Linköping University | Department of Computer and Information Science
Master's thesis, 30 ECTS | Statistics and Machine Learning
2026 | LIU-IDA/LITH-EX-A--26/001--SE

Image-to-image translation between fracture types in femur X-rays using generative adversarial networks and diffusion models

- \subtitleenglish

\titleswedish
- \subtitleswedish

Jakob Lindner

Supervisor : Anders Eklund
Examiner : Krzysztof Bartoszek

LINKÖPING UNIVERSITY

Linköpings universitet
SE–581 83 Linköping
+46 13 28 10 00 , www.liu.se

Abstract

The abstract resides in file `Abstract.tex`. Here you should write a short summary of your work.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque in massa suscipit, congue massa in, pharetra lacus. Donec nec felis tempor, suscipit metus molestie, consectetur orci. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Curabitur fermentum, augue non ullamcorper tempus, ex urna suscipit lorem, eu consectetur ligula orci quis ex. Phasellus imperdiet dolor at luctus tempor. Curabitur nisi enim, porta ut gravida nec, feugiat fermentum purus. Donec hendrerit justo metus. In ultrices malesuada erat id scelerisque. Sed sapien nisi, feugiat in ligula vitae, condimentum accumsan nisi. Nunc sit amet est leo. Quisque hendrerit, libero ut viverra aliquet, neque mi vestibulum mauris, a tincidunt nulla lacus vitae nunc. Cras eros ex, tincidunt ac porta et, vulputate ut lectus. Curabitur ultricies faucibus turpis, ac placerat sem sollicitudin at. Ut libero odio, eleifend in urna non, varius imperdiet diam. Aenean lacinia dapibus mauris. Sed posuere imperdiet ipsum a fermentum.

Nulla lobortis enim ac magna rhoncus, nec condimentum erat aliquam. Nullam laoreet interdum lacus, ac rutrum eros dictum vel. Cras lobortis egestas lectus, id varius turpis rhoncus et. Nam vitae auctor ligula, et fermentum turpis. Morbi neque tellus, dignissim a cursus sed, tempus eu sapien. Morbi volutpat convallis mauris, a euismod dui egestas sit amet. Nullam a volutpat mauris. Fusce sed ipsum lectus. In feugiat, velit eu fermentum efficitur, mi ex eleifend ante, eget scelerisque sem turpis nec augue.

Vestibulum posuere nibh ut iaculis semper. Ut diam justo, interdum quis felis ac, posuere fermentum ex. Fusce tincidunt vel nunc non semper. Sed ultrices suscipit dui, vel lacinia lorem euismod quis. Etiam pellentesque vitae sem eu bibendum. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Pellentesque scelerisque congue ullamcorper. Sed vehicula sodales velit a scelerisque. Pellentesque dignissim lectus ipsum, quis consectetur tellus rhoncus a.

Nunc placerat ut lectus vel ornare. Sed nec dictum enim. Donec imperdiet, ipsum ut facilisis blandit, lacus nisi maximus ex, sed semper nisl metus eget leo. Nunc efficitur risus ac risus placerat, vel ullamcorper felis interdum. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Duis vitae felis vel nibh sodales fringilla. Donec semper eleifend sem quis ornare. Proin et leo ut dolor consectetur vehicula. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Nunc dignissim interdum orci, sit amet pretium nibh consectetur sagittis. Aenean a eros id risus aliquam placerat nec ut lectus. Curabitur at quam in nisi sodales imperdiet in at erat. Praesent euismod pulvinar imperdiet. Nam auctor mattis nisi in efficitur. Quisque non cursus ipsum, consequat vehicula justo. Fusce varius metus et nulla rutrum scelerisque. Praesent molestie elementum nulla a consequat. In at facilisis nisi, convallis molestie sapien. Cras id ullamcorper purus. Sed at lectus sit amet dolor finibus suscipit vel et purus. Sed odio ipsum, dictum vel justo sit amet, interdum dictum justo. Quisque euismod quam magna, at dignissim eros varius in. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Acknowledgments

Acknowledgments.tex

Contents

Abstract **iii**

Acknowledgments **iv**

Contents **v**

1 Introduction **1**
1.1 Motivation 1
1.2 Research questions 2
1.3 Delimitations 2

2 Theory **3**

3 Method **4**

4 Results **5**

5 Discussion **6**
5.1 Results 6
5.2 Method 6
5.3 The work in a wider context 7

6 Conclusion **8**

Bibliography **9**

Todo list

or oversampling instead? 2
tbd 2

1 Introduction

1.1 Motivation

Atypical femoral fractures (AFFs) are a rare type of fractures in the femur. Unlike normal femoral fractures (NFFs), they are not caused by high immediate stress, but are related to long-term bisphosphonate medication prescribed to patients with osteoporosis [1]. The identification of AFFs remains challenging in clinical practice, and these fractures are frequently overlooked unless explicitly assessed. Less than 7% of all cases are correctly diagnosed in radiology reports [2]. In precision medicine and personalized healthcare, accurate diagnosis is crucial for effective treatment. AFFs have a high rate of reopertion [3] and therefore require a special focus. To avoid unecessary surgeries, reliable and early detection is essential.

To support medical doctors in improving diagnostic accuracy, deep neural networks have been trained to classify AFFs and NFFs [2]. However, the dataset used in previous studies was relatively small, consisting of approximately 1,000 images from around 375 patients. In this thesis, we utilize a larger dataset comprising approximately 4300 images from around 1200 patients. Nevertheless, this dataset presents a common challenge in medical imaging: severe class imbalance. The minority class (AFF) is significantly underrepresented, a well known issue in machine learning, that can negatively affect model performance [4]. Several strategies have been proposed to address class imbalance. Oversampling increases the number of minority class samples, often by duplicating existing images, while undersampling reduces the number of majority class samples. Although these techniques can improve performance, they also introduce potential drawbacks, such as overfitting in the case of oversampling and information loss in the case of undersampling.

With the rapid development of generative models, alternative approaches for dataset augmentation have emerged. Instead of reusing existing images, synthetic samples can be generated by learning the underlying high-dimensional data distribution.

Over the past decade, generative adversarial networks (GANs) have been the dominant architecture for generative tasks in medical imaging [5]. More recently, different versions of diffusion models showed remarkable preformance in various areas and have increasingly been applied to medical imaging tasks [6].

Generative models can be conditioned in different ways during the synthesis process. Previous thesis work used noise as input to generate synthetic images [7] in so-called noise-to-image processing. This thesis focuses on image-to-image translation, where the model

learns to translate images from one domain to another. This allows a more controlled synthesis conditioned on existing anatomical structures. A well established architecture for this task is CycleGAN [8]. We compare this approach with a diffusion-based model optimized for image-to-image translation, CycleDiff [9].

1.2 Research questions

The aim of this thesis is to establish whether synthesising data using generative models can improve the classification of rare fracture types. For this purpose, a CycleGAN and a CycleDiff model are trained on X-rays of AFFs and NFFs to create balanced datasets. The generative models are evaluated based on how well the generated images imitate the real data.

To assess whether generative augmentation improves fracture classification, a ResNet-50 classifier is trained under different conditions. First, a baseline model is trained using only the real data. Second, a classifier is trained using real data augmented with synthetically generated images. The performance of these models is then compared using standard metrics like accuracy and the F1-score and ,additionally, using statistical tests.

Furthermore, to determine whether generative augmentation provides an advantage over conventional data augmentation techniques (such as rotation, flipping, and cropping), a ResNet-50 trained on real data augmented with standard methods is compared to a ResNet-50 trained on real data augmented with generative models.

1. Does CycleGAN or CycleDiff perform better in the task of generating synthetic images by transforming between AFF and NFF images?
2. Does a ResNet-50 classifier trained on real data augmented with generated images perform significantly better than a classifier trained only on real data?
3. Does generative augmentation lead to significantly better classification performance compared to the standard augmentation techniques rotation, flipping and cropping ?

or oversampling instead?

1.3 Delimitations

tbd

This is where the main delimitations are described. For example, this could be that one has focused the study on a specific application domain or target user group. In the normal case, the delimitations need not be justified.

2 Theory

The main purpose of this chapter is to make it obvious for the reader that the report authors have made an effort to read up on related research and other information of relevance for the research questions. It is a question of trust. Can I as a reader rely on what the authors are saying? If it is obvious that the authors know the topic area well and clearly present their lessons learned, it raises the perceived quality of the entire report.

After having read the theory chapter it shall be obvious for the reader that the research questions are both well formulated and relevant.

The chapter must contain theory of use for the intended study, both in terms of technique and method. If a final thesis project is about the development of a new search engine for a certain application domain, the theory must bring up related work on search algorithms and related techniques, but also methods for evaluating search engines, including performance measures such as precision, accuracy and recall.

The chapter shall be structured thematically, not per author. A good approach to making a review of scientific literature is to use *Google Scholar* (which also has the useful function *Cite*). By iterating between searching for articles and reading abstracts to find new terms to guide further searches, it is fairly straight forward to locate good and relevant information, such as .

Having found a relevant article one can use the function for viewing other articles that have cited this particular article, and also go through the article's own reference list. Among these articles on can often find other interesting articles and thus proceed further.

It can also be a good idea to consider which sources seem most relevant for the problem area at hand. Are there any special conference or journal that often occurs one can search in more detail in lists of published articles from these venues in particular. One can also search for the web sites of important authors and investigate what they have published in general.

This chapter is called either *Theory*, *Related Work*, or *Related Research*. Check with your supervisor.

3 Method

In this chapter, the method is described in a way which shows how the work was actually carried out. The description must be precise and well thought through. Consider the scientific term replicability. Replicability means that someone reading a scientific report should be able to follow the method description and then carry out the same study and check whether the results obtained are similar. Achieving replicability is not always relevant, but precision and clarity is.

Sometimes the work is separated into different parts, e.g. pre-study, implementation and evaluation. In such cases it is recommended that the method chapter is structured accordingly with suitable named sub-headings.

4 Results

This chapter presents the results. Note that the results are presented factually, striving for objectivity as far as possible. The results shall not be analyzed, discussed or evaluated. This is left for the discussion chapter.

In case the method chapter has been divided into subheadings such as pre-study, implementation and evaluation, the result chapter should have the same sub-headings. This gives a clear structure and makes the chapter easier to write.

In case results are presented from a process (e.g. an implementation process), the main decisions made during the process must be clearly presented and justified. Normally, alternative attempts, etc, have already been described in the theory chapter, making it possible to refer to it as part of the justification.

5 Discussion

This chapter contains the following sub-headings.

5.1 Results

Are there anything in the results that stand out and need be analyzed and commented on? How do the results relate to the material covered in the theory chapter? What does the theory imply about the meaning of the results? For example, what does it mean that a certain system got a certain numeric value in a usability evaluation; how good or bad is it? Is there something in the results that is unexpected based on the literature review, or is everything as one would theoretically expect?

5.2 Method

This is where the applied method is discussed and criticized. Taking a self-critical stance to the method used is an important part of the scientific approach.

A study is rarely perfect. There are almost always things one could have done differently if the study could be repeated or with extra resources. Go through the most important limitations with your method and discuss potential consequences for the results. Connect back to the method theory presented in the theory chapter. Refer explicitly to relevant sources.

The discussion shall also demonstrate an awareness of methodological concepts such as replicability, reliability, and validity. The concept of replicability has already been discussed in the Method chapter (3). Reliability is a term for whether one can expect to get the same results if a study is repeated with the same method. A study with a high degree of reliability has a large probability of leading to similar results if repeated. The concept of validity is, somewhat simplified, concerned with whether a performed measurement actually measures what one thinks is being measured. A study with a high degree of validity thus has a high level of credibility. A discussion of these concepts must be transferred to the actual context of the study.

The method discussion shall also contain a paragraph of source criticism. This is where the authors' point of view on the use and selection of sources is described.

In certain contexts it may be the case that the most relevant information for the study is not to be found in scientific literature but rather with individual software developers and open

source projects. It must then be clearly stated that efforts have been made to gain access to this information, e.g. by direct communication with developers and/or through discussion forums, etc. Efforts must also be made to indicate the lack of relevant research literature. The precise manner of such investigations must be clearly specified in a method section. The paragraph on source criticism must critically discuss these approaches.

Usually however, there are always relevant related research. If not about the actual research questions, there is certainly important information about the domain under study.

5.3 The work in a wider context

There must be a section discussing ethical and societal aspects related to the work. This is important for the authors to demonstrate a professional maturity and also for achieving the education goals. If the work, for some reason, completely lacks a connection to ethical or societal aspects this must be explicitly stated and justified in the section Delimitations in the introduction chapter.

In the discussion chapter, one must explicitly refer to sources relevant to the discussion.

6 Conclusion

This chapter contains a summarization of the purpose and the research questions. To what extent has the aim been achieved, and what are the answers to the research questions?

The consequences for the target audience (and possibly for researchers and practitioners) must also be described. There should be a section on future work where ideas for continued work are described. If the conclusion chapter contains such a section, the ideas described therein must be concrete and well thought through.

Bibliography

[1] Jörg Schilcher, Karl Michaëlsson, and Per Aspenberg. "Bisphosphonate Use and Atypical Fractures of the Femoral Shaft". In: *New England Journal of Medicine* 364.18 (May 2011). _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMoa1010650, pp. 1728–1737. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1010650. URL: https://www.nejm.org/doi/full/10.1056/NEJMoa1010650 (visited on 02/14/2026).

[2] Georg Zdolsek, Yupei Chen, Hans-Peter Bögl, Chunliang Wang, Mischa Woisetschläger, and Jörg Schilcher. "Deep neural networks with promising diagnostic accuracy for the classification of atypical femoral fractures". In: *Acta Orthopaedica* 92.4 (July 2021). _eprint: https://doi.org/10.1080/17453674.2021.1891512, pp. 394–400. ISSN: 1745-3674. DOI: 10.1080/17453674.2021.1891512. URL: https://doi.org/10.1080/17453674.2021.1891512 (visited on 01/20/2026).

[3] H.P. Bögl, K. Michaëlsson, G. Zdolsek, J. Höijer, and J. Schilcher. "Increased rate of reoperation in atypical femoral fractures is related to patient characteristics and not fracture type. A nationwide cohort study". en. In: *Osteoporosis International* 31.5 (May 2020), pp. 951–959. ISSN: 1433-2965. DOI: 10.1007/s00198-019-05249-3. URL: https://doi.org/10.1007/s00198-019-05249-3 (visited on 02/16/2026).

[4] Arman Ghavidel and Pilar Pazos. "Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review". en. In: *Journal of Cancer Survivorship* 19.1 (Feb. 2025), pp. 270–294. ISSN: 1932-2267. DOI: 10.1007/s11764-023-01465-3. URL: https://doi.org/10.1007/s11764-023-01465-3 (visited on 02/16/2026).

[5] Mohd Ali, Mehboob Ali, Mubashir Hussain, and Deepika Koundal. "Generative Adversarial Networks (GANs) for Medical Image Processing: Recent Advancements". en. In: *Archives of Computational Methods in Engineering* 32.2 (Mar. 2025), pp. 1185–1198. ISSN: 1886-1784. DOI: 10.1007/s11831-024-10174-8. URL: https://doi.org/10.1007/s11831-024-10174-8 (visited on 01/21/2026).

[6] Yaqing Shi, Abudukelimu Abulizi, Hao Wang, Ke Feng, Nihemaiti Abudukelimu, Youli Su, and Halidanmu Abudukelimu. "Diffusion Models for Medical Image Computing: A Survey". In: *Tsinghua Science and Technology* 30.1 (Feb. 2025), pp. 357–383. ISSN: 1007-0214. DOI: 10.26599/TST.2024.9010047. URL: https://ieeexplore.ieee.org/abstract/document/10676408 (visited on 01/20/2026).

[7] Yaning Wang. *Synthetic X-ray Images of Femur Fractures*. eng. 2025. URL: https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-218223 (visited on 01/21/2026).

[8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. arXiv:1703.10593 [cs] version: 6. Nov. 2018. DOI: 10.48550/arXiv.1703.10593. URL: http://arxiv.org/abs/1703.10593 (visited on 01/20/2026).

[9] Shilong Zou, Yuhang Huang, Renjiao Yi, Chenyang Zhu, and Kai Xu. *CycleDiff: Cycle Diffusion Models for Unpaired Image-to-image Translation*. en. arXiv:2508.06625 [cs]. Jan. 2026. DOI: 10.48550/arXiv.2508.06625. URL: http://arxiv.org/abs/2508.06625 (visited on 02/11/2026).