.

E   Modeling Uncertainty in Fracture Age Estimation from Pediatric Wrist Radiographs

In clinical practice, fracture age estimation is commonly required, particularly in children with suspected non-accidental injuries. It is usually done by radiologically examining the injured body part and analyzing several indicators of fracture healing such as osteopenia, periosteal reaction, and fracture gap width. However, age-related changes in healing timeframes, inter-individual variabilities in bone density, and significant intra- and inter-operator subjectivity all limit the validity of these radiological clues. To address these issues, for the first time, we suggest an automated neural network-based system for determining the age of a pediatric wrist fracture. In this study, we propose and evaluate a deep learning approach for automatically estimating fracture age. Our dataset included 3570 medical cases with a skewed distribution toward initial consultations. Each medical case includes a lateral and anteroposterior projection of a wrist fracture, as well as patients' age, and gender. We propose a neural network-based system with Monte-Carlo dropout-based uncertainty estimation to address dataset skewness. Furthermore, this research examines how each component of the system contributes to the final forecast and provides an interpretation of different scenarios in system predictions in terms of their uncertainty. The examination of the proposed systems' components showed that the feature-fusion of all available data is necessary to obtain good results. Also, proposing uncertainty estimation in the system increased accuracy and F1-score to a final 0.906 ± 0.011 on a given task.

Bibliography

[1] Albawi, S., Mohammed, T. A., and Al-Zawi, S. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)* (2017), pp. 1–6.

[2] Bandyopadhyay, O., Chanda, B., and Bhattacharya, B. B. Entropy-Based Automatic Segmentation of Bones in Digital X-ray Images. In *PReMI* (2011).

[3] Beeran Kutty, S., Saaidin, S., Megat Yunus, P. N. A., and Abu Hassan, S. Evaluation of canny and sobel operator for logo edge detection. In *2014 International Symposium on Technology Management and Emerging Technologies* (2014), pp. 153–156.

[4] Blüthgen, C., Becker, A. S., de Martini, I. V., Meier, A., Martini, K., and Frauenfelder, T. Detection and localization of distal radius fractures: Deep learning system versus radiologists. *European journal of radiology 126* (2020), 108925.

[5] Bochkovskiy, A., Wang, C., and Liao, H. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *CoRR abs/2004.10934* (2020).

[6] Bottou, L. *Stochastic Gradient Descent Tricks*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 421–436.

[7] Cho, J., Lee, K., Shin, E., Choy, G., and Do, S. Medical Image Deep Learning with Hospital PACS Dataset. *CoRR abs/1511.06348* (2015).

[8] Cowen, A., Davies, A., and Kengyelics, S. Advances in computed radiography systems and their physical imaging characteristics. *Clinical Radiology 62*, 12 (2007), 1132–1141.

[9] Crawford, V. P. Learning the Optimal Strategy in a Zero-Sum Game. *Econometrica 42*, 5 (1974), 885–891.

[10] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 248–255.

[11] Dimitriadis, S., Liparas, D., and Tsolaki, M. N. Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate

among healhy elderly, MCI, cMCI and alzheimer's disease patients: From the alzheimer's disease neuroimaging initiative (ADNI) database. *Journal of Neuroscience Methods 302* (2018), 14–23. A machine learning neuroimaging challenge for automated diagnosis of Alzheimer's disease.

[12] Doi, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society 31*, 4-5 (2007), 198–211. 17349778[pmid].

[13] Duda, R. O., and Hart, P. E. Use of the Hough Transformation to Detect Lines and Curves in Pictures. *Commun. ACM 15*, 1 (jan 1972), 11–15.

[14] Dudley, R. M. Sample Functions of the Gaussian Process. *The Annals of Probability 1*, 1 (1973), 66 – 103.

[15] Eberly, D. Minimum-area rectangle containing a set of points. *Geometric Tools, LLC* (2015).

[16] Fushiki, T. Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing 21*, 2 (Apr 2011), 137–146.

[17] Gagunashvili, N. Chi-square tests for comparing weighted histograms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 614*, 2 (2010), 287–296.

[18] Gal, Y., and Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *Proceedings of The 33rd International Conference on Machine Learning* (New York, New York, USA, 20–22 Jun 2016), M. F. Balcan and K. Q. Weinberger, Eds., vol. 48 of *Proceedings of Machine Learning Research*, PMLR, pp. 1050–1059.

[19] Gan, K., Xu, D., Lin, Y., Shen, Y., Zhang, T., Hu, K., Zhou, K., Bi, M., Pan, L., Wu, W., et al. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta orthopaedica 90*, 4 (2019), 394–400. PMID: 30942136.

[20] Gan, K., Xu, D., Lin, Y., Shen, Y., Zhang, T., Hu, K., Zhou, K., Bi, M., Pan, L., Wu, W., and Liu, Y. Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. *Acta Orthopaedica 90*, 4 (2019), 394–400. PMID: 30942136.

[21] Geng, M., Meng, X., Yu, J., Zhu, L., Jin, L., Jiang, Z., Qiu, B., Li, H., Kong, H., Yuan, J., Yang, K., Shan, H., Han, H., Yang, Z., Ren, Q., and Lu, Y. Content-Noise Complementary

Learning for Medical Image Denoising. *IEEE Transactions on Medical Imaging 41*, 2 (2022), 407–419.

[22] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems* (2014), Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27, Curran Associates, Inc.

[23] Halabi, S. S., Prevedello, L. M., Kalpathy-Cramer, J., Mamonov, A. B., Bilbily, A., Cicero, M., Pan, I., Pereira, L. A., Sousa, R. T., Abdala, N., Kitamura, F. C., Thodberg, H. H., Chen, L., Shih, G., Andriole, K., Kohli, M. D., Erickson, B. J., and Flanders, A. E. The RSNA Pediatric Bone Age Machine Learning Challenge. *Radiology 290*, 2 (2019), 498–503. PMID: 30480490.

[24] Hao, J., Shen, Y., Xu, H., and Zou, J. A region entropy based objective evaluation method for image segmentation. In *2009 IEEE Instrumentation and Measurement Technology Conference* (2009), pp. 373–377.

[25] Hartigan, J. A., and Wong, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 28*, 1 (1979), 100–108.

[26] Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 1 (04 1970), 97–109.

[27] He, K., Zhang, X., Ren, S., and Sun, J. Deep Residual Learning for Image Recognition, 2015.

[28] Hedström, E. M., Svensson, O., Bergström, U., and Michno, P. Epidemiology of fractures in children and adolescents. *Acta Orthopaedica 81*, 1 (2010), 148–153. PMID: 20175744.

[29] Hržić, F., Janisch, M., Štajduhar, I., Lerga, J., Sorantin, E., and Tschauner, S. Modeling Uncertainty in Fracture Age Estimation from Pediatric Wrist Radiographs. *Mathematics 9*, 24 (2021).

[30] Hržić, F., Tschauner, S., Sorantin, E., and Štajduhar, I. XAOM: A method for automatic alignment and orientation of radiographs for computer-aided medical diagnosis. *Computers in Biology and Medicine 132* (2021), 104300.

[31] Hržić, F., Tschauner, S., Sorantin, E., and Štajduhar, I. Fracture Recognition in Paediatric Wrist Radiographs: An Object Detection Approach. *Mathematics 10*, 16 (2022).

[32] Hržić, F., Štajduhar, I., Tschauner, S., Sorantin, E., and Lerga, J. Local-Entropy Based Approach for X-Ray Image Segmentation and Fracture Detection. *Entropy 21*, 4 (2019).

[33] Hržić, F., Žužić, I., Tschauner, S., and Štajduhar, I. Cast suppression in radiographs by generative adversarial networks. *Journal of the American Medical Informatics Association 28*, 12 (10 2021), 2687–2694.

[34] Huang, Z., Zhang, Y., Li, Q., Zhang, T., and Sang, N. Spatially adaptive denoising for X-ray cardiovascular angiogram images. *Biomedical Signal Processing and Control 40* (2018), 131–139.

[35] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., and Ng, A. Y. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. *Proceedings of the AAAI Conference on Artificial Intelligence 33*, 01 (Jul. 2019), 590–597.

[36] Jia, W., Zhang, H., He, X., and Wu, Q. A comparison on histogram based image matching methods. In *2006 IEEE International Conference on Video and Signal Based Surveillance* (2006), IEEE, pp. 97–97.

[37] Joeris, A., Lutz, N., Wicki, B., Slongo, T., and Audigé, L. An epidemiological evaluation of pediatric long bone fractures — a retrospective cohort study of 2716 patients from two Swiss tertiary pediatric hospitals. *BMC Pediatrics 14*, 1 (Dec 2014), 314.

[38] Kaur, H., Pannu, H. S., and Malhi, A. K. A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions. *ACM Comput. Surv. 52*, 4 (aug 2019).

[39] Kavitha, M. S., Shanthini, J., and Sabitha, R. ECM-CSD: An Efficient Classification Model for Cancer Stage Diagnosis in CT Lung Images Using FCM and SVM Techniques. *Journal of Medical Systems 43*, 3 (Feb 2019), 73.

[40] Khairandish, M., Sharma, M., Jain, V., Chatterjee, J., and Jhanjhi, N. A Hybrid CNN-SVM Threshold Segmentation Approach for Tumor Detection and Classification of MRI Brain Images. *IRBM* (2021).

[41] Kim, D., and MacKinnon, T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clinical Radiology 73*, 5 (2018), 439–445.

[42] Kim, S., and Lee, W. Does McNemar's test compare the sensitivities and specificities of two diagnostic tests? *Statistical Methods in Medical Research 26*, 1 (2017), 142–154. PMID: 24996898.

[43] Kingma, D. P., and Ba, J. Adam: A Method for Stochastic Optimization, 2017.

[44] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM 60*, 6 (may 2017), 84–90.

[45] Kwon, G., Han, C., and Kim, D.-s. Generation of 3d brain mri using auto-encoding generative adversarial networks. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* (Cham, 2019), D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds., Springer International Publishing, pp. 118–126.

[46] Le Cam, L., LeCam, L. M., and Yang, G. L. *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media, 2000.

[47] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324.

[48] Lee, J. H., Kim, Y. J., and Kim, K. G. Bone age estimation using deep learning and hand X-ray images. *Biomedical Engineering Letters 10*, 3 (Aug 2020), 323–331.

[49] Lee, J. H., Lee, S., Zhang, G., Lim, J., Chung, W. K., and Suh, I. H. Outdoor place recognition in urban environments using straight lines. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (2014), pp. 5550–5557.

[50] Li, C., Yao, J., and Jiang, T. Retinal vessel segmentation network based on patch-gan. In *Intelligent Life System Modelling, Image Processing and Analysis* (Singapore, 2021), M. Fei, L. Chen, S. Ma, and X. Li, Eds., Springer Singapore, pp. 43–53.

[51] Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R., et al. Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences 115*, 45 (2018), 11591–11596.

[52] Liu, J., Zhao, G., Fei, Y., Zhang, M., Wang, Y., and Yu, Y. Align, Attend and Locate: Chest X-Ray Diagnosis via Contrast Induced Attention Network With Limited Supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2019).

[53] Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. Path Aggregation Network for Instance Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018).

[54] Lou, S., Jiang, X., and Scott, P. J. Applications of Morphological Operations in Surface Metrology and Dimensional Metrology. *Journal of Physics: Conference Series 483* (mar 2014), 012020.

[55] Luo, H., Hao, W., Foos, D., and Cornelius, C. Automatic image hanging protocol for chest radiographs in PACS. *IEEE Transactions on Information Technology in Biomedicine 10*, 2 (2006), 302–311.

[56] Ma, Y., Liu, J., Liu, Y., Fu, H., Hu, Y., Cheng, J., Qi, H., Wu, Y., Zhang, J., and Zhao, Y. Structure and Illumination Constrained GAN for Medical Image Enhancement. *IEEE Transactions on Medical Imaging 40*, 12 (2021), 3955–3967.

[57] Marín-Reyes, P. A., Lorenzo-Navarro, J., and Santana, M. C. Comparative study of histogram distance measures for re-identification. *CoRR abs/1611.08134* (2016).

[58] Mayerhoefer, M. E., Materka, A., Langs, G., Häggström, I., Szczypiński, P., Gibbs, P., and Cook, G. Introduction to Radiomics. *Journal of Nuclear Medicine 61*, 4 (2020), 488–495.

[59] Nagy, E., Janisch, M., Hržić, F., Sorantin, E., and Tschauner, S. A pediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning. *Scientific Data 9*, 1 (May 2022), 222.

[60] Neitzel, U. Status and prospects of digital detector technology for CR and DR. *Radiation Protection Dosimetry 114*, 1-3 (05 2005), 32–38.

[61] Nose, H., Unno, Y., Koike, M., and Shiraishi, J. A simple method for identifying image orientation of chest radiographs by use of the center of gravity of the image. *Radiological Physics and Technology 5*, 2 (Jul 2012), 207–212.

[62] Olczak, J., Fahlberg, N., Maki, A., Razavian, A. S., Jilert, A., Stark, A., Sköldenberg, O., and Gordon, M. Artificial intelligence for analyzing orthopedic trauma radiographs: deep learning algorithms—are they on par with humans for diagnosing fractures? *Acta orthopaedica 88*, 6 (2017), 581–586. PMID: 28681679.

[63] Raisuddin, A. M., Vaattovaara, E., Nevalainen, M., Nikki, M., Järvenpää, E., Makkonen, K., Pinola, P., Palsio, T., Niemensivu, A., Tervonen, O., et al. Critical evaluation of deep neural networks for wrist fracture detection. *Scientific reports 11*, 1 (2021), 6006.

[64] Redmon, J., and Farhadi, A. YOLOv3: An Incremental Improvement. *CoRR abs/1804.02767* (2018).

[65] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Cham, 2015), N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Springer International Publishing, pp. 234–241.

[66] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. ImageNet Large Scale Visual

Recognition Challenge. *International Journal of Computer Vision 115*, 3 (Dec 2015), 211–252.

[67] Santos, A. M., de Carvalho Filho, A. O., Silva, A. C., de Paiva, A. C., Nunes, R. A., and Gattass, M. Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM. *Engineering Applications of Artificial Intelligence 36* (2014), 27–39.

[68] Santos, M. K., Ferreira, J. R., Wada, D. T., Tenório, A. P. M., Barbosa, M. H. N., and Marques, P. M. d. A. Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine. *Radiologia brasileira 52* (2019), 387–396.

[69] Simonyan, K., and Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015.

[70] Tan, M., and Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning* (09–15 Jun 2019), K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97 of *Proceedings of Machine Learning Research*, PMLR, pp. 6105–6114.

[71] Thian, Y. L., Li, Y., Jagmohan, P., Sia, D., Chan, V. E. Y., and Tan, R. T. Convolutional neural networks for automated fracture detection and localization on wrist radiographs. *Radiology: Artificial Intelligence 1*, 1 (2019), e180001.

[72] Ur Rehman, H. Z., and Lee, S. Automatic Image Alignment Using Principal Component Analysis. *IEEE Access 6* (2018), 72063–72072.

[73] van Dokkum, P. G. Cosmic-Ray Rejection by Laplacian Edge Detection. *Publications of the Astronomical Society of the Pacific 113*, 789 (nov 2001), 1420–1427.

[74] Vidya Bhargavi, M., Mudunuru, V. R., and Veeramachaneni, S. Colon cancer stage classification using decision trees. In *Data Engineering and Communication Technology* (Singapore, 2020), K. S. Raju, R. Senkerik, S. P. Lanka, and V. Rajagopal, Eds., Springer Singapore, pp. 599–609.

[75] Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience 2018* (Feb 2018), 7068349.

[76] Wager, S., Wang, S., and Liang, P. S. Dropout Training as Adaptive Regularization. In *Advances in Neural Information Processing Systems* (2013), C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., vol. 26, Curran Associates, Inc.

[77] Wang, C.-W., and Chen, H.-C. Improved image alignment method in application to X-ray images and biological images. *Bioinformatics 29*, 15 (05 2013), 1879–1887.

[78] Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (June 2020).

[79] Wang, Z., Wu, F., and Hu, Z. MSLD: A robust descriptor for line matching. *Pattern Recognition 42*, 5 (2009), 941–953.

[80] Widek, T., Genet, P., Ehammer, T., Schwark, T., Urschler, M., and Scheurer, E. Bone age estimation with the Greulich-Pyle atlas using 3T MR images of hand and wrist. *Forensic Science International 319* (2021), 110654.

[81] Yahalomi, E., Chernofsky, M., and Werman, M. Detection of distal radius fractures trained by a small set of x-ray images and faster r-cnn. In *Intelligent Computing* (Cham, 2019), K. Arai, R. Bhatia, and S. Kapoor, Eds., Springer International Publishing, pp. 971–981.

[82] Yan, Z., Yang, X., and Cheng, K.-T. A Three-Stage Deep Learning Model for Accurate Retinal Vessel Segmentation. *IEEE Journal of Biomedical and Health Informatics 23*, 4 (2019), 1427–1436.

[83] You, C., Li, G., Zhang, Y., Zhang, X., Shan, H., Li, M., Ju, S., Zhao, Z., Zhang, Z., Cong, W., Vannier, M. W., Saha, P. K., Hoffman, E. A., and Wang, G. CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE). *IEEE Transactions on Medical Imaging 39*, 1 (2020), 188–203.

[84] Zhao, J., Li, D., Kassam, Z., Howey, J., Chong, J., Chen, B., and Li, S. Tripartite-GAN: Synthesizing liver contrast-enhanced MRI to improve tumor detection. *Medical Image Analysis 63* (2020), 101667.

[85] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (Oct 2017).

[86] Zimmerman, D. W. Teacher's Corner: A Note on Interpretation of the Paired-Samples t Test. *Journal of Educational and Behavioral Statistics 22*, 3 (1997), 349–360.

List of Figures

List of Tables

List of Publications

- Hržić F., Tschauner S., Sorantin E., Štajduhar I. XAOM: A method for automatic alignment and orientation of radiographs for computer-aided medical diagnosis. Comput Biol Med. 2021 May;132:104300. doi: 10.1016/j.compbiomed.2021.104300. Epub 2021 Mar 3. PMID: 33714842.

- Hržić, F., Žužić, I., Tschauner, S., and Štajduhar, I., Cast suppression in radiographs by generative adversarial networks. Journal of the American Medical Informatics Association 28, 12 (10 2021), 2687–2694. https://doi.org/10.1093/jamia/ocab192

- Hržić, F., Štajduhar, I., Tschauner, S., Sorantin, E., and Lerga, J., Local-Entropy Based Approach for X-Ray Image Segmentation and Fracture Detection. Entropy 2019, 21, 338. https://doi.org/10.3390/e21040338

- Hržić F., Janisch M., Štajduhar I., Lerga J., Sorantin E., Tschauner S., Modeling Uncertainty in Fracture Age Estimation from Pediatric Wrist Radiographs. Mathematics. 2021; 9(24):3227. https://doi.org/10.3390/math9243227

- Hržić F., Tschauner S., Sorantin E., Štajduhar I., Fracture Recognition in Paediatric Wrist Radiographs: An Object Detection Approach. Mathematics. 2022; 10(16):2939. https://doi.org/10.3390/math10162939

- Sorantin, E., Grasser, M.G., Hemmelmayr, A. et al. The augmented radiologist: artificial intelligence in the practice of radiology. Pediatr Radiol (2021). https://doi.org/10.1007/s00247-021-05177-7

- Lopac N., Hržić F., Vuksanović I. P. and Lerga J. , Detection of Non-Stationary GW Signals in High Noise From Cohen's Class of Time–Frequency Representations Using Deep Learning. in IEEE Access, vol. 10, pp. 2408-2428, 2022; doi: 10.1109/ACCESS.2021.3139850.

- Ljubic S., Hržić F., Salkanovic A., Štajduhar I. Augmenting Around-Device Interaction by Geomagnetic Field Built-in Sensor Utilization. Sensors. 2021; 21(9):3087. https://doi.org/10.3390/s21093087

51

- Nagy E., Janisch M., Hržić F. et al. A pediatric wrist trauma X-ray dataset (GRAZPEDWRI-DX) for machine learning. Sci Data 9, 222 (2022). https://doi.org/10.1038/s41597-022-01328-z

- Nagy, E., Marterer, R., Hržić, F., Sorantin, E., and Tschauner, S. (2022). Learning rate of students detecting and annotating pediatric wrist fractures in supervised artificial intelligence dataset preparations. Plos one, 17(10), e0276503.

- Hržić F., Jansky V., Sušanj D., Gulan G., Kožar G. and Jeričević D. Ž., Information entropy measures and clustering improve edge detection in medical X-ray images. 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 0164-0166, doi: 10.23919/MIPRO.2018.8400032.

- Dumenčić S., Tschauner S., Hržić F. and Štajduhar I., Automatic extraction of multiple-study X-ray images. 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), 2021, pp. 1-6, doi: 10.1109/INISTA52262.2021.9548551.

- Štajduhar, I. et al. (2021). Analysing Large Repositories of Medical Images. In: Rojas, I., Castillo-Secilla, D., Herrera, L.J., Pomares, H. (eds) Bioengineering and Biomedical Signal and Image Processing. BIOMESIP 2021. Lecture Notes in Computer Science(), vol 12940. Springer, Cham. https://doi.org/10.1007/978-3-030-88163-4_17

- Hržić F., Poljančić D. and Grbac T. G., Secure operations as congestion control mechanism within OpenStack based cloud laboratory. 2016 International Conference on Smart Systems and Technologies (SST), 2016, pp. 129-134, doi: 10.1109/SST.2016.7765646.

Curriculum Vitae

Franko Hržić was born in 1993 in Rijeka, Croatia. He finished elementary school in Malinska and high school in Rijeka. In 2015 at Faculty of Engineering University of Rijeka, he obtained a bachelor's degree in computer engineering, and in 2017 at the same university, he obtained a master's degree in computer engineering. In the same year, he enrolled a postgraduate university doctoral study in Electrical Engineering under the supervision of prof. dr. sc. Ivan Štajduhar. Later, when a postgraduate university doctoral study in Computer science was established, he transferred to it.

Since 2017, he has been an employee of the Faculty of Engineering University of Rijeka, where he works as a teaching assistant in the Department of computer engineering. He teaches courses Introduction to Programming and Introduction to Artificial Intelligence at an undergraduate, and Machine learning at graduate university study of computer engineering. Besides teaching, his scientific research focuses on developing and applying artificial intelligence in medicine. During his PhD, he established strong cooperation with the Medical University Of Graz, Austria, Division of Pediatric Radiology, Department of Radiology, where he cumulatively spent nearly half a year studying medical routines in patients' treatment. Since 2020 he is also a member of Center for Artificial Intelligence and Cybersecurity at University of Rijeka, where he participates in discussions and interdisciplinary research based on machine learning.

During his PhD he was a member and a researcher on several projects, including ERASMUS+ and bilateral projects with Slovenia. Since 2021, he has been a researcher on a project: "Machine Learning for Knowledge Transfer in Medical Radiology," financed by the Croatian Scientific Foundation. On the project, he is developing and testing machine learning models focusing on knowledge distillation and transfer learning. Besides this project, it is worthy to mention the "BLISS: Blended Learning Implementation for reSilient, acceSsible and efficient higher education" project where he is a cooperator.

His areas of scientific interest and research include but are not limited to machine learning, medical image analysis, explainable artificial intelligence, computer vision, and computer-aided diagnosis (CADx). Since 2018, he has published nine papers in international peer-reviewed scientific journals. He is the first author on five of those nine papers and the corresponding author on one. Furthermore, he was the first author of one scientific paper and co-author of another two scientific papers that have been reviewed and published in the proceedings of international scientific conferences.

Part II

Included Publications