Few-shot learning for (medical) image analysis: An overview of recent techniques with a focus on classification

Seminar Software Engineering
Winter Semester 2023/2024

Philipp Treutlein

University of Würzburg, Germany
philipp.treutlein@stud-mail.uni-wuerzburg.de

Abstract. Diseases are often diagnosed through a wide array of imaging techniques. Normally, trained medical personnel is needed for the analysis of such medical images. On the other hand, deep learning approaches show great results in image classification tasks.
However, these techniques are often immensely data-dependent, while large medical datasets are often sparse. Few-shot learning approaches try to reduce this reliance on large datasets while maintaining accuracy.
This paper investigates different few-shot learning techniques and examines the results of recent models based on those. *TODO: Ergebnis*

1 Introduction

Imaging procedures are widely used for medical diagnosis today. X-ray scans, MRI, CT-scans and many more procedures are used to create special images of the human body, which are indispensable for detecting the cause of diseases.
On the other hand, image classification is one of the most well-known (and highly researched) machine-learning tasks.
Naturally, using machine-learning for the classification of medical images would be an obvious connection. Neural networks could be utilized to aid physicians in this regard by quickly giving a first assessment. It also might detect a very rare disease the doctor has never seen before and is therefore unable to diagnose correctly. Therefore it might not only save time, but also increase the accuracy of diagnosis.
The problem here lies in the high data-dependency of common ML approaches. Image classification only works well in practice if the model is trained on huge amounts of labelled data. But some diseases are very rare and/or the labelling of large amounts of data is time-consuming and expensive. For this reasons, the needed datasets for many medical tasks simply do not exist.
To battle this problem, few-shot learning could be utilized. Few-shot learning (FSL) is the generic term for different techniques which aim to provide well-performing models with a limited amount of training data.
This paper aims to provide an overview over FSL in image classification, especially in the medical domain. The different techniques of FSL are categorized and examples for them are discussed.

2 Few-shot learning in general

As briefly mentioned in the introduction, few-shot learning (FSL) is the general term for a group of techniques, which train a neural network on a small data-set contrary to "classical" approaches which use large data-sets with possibly hundreds of thousands data points.

FSL can be used for different tasks, like image segmentation, face recognition (especially for "face unlock" features)[4], handwriting deciphering[4], and many more, but this paper focuses on image classification, especially for medical analysis.
In image classification, FSL could be applied in different ways, either to train a network from scratch for a specific task or to add new classes to existing models.
The latter can be quite difficult in traditional approaches, since the classes to add also need a large amount of data samples to not cause class imbalance.[23] or might be hard to come by. Moreover, it also takes a substantial amount of time to train the model with the new data.
Few-shot learning could also be employed in such a case to reduce the data-dependency.
When talking about few-shot learning in classification tasks, a common notation for the general problem is "k-shot n-way", meaning that the model is trained on k samples per class and tries to sort the new sample into one of n classes[8][18].
Some common approaches for FSL, which this paper will give an overview over, are data augmentation, prototypical networks and meta-learning.

2.1 Data augmentation

The basic idea in data augmentation is the use of different techniques to generate new training samples, this can either be done by hand or with another neural network. These can then be used for training a NN for the actual task. The prerequisite is a training set of very high quality. Furthermore, the sample needs the ability to generalize well over other members of its corresponding class, meaning its key features need to be a good representation of its corresponding class.[8]
Data augmentation is often done in combination with other techniques, not just as a standalone strategy.
The reason for this is the simplicity of combining both, since data augmentation does not have any influence on the specific model employed, but only on the training data for it.
Also, data augmentation can reduce overfitting[11] and is therefore often used for different strategies as a pre-processing step on the training dataset.
In the most simple case, data augmentation is done by applying transformations like rotation[8][11] or flipping[11].
The idea with these methods often is the simulation of images of worse quality, meaning partially occluded, taken from weird angles, not correctly focused etc..
In the following, a short overview over some common techniques for data augmentation is given.

Rotation Like already mentioned, rotating is an easy data augmentation method. It could be done either by a 90° angle or only by a slight angle. Contrary to the first case, the latter automatically introduces some noise[11], since some parts of the image will "rotate out of the view" (which means some kind of cropping2.1 is also applied) or the image has to be resized. In both cases, some of the background will be shown instead (like seen in 1).

Cropping Cropping means cutting out some parts from the edges of the image. Since the neural network needs the input in a specified size, the cropped result is then scaled up to match the right dimensions. This is often done randomly[11].

Gaussian blur But more complex methods for data augmentation also exist, for example "Gaussian blur" (example see Fig. 2), which uses the Gaussian distribution function to change the weight of each pixel by factoring in the weights of the surrounding pixels. This technique mimics images where the target is out of focus.[17]

Fig. 1. Example for rotation by Song et al.[17]

Fig. 2. Example for Gaussian blur by Apple Inc.[1]

The techniques presented here are just general examples, which might not be sensible for all medical imaging procedures.

2.2 Meta-learning

Another often used approach is meta-learning, which means "learning to learn"[24][8].
This approach tries to mimic the learning-abilities of the human brain, which is way less data-dependent than neural networks, especially when it comes to classification tasks. For instance, Children only need a few examples each to be able to distinguish between different kinds of animals[8], while a deep neural network requires data of several orders of magnitude more to reach a comparable accuracy.
The neural network is trained on one (or more) tasks and then given a different task with only a limited amount of samples. When successful, the model should be able to "learn" the new task much easier than the old one it was trained on.
...
Meta-learning in general has the problem, that it often does not work well if there is a significant domain shift between the tasks[6]. For instance, a model trained on images of plants might not be able to generalize well and thus show poor results when classifying animals (using a small number of examples).

Internally, meta-learning often means training a network to extract key features of the images, which can then be used for measuring the similarity between a sample and each of the classes. This can be seen in Fig.3.
In general, good image embeddings, that capture these key features well, are more important for this technique than highly optimizing the meta-learning algorithm itself[20].

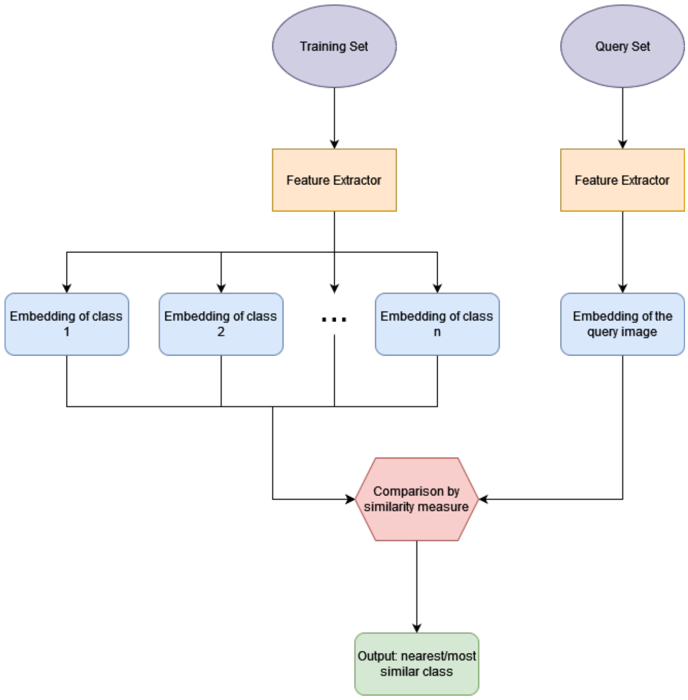

Fig. 3. basic principle of Meta learning: extraction of class features and classification by similarity

Nevertheless, meta-learners might also take the model parameters and configurations that performed well in the training task into consideration.[21]

Model-agnostic Meta Learning (MAML) ... MAML, or Model-agnostic Meta Learning is a special meta-learning algorithm, which aims to generalize well and provide the ability to be easily applied to any task or model by fine-tuning.[5]
To reach this goal, the algorithm has to obtain only transferrable representations. This is done by finding parameters, that are susceptible to task changes, which means that altering those improves the target task by a large margin.[5]
MAML does not make many assumptions about the employed model, it is only necessary to ensure that parameterization is done by some parameter vector and gradient learning can be applied on it.[5]
Furthermore, parameter updates when learning are done by gradient descent and the model is optimized to need as few of those steps as possible for a new task. [5]

2.3 Prototypical Networks

A third approach for few-shot learning employs prototypical networks. This kind of network is only fed positive samples for each class. Next, the network creates a prototype for this class from the most important features found in the samples. This prototype serves as a baseline for further classification tasks.[8][16]
Internally, this is handled by embeddings, which are low dimensional representations of the images. These embeddings are created by a neural network, which maps the input in a specific embedding space[16]. Each image is described by a single vector. These vectors are then combined into the

embedding of the whole class by calculating the centroid of the instances.[8]
New classification examples are then compared to these trained class prototypes and classified

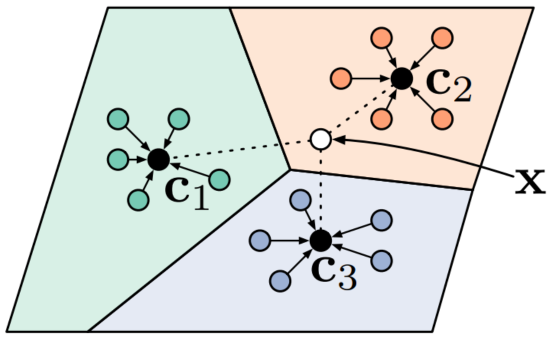

Fig. 4. Representation of a prototypical network by Snell et al.[16]
c_1, c_2, c_3 are the learned class prototypes computed by combining the embeddings of the colored dots around them. X is a new instance to classify, which is assigned to c_2, since it is the nearest.

by some similarity measure[8] such as cosine similarity or (squared) euclidean distance, which outperforms the former according to Snell et al.[16].

3 Categorization and state-of-the-art

Since some of the different approaches to few-shot-learning are now clear, an overview over a few selected examples of the state-of-the-art for those methods in the medical domain will be given.

3.1 Data augmentation

Electrocardiography/Electrocardiograms (ECGs) Electrocardiography is one of the most common medical imaging techniques utilized for the detection of cardiac diseases. A so called "electrocardiagram" (or short ECG) is produced, which shows the hearts activity over multiple cycles.
There have been multiple works dedicated to applying few-shot learning onto this task to identify these ECGs automatically and therefore save time and personal resources.
Li et al.[9] proposed a one-dimensional siamese network (for the base concepts of siamese networks see 3.2) to reduce the data-dependency in comparison to other successful models.
The network is trained with a small labeled dataset consisting of N samples for each of the four categories. Correspondingly, the value of N was set to four, 20, 40, 120 and 200 respectively in different tests. Moreover, training input is always given in pairs, half of them from the same class, half from different classes. Further, feature vectors are extracted from these inputs, the (manhatten) distance between them is calculated and the results assembled into a matrix, from which the probability of the two images being from the same class can be computed. [9]
This input pair approach helps to counter the small dataset. Since many different pairs can be generated, even for small Ns, the amount of training data is increased substantially. This is enough to prevent overfitting.[9]
Each ECG signal is divided into the heartbeats and shortened to include 250 sample points.[9]
Since the selection of the training set has a big impact on the accuracy of the model, Li et al.[9]

ran the experiment multiple times with the training set being randomized and therefore different every time. Furthermore, the number of shots was set to different values between one and 50, with accuracies ranging from 82.36% to 92.42%. Obviously, adding more shots always enhanced the result.[9]
The approach can be seen as some form of data augmentation, since additional training data is generated from the small dataset. Still, the classification of the used technique is not clear-cut, since the feature extraction and distance measurement is very similar to meta-learning as seen in 2.2. But like already stated in 2.1, data augmentation is more of a supporting technique than a fully fledged method.
Nevertheless, this work shows, that in some cases, even relatively simple techniques can achieve great results in a few-shot setting. But it also has to be noted, that ECGs are relatively simple compared to images produced by other medical imaging techniques, which might result in lower accuracy for other tasks.

3.2 Meta-learning

Electrocardiography/Electrocardiograms (ECGs) Unlike Li et al.'s[9] data augmentation approach, Liu et al.[10] proposed a meta-learning based method for classification of ECG signals for arrhytmia detection, especially in regard to data created by wearable devices. The computing capabilities of these devices, which often employ edge computing, limit the model to smaller datasets.[10]
For their method, Liu et al.[10] present a special pre-processing strategy which converts the ECG into so called "time-frequency spectrograms". These spectrograms are then utilized as the representation fed into the classifying 2D-CNN.[10]
Instead of using the Euclidean distance, the Mahalanobis distance was employed as a similarity measure, since it is much better suited for measuring the distance between a data point and a distribution[14] representing a class.
Further, the feature extractor, based on the EfficientNet network, is trained on a resampled variant of the PTB dataset with 10000 entries. Also, 3000 samples are used for validation and 1000 for testing. The EfficientNet[19] approach aids with reducing the amount of parameters, which according to the authors helps for few-shot learning. [10]
Moreover, the network was fine-tuned with randomly selected 2-way/4-way classification tasks from the dataset with k samples for k-shot learning.[10]
For evaluation, a two-way (healthy and unhealthy) and a four-way (which type of arrhythmia) classification were tested with a k value of 1, 5 and 10 shots each. Also, five, ten or 15 records were merged into a single spectrogram to represent different sampling periods. The samples for this testing were taken from the MIT-BIH dataset.[10]
The experiment showed mean accuracies from 93.3-98.5% for the two-way classification task and 90.9-97.0% for the four-way classification with a confidence interval of 95%.

Looking at the result, this approach seems to be even better than Liu et al's[10] (seen in 3.1), when purely going off the accuracy measures. However, Liu et al's[10] model did not leverage such a big dataset for training the feature extractor...
TODO: Kommentar/Bewertung/Vergleich erweitern

Disease detection in plant leaves ... Argüeso et al.[2] also used a meta-learning approach for disease-detection in plant leaves. In their work, they proposed using Siamese networks for few-shot learning. The goal was adding new classes with only a few images.
The model was first trained on a large data-set of 54,303 images for 38 classes of leaf/disease types. After learning image embeddings on 32 of those classes, the fully connected layers of the neural

network were replaced by a SMV classifier. Further, it was retrained on only a few samples (one to 140 images per class, to test the scaling of more training data) of the other six classes and then tasked to classify images sampled from them.[2]

A classical Siamese networks consist of two identical subnets, meaning they have the exact same weights, which are each given an input simultaneously[3][2]. The network then tries to minimize a loss function for the given input pair. The (Contrastive) loss function is designed to minimize distances between similar classes, while also maximizing the distances between unrelated ones.[2]

Argüeso et al.[2] used such networks for learning the image embeddings.
In addition to the two subnet version, a three subnet Siamese network was also employed. In that version, a so called "anchor" is fed into the network, as well as a positive and a negative sample. The anchor is the baseline of the comparison, the positive image has the same class, while the negative image is of another class. The used loss function is a Triplet loss function, which minimizes the distance between positive image and anchor while maximizing the distance between anchor and negative image. [2]
Argüeso et al.'s approach yielded promising results in their testing, with the Triplet loss version reaching accuracies of 80% when using 15 images per class and 90% for 80 respectively.[2]
When looking at these results, it is clear that scaling the data-set with new diseases works well with their method. But this test only showed meta-learning transferred between very similar domains. To be more precise, only new classes of the same domain were added with few-shot learning. Which means, it was still needed for the networks to train on a large data-set of similar classes to be able to learn new ones with only a few shots.
In a medical context, this might not be a problem for some imaging techniques like chest x-rays, for which large annotated data-sets exist[12], but could be problematic for others.
Still, this learning approach has great potential when trying to detect very rare conditions, since it would be possible to learn the network on common cases (under the premise a corresponding data-set exists) first and then use this few-shot learning method to train for the uncommon ones. If that approach also works for cross-domain application is unclear from the paper and could be an interesting path to explore in future work.

Human cell analysis Another medical classification task with the need for few-shot learning due to the lack of large annotated data-sets is the analysis of human cells. Due to this fact, cross-domain meta learning approaches were explored in the past.
Relatively recently, Walsh et al.[23] evaluated state-of-the-art techniques for this use-case and tried to optimize the best performing ones by using different backbone architectures, data augmentation methodologies and training schemes. Still, they found the accuracy of even the best, optimized models to be quite lackluster when it comes to transferring from a non-medical data-set to a medical one.[23]
For their experiments, Walsh et al.[23] firstly trained nine different techniques on the mini-ImageNet[22] dataset. Secondly the models were tested on a medical dataset and the performance of each of them was compared.[23]
Moreover, a 5-way 5-shot strategy was employed in testing, as well as training. Also, different data-augmentation (see 2.1 for the basic idea) methods were applied to the training data to prevent overfitting.[23]
The results show a loss of accuracy of at least 30% when switching domain to the medical dataset. Overall, the Reptile[13] model, which is based on MAML, achieved the best results, reaching the highest accuracy on the BCCD dataset and performing well on HEp-2.[23]
TODO: Kommentar/Bewertung

3.3 Prototypical Networks

Dermatological disease diagnosis Like with many other types of diseases, many dermatological conditions[1] are quite rare. Additionally, many common ones are seldom recorded[15]. Hence, corresponding data-sets are following a long tailed distribution[15], meaning there are only few data-points for these conditions.
For this reason, Prabhu et al.[15] examined a prototypical network based method for few-shot learning to counteract this data limitation. Further, they presented an extension of this approach named "Prototypical Clustering Networks" (PCN) to accurately model the high intra-class variability often found in dermatological diseases due to the differences between affected areas.[15]
Compared to typical prototypical networks, the PCNs represent classes by more than a single prototype. Accordingly, the distance measures use a weighted combination of these sets of prototypes instead of a single class embedding.[15]
Using this addition to their model consistently yielded more accurate classifications than the comparison PN. In their testing, Prabhu et al[15] trained the model with all their training data for the 150 largest classes, which they called "base classes" and only five or ten examples for the other 50 ("novel") classes. When classifying five test examples of these novel classes, the PCN reached mean per-class accuracies of around 30 for five training shots and around 50 for ten.[15]
Important to note is the comparison value for the base classes only, which was around 51[15]. This means that the model with only ten examples was almost as accurate for novel classes as it was for base classes.
...
TODO: vllt noch Zhang Paper[25], da nochmal neuer (2023) + verwendet beides, meta-learning und prototypical network, vllt noch ne subsection mit "Hybrids" o.Ä?

4 Conclusions

TODO: ausformulierung, ergänzung

- meta learning: cross-domain is often still problematic, but reasearch is done to fix that problem, one possible method is feature augmentation[7]; transfer between similar domains works pretty well
- comparing between results is hard, different datasets and methodologies are used; not many papers testing multiple recent approaches under same conditions
- many different approaches can show great results, but often only after training on large, similar datasets to learn feature extraction
- ...

[1] Conditions affecting the skin

References

1. Apple Inc.: Gaussian blur filter in motion, `https://support.apple.com/en-ge/guide/motion/motn169fe6e6/mac`, accessed on 03.12.2023
2. Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M.G., Bereciartua, A., Alvarez-Gila, A.: Few-shot learning approach for plant disease classification using images taken in the field. Computers and Electronics in Agriculture 175, 105542 (2020)
3. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a" siamese" time delay neural network. Advances in neural information processing systems 6 (1993)
4. Dilmegan, C.: What is few-shot learning? methods & applications in 2023, `https://research.aimultiple.com/few-shot-learning/`, accessed on 08.12.2023
5. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. pp. 1126–1135. PMLR (2017)
6. Hu, S., Tomczak, J., Welling, M.: Meta-learning for medical image classification (2018)
7. Hu, Y., Ma, A.J.: Adversarial feature augmentation for cross-domain few-shot classification. In: European Conference on Computer Vision. pp. 20–37. Springer (2022)
8. Kotia, J., Kotwal, A., Bharti, R., Mangrulkar, R.: Few shot learning for medical imaging. Machine learning algorithms for industrial applications pp. 107–132 (2021)
9. Li, Z., Wang, H., Liu, X.: A one-dimensional siamese few-shot learning approach for ecg classification under limited data. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 455–458. IEEE (2021)
10. Liu, T., Yang, Y., Fan, W., Wu, C.: Few-shot learning for cardiac arrhythmia detection based on electrocardiogram data from wearable devices. Digital Signal Processing 116, 103094 (2021)
11. Maharana, K., Mondal, S., Nemade, B.: A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings 3(1), 91–99 (2022)
12. Nguyen, H.Q., Lam, K., Le, L.T., Pham, H.H., Tran, D.Q., Nguyen, D.B., Le, D.D., Pham, C.M., Tong, H.T., Dinh, D.H., et al.: Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. Scientific Data 9(1), 429 (2022)
13. Nichol, A., Achiam, J., Schulman, J.: On first-order meta-learning algorithms. arXiv preprint arXiv:1803.02999 (2018)
14. Prabhakaran, S.: Mahalanobis distance – understanding the math with examples (python), `https://www.machinelearningplus.com/statistics/mahalanobis-distance/`, accessed on 08.12.2023
15. Prabhu, V., Kannan, A., Ravuri, M., Chaplain, M., Sontag, D., Amatriain, X.: Few-shot learning for dermatological disease diagnosis. In: Machine Learning for Healthcare Conference. pp. 532–552. PMLR (2019)
16. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in neural information processing systems 30 (2017)
17. Song, C., Xu, W., Wang, Z., Yu, S., Zeng, P., Ju, Z.: Analysis on the impact of data augmentation on target recognition for uav-based transmission line inspection. Complexity 2020, 1–11 (2020)
18. Song, Y., Wang, T., Cai, P., Mondal, S.K., Sahoo, J.P.: A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. ACM Computing Surveys (2023)
19. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
20. Tian, Y., Wang, Y., Krishnan, D., Tenenbaum, J.B., Isola, P.: Rethinking few-shot image classification: a good embedding is all you need? In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16. pp. 266–282. Springer (2020)
21. Vanschoren, J.: Meta-learning. Automated machine learning: methods, systems, challenges pp. 35–61 (2019)
22. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. Advances in neural information processing systems 29 (2016)
23. Walsh, R., Abdelpakey, M.H., Shehata, M.S., Mohamed, M.M.: Automated human cell classification in sparse datasets using few-shot learning. Scientific Reports 12(1), 2924 (2022)
24. Wang, J.X.: Meta-learning in natural and artificial intelligence. Current Opinion in Behavioral Sciences 38, 90–95 (2021)

25. Zhang, B., Gao, B., Liang, S., Li, X., Wang, H.: A classification algorithm based on improved meta learning and transfer learning for few-shot medical images. IET Image Processing 17(12), 3589–3598 (2023), https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ipr2.12889, accessed on 12.12.2023

List of Figures

Ich versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich keiner anderer als der in den beigefügten Verzeichnissen angegebenen Hilfsmittel bedient habe. **Außerdem erkläre ich ausdrücklich, dass für die vorliegende Arbeit, mit Ausnahme reiner Übersetzungs- und Grammatiktools, keine text- und mediengenerierenden KIs (wie ChatGPT) oder ähnliche elektronische Programme zum Einsatz kamen.** Alle Textstellen, die wörtlich oder sinngemäß aus Veröffentlichungen Dritter entnommen wurden, sind als solche kenntlich gemacht. Alle Quellen, die dem World Wide Web entnommen oder in einer digitalen Form verwendet wurden, sind der Arbeit beigefügt.

Weitere Personen waren an der geistigen Leistung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines Ghostwriters oder einer Ghostwriting-Agentur in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar Geld oder geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Arbeit stehen.

Der Durchführung einer elektronischen Plagiatsprüfung stimme ich hiermit zu. Die eingereichte elektronische Fassung der Arbeit ist vollständig. Mir ist bewusst, dass nachträgliche Ergänzungen ausgeschlossen sind.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht. Ich bin mir bewusst, dass eine unwahre Erklärung zur Versicherung der selbstständigen Leistungserbringung rechtliche Folgen haben kann.

Würzburg, XX. Month 20YY

....

(Philipp Treutlein)