Aus dem Interdisziplinären Schlafmedizinischen Zentrum
der Charité – Universitätsmedizin Berlin

DISSERTATION

Towards Interpretability and Reproducibility of Machine Learning Applications in Clinical Research

– Assessment of the Physiological Network in Sleep Apnea –

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Christoph Jansen

aus Trier

Datum der Promotion: 18.12.2020

Table of Contents

1 Table of Acronyms

2 Abstract

Objective: Machine Learning models, in particular Artificial Neural Networks, have shown to be applicable in clinical research for tumor detection and sleep phase classification. Applications in systems medicine and biology, for example in Physiological Networks, could benefit from the ability of these methods to recognize patterns in high-dimensional data, but decisions of an Artificial Neural Network cannot be interpreted based on the model itself. In a medical context this is an undesirable characteristic, because hidden age, gender or other data biases negatively impact the model quality. If insights are based on a biased model, the ability of an independent study to come to similar conclusions is limited and therefore an essential property of scientific experiments, known as results reproducibility, is violated. Besides results reproducibility, methods reproducibility allows others to reproduce exact outputs of computational experiments, but requires data, code and runtime environments to be available. These challenges in interpretability and reproducibility are addressed as part of an assessment of the Physiological Network in Obstructive Sleep Apnea.

Approach: A research platform is developed, that connects medical data, code and environments to enable methods reproducibility. The platform employs a compute cluster or cloud to accelerate the demanding model training. Artificial Neural Networks are trained on the Physiological Network data of a healthy control group for age and gender prediction to verify the influence of these biases. In a subsequent study, an Artificial Neural Network is trained to classify the Physiological Networks in Obstructive Sleep Apnea and a healthy control group. The state-of-the-art interpretation method DeepLift is applied to explain model predictions.

Results: An existing collaboration platform has been extended for sleep research data and modern container technologies are used to distribute training environments in compute clusters. Artificial Neural Network models predict the age of healthy subjects in a resolution of one decade and correctly classify the gender with 91% accuracy. Due to the verified biases, a matched dataset is created for the classification of Obstructive Sleep Apnea. The classification accuracy reaches 87% and DeepLift provides biomarkers as significant indicators towards or against the disorder. Analysis of misclassified samples shows potential Obstructive Sleep Apnea phenotypes.

Significance: The presented platform is extensible for future use cases and focuses on the reproducibility of computational experiments, a concern across many disciplines. Machine

learning approaches solve analysis tasks on high-dimensional data and novel interpretation techniques provide the required transparency for medical applications.

3 Kurzfassung

Ziel: Methoden des maschinellen Lernens, insbesondere künstliche neuronale Netze, finden Anwendung in der klinischen Forschung, um beispielsweise Tumorzellen oder Schlafphasen zu klassifizieren. Anwendungen in der Systemmedizin und -biologie, wie physiologische Netzwerke, könnten von der Fähigkeit dieser Methoden, Muster in großen Merkmalsräumen zu finden, profitieren. Allerdings sind Entscheidungen eines künstlichen neuronalen Netzes nicht allein anhand des Modells interpretierbar. In einem medizinischen Kontext ist dies eine unerwünschte Charakteristik, weil die Daten, mit denen ein Modell trainiert wird, versteckte Einflüsse wie Alters- und Geschlechtsabhängigkeiten beinhalten können. Erkenntnisse, die auf einem beeinflussten Modell basieren, sind nur bedingt durch unabhängige Studien nachvollziehbar, sodass keine Ergebnisreproduzierbarkeit gegeben ist. Neben der Ergebnisreproduzierbarkeit bezeichnet Methodenreproduzierbarkeit die Möglichkeit exakte Programmausgaben zu reproduzieren, was die Verfügbarkeit von Daten, Programmcode und Ausführungsumgebungen voraussetzt. Diese Promotion untersucht Veränderungen im physiologischen Netzwerk bei obstruktivem Schlafapnoesyndrom mit Methoden des maschinellen Lernens und adressiert dabei die genannten Herausforderungen der Interpretierbarkeit und Reproduzierbarkeit.

Ansatz: Es wird eine Forschungsplattform entwickelt, die medizinische Daten, Programmcode und Ausführungsumgebungen verknüpft und damit Methodenreproduzierbarkeit ermöglicht. Die Plattform bindet zur Beschleunigung des ressourcenintensiven Modelltrainings verteilte Rechenressourcen in Form eines Clusters oder einer Cloud an. Künstliche neuronale Netze werden zur Bestimmung des Alters und des Geschlechts anhand der physiologischen Daten einer gesunden Kontrollgruppe trainiert, um den Einfluss der Alters- und Geschlechtsabhängigkeiten zu untersuchen. In einer Folgestudie werden die Unterschiede im physiologischen Netzwerk einer Gruppe mit obstruktivem Schlafapnoesyndrom und einer gesunden Kontrollgruppe klassifiziert. DeepLift, eine Interpretationsmethode nach aktuellem Stand der Technik, wird zur Erklärung der Modellvorhersagen angewendet.

Ergebnisse: Eine existierende Forschungsplattform wurde für die Verarbeitung schlafbezogener Forschungsdaten erweitert und Containertechnologien ermöglichen die Bereitstellung der Ausführungsumgebung eines Experiments in einem Cluster. Künstliche neuronale Netze können

anhand der physiologischen Daten das Alter einer Person bis auf eine Dekade genau bestimmen und eine Geschlechtsklassifikation erreicht eine Genauigkeit von 91%. Die Ergebnisse bestätigen den Einfluss der Alters- und Geschlechtsabhängigkeiten, sodass für Schlafapnoeklassifikationen zunächst eine Datenbasis geschaffen wird, in der die Geschlechts- und Altersverteilung zwischen gesunden und kranken Gruppen ausgeglichen ist. Die resultierenden Modelle erreichen eine Klassifikationsgenauigkeit von 87%. DeepLift weist auf Biomarker und mögliche physiologische Schlafapnoe-Phänotypen im Tiefschlaf hin.

Signifikanz: Die vorgestellte Plattform ist für zukünftige Anwendungsfälle erweiterbar und ermöglicht Methodenreproduzierbarkeit, was über den Einsatz in der Medizin hinaus auch in anderen Disziplinen von Bedeutung ist. Maschinelles Lernen bietet sinnvolle Ansätze für die Analyse hochdimensionaler Daten und neue Interpretationstechniken schaffen die notwendige Transparenz für medizinische Anwendungszwecke.

4 Introduction

In recent years, Machine Learning (ML) models have become a popular tool in clinical research to improve or enable computer-aided diagnostics. Prominent examples, where ML models have outperformed other solutions, are the detection of metastatic tissue in pathological images [1] of lymph nodes as an indicator for breast cancer [2] [3], as well as automatic sleep scoring [4] and classification of Type 1 Narcolepsy [5] based on polysomnographic biosignal recordings. These advances are possible, because computational resources are more widely available and large amounts of patient data are collected. In contrast to a static computer program, ML models do not follow handcrafted rules representing the knowledge of domain experts. ML models represent mathematical functions with unknown, randomly initialized parameters. Based on existing data samples, these parameters are optimized in an iterative, computationally expensive training process, until the model closely predicts the desired output for the given samples. A well optimized and generalized ML model cannot only predict the outcome for training data samples, but also for unknown test data with an acceptable error rate. Depending on the number of parameters and the structure, ML models can solve different problems of varying complexity. Artificial Neural Networks (ANN) are a type of ML model with a very flexible architecture, that is constructed from several layers of artificial neurons with non-linear mathematical operations between each layer [6]. They can work with high-dimensional data and can adapt to different types of input data using different layer architectures. For example, cancer detection in high-resolution images uses Convolutional Neural Networks (CNN) that consist of layers performing image convolution operations [7] [8]. On the other hand, the narcolepsy classification uses Long Short-Term Memory (LSTM) layers [9] [10], that are capable of processing temporal data like biosignals. Despite the structural differences of the input data, the data values can be referred to as features, that together form a feature space. A feature can be an individual pixel of an image, a measurement value in a biosignal time series, or any kind of engineered feature, like the statistical sleep parameters used in sleep research [11]. Based on such features, ANNs can be trained to solve various types of problems. In case of a classification problem, an ANN can learn to transform a high-dimensional feature space into a low-dimensional representation. This representation is linearly separable in the ANNs output layer to classify the data. In the context of this thesis, any type of algorithmic analysis that relies on digital assets, like an ANN classification training, is referred to as computational experiment.

Although ANNs can outperform other methods in terms of a decreased prediction error, they impose various challenges, that need to be addressed to be a viable solution in a medical context. Compared to some simpler ML models, like Decision Trees [12], ANNs do not have the intrinsic ability to be interpretable with respect to existing domain knowledge [13] [14], because a non-linear mathematical model with thousands of parameters is incomprehensible for humans. If, for example, the diagnostic prediction of an ANN leads to a certain treatment, it would be desirable to understand the decision process [15]. This problem is even more amplified, when it comes to hidden biases in the data [16], which can be learned by ML models to distinguish patient groups by features not related to the actual disease. The result is a misclassification of patients, that do not share the typical secondary features of other patients with the same disease. As an example, Obstructive Sleep Apnea (OSA) is more prevalent in elderly men [17] and therefore datasets contain more samples of this group. An ANN model trained on such a skewed dataset, could rely on age- and gender-related features to support its predictions without this misbehavior being discovered. To understand or visualize how important individual features are for an ANN prediction, new external interpretation methods for various ANN architectures have been developed in recent years [18] [19] [20] [21] [22]. These methods can be compared to each other in benchmarks [23] [24], but also allow ANNs to compete with feature selection algorithms [25] [26], statistical testing on group differences [27] and tree-based ML models [12] [28] [29], that are intrinsically interpretable. Furthermore, the problem of data biases does not only relate to interpretability, but also to reproducibility [30] [31] [32]. Achieving results reproducibility, as defined by Goodman et al. [33], requires other researchers to being able to obtain similar results in an independent study that supports the original findings. If the original study did not account for biases, for example by not age matching data samples between patient groups, the results might not be reproducible. Compared to results reproducibility, a more technical aspect is methods reproducibility [33]. It requires the actual resources that have been used in the experiment to be available to other researchers, to allow them to reproduce exact outcomes, like statistics or plots, as they have been published by the original authors. In the context of computational experiments, relevant resources are data, code and environment [34]. Code can be shared publicly on popular code hosting platforms like GitHub, where for example the ANN models for Narcolepsy classification [5] can be found[1]. In order to execute this code, many software dependencies like an operating system, programming libraries and script interpreters must be installed and configured on a computer system. These environments can be prepared and distributed to others using modern

[1] https://github.com/Stanford-STAGES/stanford-stages

virtualization and container technologies [35]. While sharing code and environments is feasible, the public accessibility of medical data is limited due to ethical considerations. Here, medical Data Management Systems (DMS) [36], like the Extensible Neuroimaging Archive Toolkit (XNAT) [37], that allow for data access restrictions, can be used to share data with authorized users over a network. Although various technical systems and platforms exist to make data, code and environment available, methods reproducibility can only be achieved if these systems follow the FAIR Guiding Principles [38], that require resources to be findable, accessible, interoperable and reusable. These principles require the usage of global identifiers, support for standard network transmission and authentication protocols, to store data in open formats and to follow community standards. Furthermore, combining the described resources to run a defined computational experiment is a rather difficult technical hurdle. Compared to established grid computing infrastructures and science gateways [39] [40] [41] [42], modern research platforms and workflow runtimes [43] [44] leverage container technologies for reproducible environments. While these tools provide platform-specific ways of handing data into these environments, they are not flexible enough to work well with external data sources like XNAT, that have complex programming interfaces. In addition, they are not designed with ML applications in mind and therefore do not explicitly support the usage of Graphics Processing Units (GPU) that are hardware devices typically used to accelerate the training of ANNs [45].

It is important to address the described challenges, because advances in the fields of systems medicine and biology [46] [47], that study the complex interactions in the human body, rather than attributing problems to individual organs, require "new computational and analytical approaches to extract information" [48]. In particular multidimensional biosignals recorded during sleep, known as polysomnographies (PSG), provide a rich data source to analyze Physiological Network (PN) interactions across the human body and how these interactions are altered in the context of sleep disorders. In a PN, each network node represents an organ system or specific characteristic of it. The edges of the network, that connect each pair of nodes represent the coupling of these organ systems. ML methods, like ANNs, can be applied to classify diseases based on these high-dimensional feature spaces formed by the edges in a PN, to discover underlying patterns or even novel biomarkers [49]. In the course of this dissertation, ML methods have been applied, to discover physiological differences in patients suffering from OSA. OSA is a disorder with high prevalence [50], that has the potential to disrupt physiological processes and the common pattern of sleep cycles, when compared to healthy individuals [51]. Affected people stop breathing (apnea) or have a reduced airflow (hypopnea) for more than 10 seconds [52], due to a "clinical condition

in which the throat narrows or collapses repeatedly during sleep" [53]. The Apnea Hypopnea Index (AHI) indicates the number of apnea or hypopnea events per hour during a night. Depending on this index, OSA is classified as mild (AHI 5–15), moderate (AHI 15–30) or severe (AHI >30) [52]. Senaratna et. al have reported OSA prevalence in a recent meta-study: "At ≥5 events/h apnea-hypopnea index (AHI), the overall population prevalence ranged from 9% to 38% and was higher in men. It increased with increasing age and, in some elderly groups, was as high as 90% in men and 78% in women. At ≥15 events/h AHI, the prevalence in the general adult population ranged from 6% to 17%, being as high as 49% in the advanced ages." [17]. As suggested by these numbers, the prevalence increases with age and is higher in men than in women. Disturbed sleep often leads to daytime sleepiness, but not every affected person suffers from negative conditions. Although the prevalence of OSA is higher in elderly people, daytime sleepiness in this age group is less of a problem compared to young patients [54]. In terms of known network effects, an increased cardio-respiratory coordination in OSA patients has been reported [55].

To enable interpretability and reproducibility of the conducted experiments, the previously described challenges are addressed on different levels. First, a collaborative research platform centered around XNAT is introduced, that supports the handling of medical data and connects to a private cloud or compute cluster for efficient and distributed processing of experiments. The platform is further evaluated and improved to enable methods reproducibility and to follow the FAIR principles. External data sources and GPUs for ANN training acceleration are explicitly supported. An analysis conducted in the platform is the assessment of age- and gender-related differences [56] [57] [58] in the PNs of a healthy control group during sleep. As part of the study, ANNs are trained to predict age and gender based on the PN data, because the contained information could function as an underlying bias in other classification tasks. In a subsequent study, ANNs are trained to classify OSA based on the same type of PN and such that the ANN classification decisions can be explained using a state-of-the-art interpretation method called DeepLift [21]. The performance of the ANN classifiers is compared to other interpretable ML methods and the relevance of features is verified using statistical tests. For these methods not to rely on the described biases, the training is performed on an age and gender matched dataset. Countermeasures against data biases and the interpretation of models contribute towards an improved results reproducibility of the presented computational experiments.

5 Methods

5.1 SIESTA Database

The SIESTA database [59] contains PSGs from various patient groups, including 100 recordings from 50 OSA patients and 393 recordings from 197 healthy controls. The healthy control group contains 103 female and 94 male subjects. Two consecutive nights have been recorded for most subjects. Eight European sleep labs have taken part in the SIESTA study. Included PSGs contain at least six EEG channels to capture brain activity, EOG measuring activity of the eyes, a chin EMG, an ECG, as well as three respiratory signals (nasal, chest, abdomen) and oxygen saturation. Each EEG channel is split into delta, theta, alpha, sigma and beta frequency bands [60], resulting in 38 signals in total. Associated with each PSG is a hypnogram created by a human scorer following the Rechtschaffen and Kales scoring rules [61]. The hypnograms describes the sleep stage transitions during the course of a recorded night, containing labels for wake (W), rapid eye movement (REM) and non-REM (N1, N2, N3, N4) stages. Since the distinction between N3 and N4 deep sleep stages is not made in the newer AASM scoring criteria [62] [63], the N4 stages were labeled as N3 stages for the purpose of the presented studies. All PSGs are stored in the standardized European Data Format (EDF) [64].

5.2 Data Normalization

EDF+ is an addition to the original EDF standard, that defines strict formatting rules for EDF header fields [65]. Since the EDFs contained in the SIESTA database have been created before the release of EDF+ and have been exported from different recording devices, they do not always contain consistent information and strictly formatted header fields. For the purpose of this work, an existing open-source MATLAB application[2] was employed to convert individual EDFs to contain normalized channel names and header fields.

5.3 Physiological Networks

Physiological Networks (PN) describe the physiological interactions between organ systems. The coupling of organs can be calculated using various methods. As a foundation for the presented experiments, the Time Delay Stability (TDS) [66] method has been applied. TDS is based on cross-correlation to determine if two signals are coupled at a certain point in time. The cross-correlation

[2] Dagmar Krefting, https://github.com/somnonetz/psgScan2edfData

determines a certain time-delay of one signal to the other, where correlation is maximized. If this time-delay does not change across five subsequent time steps, the coupling is considered stable. When first introduced, the concept was used to find changes in physiological interactions during sleep stage transitions. This universal method is applicable to combinations of many organ systems, including brain EEG frequency bands, EOG, chin and leg EMG, respiration and heart rate. A follow-up study has identified rules for brain-brain, brain-periphery and periphery-periphery systems in healthy subjects [67]. Concerning brain-brain interactions, intra- and inter-channel communication is distinguished. In a single brain hemisphere, inter-channel couplings between frontal and central EEG are stronger than couplings between frontal and occipital EEG. Across hemispheres, the coupling of frontal channels is stronger than the coupling of central channels, which again is stronger than occipital coupling. Concerning brain-periphery interactions, the EOG is coupled with the frontal EEG. In general, during W and N2 stages couplings occur more often and are stronger than during N3 and REM stages.

Using the TDS method [66], the PN topology changes over the course of the recording time. The time dimension of a signal pair can be summarized by taking the average value per sleep stage. Based on the 38 preprocessed PSG signals, this results in 2812 feature values for the sleep stages W, N2, N3 and REM. This high-dimensional feature space is a summarized description of the PN topology based on a single PSG. The TDS method shows great flexibility, because it is applicable to all standard PSG channels contained in the SIESTA EDFs. Again, an existing MATLAB application[3] is used to generate TDS values from previously normalized EDFs.

5.4 Infrastructure

5.4.1 Container Appliances

While data normalization and TDS algorithms are provided as existing MATLAB applications, other programs, including ANN models, had to be implemented for the presented studies. The implementation of ANNs has been carried out in Python, a general-purpose programming language with support for popular ML frameworks. Tree-based ML methods have been implemented using the software ecosystem of the R language. MATLAB, Python and R applications each require specific runtimes and additional software libraries to be installed on the Linux host operating system (OS). To enable the execution of these applications on remote servers,

[3] Dagmar Krefting, https://github.com/somnonetz/physiological-networks-tds

without the requirement of setting up the specific runtime environment on the given server, the applications and all their respective dependencies can be installed in Docker container images. Container technologies provide process isolation from the host OS, which means that instead of accessing the host's file system, they provide their own minimalistic file systems in the form of a container image. A bundled application, also referred to as appliance, can be uploaded to a Docker registry. Given the required authorization credentials, any computer running a docker-engine on top of a Linux operating system can pull the appliance from the registry using a URL and execute it without any further modifications.

5.4.2 Common Workflow Language

A Docker appliance on its own, does only solve the problem of distributing a given software program, but does not provide a standardized way of executing it. Containers can either contain long-running server processes, which we are not interested in for the given use case, or can execute command-line interface (CLI) applications, that terminate after finishing an analysis. Every CLI application accepts certain parameters, for example a path to an input file or a parameter to change the behavior of the program, to produce certain outputs. Since every CLI application is different, a person or automation process trying to run the software, does not know how to do so. The Common Workflow Language (CWL) [68] solves this problem, by providing a CLI specification language, which allows for formalization of the process execution.

5.4.3 OpenStack

Compared to Grid computing infrastructures, modern cloud computing technologies use virtual machines (VM) and virtual networks to partition different kinds of physical resources, like storage, network and processing time. This allows for fine-grained control in a multi-tenant environment. The OpenStack Nova client API allows for programmatic control of the cloud resources. Docker containers can be started inside of VMs using a Docker client or can be launched outside of VMs under control of OpenStack Nova. While many commercial cloud infrastructures are being operated by international corporations, the OpenStack[4] cloud platform is an open source project and can be deployed on premise, for example in clinics or university networks.

[4] https://www.openstack.org/

5.5 Artificial Neural Networks

While different ANN architecture for specific use-cases exist, this section only discusses simple feed-forward ANNs for classification and regression problems. ANNs can incorporate combined information from large feature-spaces into their predictions. They are organized in layers of artificial neurons to apply subsequent non-linear transformations to the data. Each layer is determined by a matrix of weights. In a so-called forward pass, the matrix of one or more data samples is passed to the first layer, where it is multiplied with the weights of the first layer and then transformed by a non-linear activation function. The result is passed to next layer where this procedure repeats. The values of the weights are initialized randomly and optimized during training. For a given input, a loss function calculates an error value, by comparing the actual output of the network with the desired output. A stochastic optimizer, like Adam [69], iteratively adapts the weights to improve the ANN output based on the calculated loss. A typical ANN has one input layer, where the number of input neurons corresponds to the number of features in the data. This input layer is followed by an arbitrary number of hidden layers, with an arbitrary number of neurons per layer. The actual number of layers and neurons are hyperparameters, that determine the complexity of the model. If the hyperparameters are defined in a way such that each layer has an equal or smaller number of neurons than its predecessor, the ANN can effectively transform high-dimensional data representations into low-dimensional representations. Throughout the experiments, the well-known Rectified Linear Units (ReLU) [70] activation function has been used for input and hidden layers. The last hidden layer is followed by a single output layer, that can have different properties, depending on the problem at hand.

For binary classifications, e.g. male and female classes or healthy and OSA classes, two output layer options exist. The first option is, to encode the output as a single neuron with a sigmoid activation function [71]. This function always yields a response in the range]0, 1[, where 0 and 1 represent the classes. The sigmoid activation will never be exactly 0 or 1, instead a classification decision is determined at a classification boundary, usually at 0.5. Such a model can be trained using the binary cross-entropy loss function. The second option is to encode the classes in two separate neurons. The softmax activation function is applied to the output layer, such that the sum of all neurons is one [72]. The value of each neuron can then be interpreted as a class probability. The classification is therefore determined by the highest probability. This model configuration can be trained using the categorical cross-entropy loss function.

In a regression problem, for example predicting a continuous age value, no activation function is applied to the output layer. The difference between the continuous output value and the continuous target value is determined by the Mean Squared Error (MSE) loss function.

5.5.1 Architecture and Hyperparameters

The input data to the model are 2812 TDS features, each sample representing a whole night. As described in section 5.3, all features are average values, that summarize the coupling during a sleep stage. Due to taking the average values, there is no temporal information remaining in the PN features. Therefore, a standard feed-forward ANN architecture can be used, where all features are passed into the network at once. Working with temporal data would otherwise require a more complex architecture, for example with CNN or LSTM layers. For some experiments we have conducted a hyperparameter grid-search, where ANNs are trained on all possible combinations from a list of predetermined hyperparameter choices to yield optimal results [73]. A grid-search is easy to implement, but, due to the potentially large number of combinations, requires a compute cluster to execute the experiments in a reasonable amount of time.

5.5.2 Overfitting Avoidance

The number of data samples in the SIESTA database is an order of magnitude lower than the number of features per sample. ANN models represent complex, non-linear mathematical functions, which are likely to overfit on datasets with small sample sizes. An overfitted model has learned to identify individual samples and therefore to separate these specific samples into their respective classes. Conversely this means, that the trained model does not represent general knowledge about the given classes, for example healthy and OSA, and is not able to classify validation data correctly. Analyzing the feature importance on basis of such a model would lead to false results and must be avoided. Four complementary strategies for overfitting avoidance are employed. A) Data is matched by external properties, like age and gender, if these properties are not the classification target [74]. In consequence, age and gender are equally distributed for each target class. B) Training and validation data splits are stratified for equal class distribution [75]. C) With Dropout regularization a certain percentage of neurons per layer is randomly set to zero at each training step [76]. Dropout influences the training process, such that an ANN is forced not to rely on individual features and instead to learn general patterns in the data. The Dropout percentage per layer are additional hyperparameters. D) A 5-fold Cross-Validation (CV) strategy

is employed, where the classification performance is evaluated across five different training and validation data splits [77].

5.5.3 Model Evaluation

The most basic, but easy to interpret, evaluation score for classifiers is accuracy, the ratio of correctly classified samples to the total number of samples. Since the focus of this work is the interpretation of models and not model performance, fine-grained scores like precision, recall (sensitivity), F-score and specificity were not required. For evaluation of continuous age predictions in a regression problem the mean absolute error (MAE) is used. The error term is calculated by subtracting the predicted value from the target value and taking the absolute of the result. The mean is then calculated over all samples. In contrast to the MSE function, that is used during model training, the MAE value is easier to interpret as an absolute age difference.

Fleiss' Kappa is a score for interrater reliability, where the agreement in decisions from different raters is measured [78]. The score ranges from -1 to 1, with 1 denoting perfect consensus in all decisions and -1 denoting complete disagreement. Compared to Cohen's Kappa [79], Fleiss' Kappa is applicable to an arbitrary number of raters. In an equivalent way, Fleiss' Kappa can be applied to classification models. In a 5-fold cross-validation approach five models of the same type are trained on five different subsets of the data. If the models are not overfitting on a specific training subset of the data, each model of the same type should provide very similar or ideally the exact same classification decisions on the full dataset with the Fleiss' Kappa score being 1. In this case, we interpret the score as the intra-type consistency. For comparison of different types of classifiers, the score can as well be interpreted as an inter-type agreement.

5.5.4 DeepLift

DeepLift [21] is a method to determine the importance of each neuron in an ANN. A forward pass through the network is calculated to provide a classification decision for a given sample. From the classification output a backward pass determines the importance of neurons one layer at a time, starting with the last hidden layer and finishing with the input layer. Importance values are calculated in reference to a neutral data input, for example a zero vector. We have applied DeepLift to a binary classification problem with a single output neuron. The importance value of a neuron can be positive or negative, with positive values indicating that higher inputs to the neuron

contribute more towards class 1 and negative values indicating that larger inputs to the neuron contribute more towards class 0.

5.6 Age- and Gender-Related Group Differences

Since the prevalence of OSA is increased in elderly men, the distribution of subjects in the SIESTA database is skewed towards this group. If not conducted carefully, this underlying bias could potentially lead to misguided results. OSA classifiers trained on a skewed dataset, might learn to differentiate the disorder by age- or gender-related differences instead of actual OSA features. Therefore, in a baseline analysis, we asses related studies and compare reported age and gender dependencies to statistical group differences in PNs of healthy controls. To validate the findings of the employed tests, different ANN architectures for regression and classification tasks are trained. This demonstrates the capabilities of the models to predict age (regression) and gender (classification) based on the TDS features.

6 Implementation and Results

This section summarizes the implementation and results of three subsequent studies: “Multicenter data sharing for collaboration in sleep medicine” [80] (including associated studies [81] [82] [83] [84] [85]), “Age and gender dependency of physiological networks in sleep” [86] and “Feature relevance in physiological networks for classification of obstructive sleep apnea” [87].

6.1 Multicenter Data Sharing for Collaboration in Sleep Medicine

The proposed collaboration platform is centered around XNAT as a data management system. Development started in an earlier project [81], with the aim to enable automated quality assessment of retinal Optical Coherence Tomography (OCT) scans. Due to XNAT’s extensibility, the system was then adapted to store PSGs in EDF format, as well as proprietary vendor formats [80] [84]. Custom XNAT data types have been developed to hold sleep research specific metadata, including biosignal channel information. Data is organized in projects with access restrictions for unauthorized users, but inter-project collaboration is supported by sharing selected patient data or recording sessions. Data storage and processing have different server hardware requirements. The XNAT server provides high availability and storage capacities, but only moderate processing power. Fully utilizing the server’s CPU with data processing jobs would lower the availability of the data. The collaboration platform therefore connects the XNAT data storage to a separate cluster

with multiple compute nodes for processing tasks. Due to the redundancy in the cluster, high availability is a not a concern for individual nodes. Available hardware resources are managed by the OpenStack cloud platform. A custom Job Management System (JMS) schedules processing tasks in the cluster. These jobs are triggered by XNAT pipelines, either manually through XNAT's web interface or automatically when new data is uploaded. Output data generated by a job is stored back to XNAT. Via the JMS, different Docker appliances can be launched to process data, for example to normalize data, to extract TDS features from the previously normalized data or to train ANN models on TDS features.

6.1.1 Job Management System

The Job Management System (JMS) is a server-side software component with a network-based programming interface [88] to receive job information from XNAT. It is loosely coupled with XNAT and can be integrated with other Data Management Systems (DMS). Originally, the JMS was developed to connect with OpenStack Nova to launch Virtual Machines (VM) or Docker containers and was therefore not suited to work outside of an OpenStack environment. Support for Nova has since been dropped in favor of directly connecting with Docker [82]. This allows for more flexibility, by either running containers inside of OpenStack VMs or by running them directly on a server's host Operating System (OS).

The core concept of the JMS is to control the temporary availability of data in the compute cluster and to maximize the security of the medical data. For this purpose, the JMS never directly starts an application process in a container, but instead starts a supervisor process. This supervisor process downloads input data from a DMS directly into the container file system, launches the intended application to process the data and uploads the output data back to the DMS. As soon as the supervisor process terminates, the container and its file system are deleted from the compute node, leaving no remnant data on the server. Data transfer is handled by so-called connector programs, that must be included in the Docker appliance. Besides data transfer with XNAT, different connector implementations can enable communication via standard transmission and authentication protocols like HTTPS or SSH.

6.1.2 Methods Reproducibility

Since reproducibility is a significant concern in computational research, the components of the collaboration platform have been evaluated in terms of the FAIR Guiding principles during a

subsequent study [83]. XNAT implements or supports all of these principles, because the data is findable via URLs or the integrated metadata search engine, it is accessible via a standard REST interface, the implemented data exchange formats XML, JSON and DICOM are interoperable and reusable. At the time of the study, the JMS fulfilled most of the principles, by providing a REST interface with URLs to job metadata, and by implementing support for standard transmission protocols via connectors. In terms of the reusability aspect it fell short, due to a custom job description format, that did not follow community standards. The job format has since been replaced by Reproducible Experiment Descriptions (RED) [85], a new format based on the open CWL standard. A RED file contains an application CLI description in CWL format, input parameters, connector references for input and output data, container engine settings, including a reference to the Docker appliance in a registry, as well as execution engine settings. This complete description of an experiment can be used to reproduce results on any computer, that has a Linux OS and Docker installed, if the hardware requirements are satisfied. The JMS is now known as CC-Agency and has been published as part of the Curious Containers[5] open source project. CC-Agency is one of two available RED execution engines. Docker appliances created to be compatible with Curious Containers, are also compatible with CWL runtimes to prevent experiments from being locked in the collaboration platform. Besides the sleep research related workloads, the RED execution engines have been successfully evaluated to support a computationally demanding CNN workload for cancer detection in pathological images [85]. This workload requires access to large amounts of training data, that may be too large to be stored on a local computer. To circumvent this problem, support for Filesystem in Userspace (FUSE) [89] has been added to the software. FUSE network filesystems allow remote data directories, that are located on a dedicated storage server, to be mounted on top of a local filesystem. The remote data is then transparently accessible on the local computer or within a Docker container. The FUSE implementation SSHFS[6] does not store data, but transfers chunks of data on demand as they are requested by an application. SSHFS enables efficient network access to a CAMELYON16 training database, that contains more than a terabyte of tissue image tiles. Additionally, support for the Nvidia-Docker technology was implemented to accelerate the process using GPUs and therefore reduce CNN training durations. For experiments that rely on these technologies, the filesystem and GPU requirements are documented in their respective RED files.

[5] https://www.curious-containers.cc/
[6] https://github.com/libfuse/sshfs

6.2 Age and Gender Dependency of Physiological Networks in Sleep

The second study investigates age- and gender-related differences in PNs of healthy controls from the SIESTA database. The PNs are based on the TDS algorithm, with 703 calculated biosignal couplings for each sleep stage (W, N2, N3, REM). The study conducts a statistical analysis and verifies the results using ANNs for non-linear classification and regression.

Results from the statistical analysis show, that women have significantly higher couplings in different central and occipital brain regions during N2 and REM sleep stages. Other differences in men and women, that have been detected, are not statistically significant. With increased age, a significant decrease in coupling during N2, N3 and W sleep stages for male and female subjects was found.

Five types of ANN classifiers have been trained on 703 TDS features from the four individual sleep stages and a combined input containing all 2812 features. The ANNs are configured with two output neurons representing both classes and use the softmax activation function. In a hyperparameter grid-search, 1800 hyperparameter combinations have been trained for each type of ANN, resulting in a total of 9000 ANN training experiments. The set of possible hyperparameters contains different Dropout values for input and hidden layers. All experiments have been carried out in the collaboration platform's job management system. The highest test classification accuracy of 0.91 was achieved with the combination of all features. On separated sleep stage input data, the classification accuracies are 0.89 for N2, 0.88 for R, 0.79 for W and 0.77 for N3. These results are in line with the statistical analysis, where some brain couplings are significantly higher in women during N2 and REM sleep stages.

Equivalent to the ANN gender classification task, a hyperparameter grid-search for five types of ANNs has been carried out to perform a non-linear regression for age prediction using a single output neuron. With a MAE of 4.79 years, the best results are achieved on the N2 sleep stage data. The result for combined features from all sleep stages is slightly lower and reaches a MAE of 5.42 years. This result indicates, that the predictive power of the data lies in N2 and that the other sleep stages add noise to the combined features that lower the overall ANN performance. The analysis was repeated for male and female subgroups respectively. The age prediction error in N2 was slightly lower for the male subgroup (MAE = 4.50 years), than for the female subgroup (MAE = 5.16 years).

ANN classification and regression results show, that ANNs can be successfully trained on TDS features using Dropout for overfitting avoidance, even though the number of input features is an order of magnitude larger than the number of data samples. This information is required for the subsequent OSA study, where age and gender-related biases in the data might influence the results, if no countermeasures are applied.

6.3 Feature Relevance in Physiological Networks for Classification of Obstructive Sleep Apnea

The third study compares the PN topologies of OSA patients to healthy controls during sleep. As shown for age and gender differences, ANNs can be trained to solve classification and regression tasks based on TDS features. Similarly, ANNs are trained on TDS features to classify OSA and the interpretation technique DeepLift is applied to calculate the relevance of features based on a given classification decision. DeepLift feature relevance scores are compared with scores obtained from Decision Trees, Random Forests and statistical testing on group differences.

For a meaningful OSA classifier it is crucial not to rely on age and gender information, because the age distribution of OSA patients in the SIESTA dataset is skewed towards older patients and most of them are male. To avoid these biases, 48 healthy controls have been age and gender matched with 48 OSA patients and stratified training, validation and test data splits have been created from the matched samples. This bias avoidance strategy improves result reproducibility of the experiments. The final dataset contains 188 PSGs from the 48 subject-pairs. In contrast to the binary classification ANNs of the second study, the ANNs are configured with a single output neuron using the sigmoid activation function. DeepLift is applied to obtain relevance scores from a classification decision of this single output neuron, where otherwise different DeepLift scores would be obtained from two output neurons.

Due to the small number of samples, the ANN classifier was prone to overfit on the training data. In order to reduce overfitting as much as possible, we found a very high Dropout rate of 95% applied to the ANN input layer to yield the best results (ANN95). The resulting ANN95 models reach a mean test accuracy of 87% during cross-validation, compared to a mean accuracy of 96% on the training data. The Fleiss' Kappa score measuring intra-model consistency is 0.91. In our experiments the ANN95 models reach, compared to ANNs with 20% input Dropout, as well as Decision Tree and Random Forest models, the highest accuracy and intra-model consistency.

Feature relevance scores have been obtained for all data samples from the five cross-validation ANN95 models using DeepLift. From the individual sample scores, the mean scores have been calculated across the ANN95 models for each classification group, OSA and healthy. The results show, that during N2 increased physiological couplings from breathing rate or chin EMG signals to most other signals, including brain EEGs, are an indication for OSA. During N3 on the other hand, mostly decreased coupling is an indication for OSA. Relevant decreases are found from the right EOG to the right-side EEG bands and to O1 and C4 delta EEG bands. Additionally, a decreased coupling between both brain hemispheres in C3 and C4 sigma EEG bands is a relevant feature for OSA classification. The alternative ML models, Decision Trees and Random Forests, each provide slightly different rankings of the 2812 features, but none of the results contradict each other. The five most relevant features of each model are at least in the top 10% of most relevant features from every other model.

A useful property of DeepLift is, that an individual assessment of misclassified samples is possible, to find the relevant features for the wrong classification outcome. Some of the OSA patients, that have been misclassified as healthy, show typical couplings with breathing and chin EMG signals during N2, but do not show the decreased coupling in brain EEG signals during N3. Two of three misclassified healthy controls show OSA related couplings in N2 and have an AHI of ~20 and ~40 respectively.

7 Discussion

The collaboration platform has been set up successfully, with EDF raw data and processing results being stored in XNAT. In subsequent studies the JMS was improved to be compatible with the existing CWL standard and to focus on methods reproducibility.

In order to find the best ANN models for gender classification and age prediction, the scheduling capabilities of the JMS were used to run a hyperparameter grid-search optimization in a compute cluster. Based on PN data, such an optimized ANN model can predict the age of a subject in the range of a decade. The subject's gender was classified correctly in 91% of the cases. These results confirm, that age and gender information is contained in PNs and that ANNs can learn to use these data patterns for predictive tasks.

ANNs and other ML models have been trained to identify relevant PN features for OSA classification. For this approach to yield meaningful results, it is crucial that the trained models do

not rely on unrelated biases, that might be present in the data. Based on our previous findings, we have only used age and gender matched data samples for OSA classification training, making it unlikely for the classifier to rely on these secondary attributes. As shown in the study, the relevance values obtained from ANN models via DeepLift are consistent with other types of models and statistical tests. The identified features are in good agreement with the sleep medicine and physiology literature. In addition, we have identified potential differences in PN phenotypes, where some OSA patients do not show reduced brain EEG couplings during N3. However, in young OSA patients, the strength in sigma C3 and C4 couplings seems to diverge more strongly from healthy subjects in the same age group.

8 Conclusion

Different PN analysis tasks based on a high-dimensional feature-space were implemented successfully. The development of a collaboration platform allowed for structured data management and the usage of distributed compute resources. The platform has specific extensions for sleep data support, but is not limited to these data types, as has been shown in an earlier OCT quality analysis and a cancer detection use case. The JMS was published as part of the Curious Containers open source software project and is still in ongoing development.

ANN models have been successfully trained on PN data, but strict overfitting avoidance was required. DeepLift is an interpretation technique for ANNs that was applied to OSA classifiers. The method provides meaningful results in the context of high-dimensional feature-spaces and relevance scores are consistent with established ML methods. Due to its generic nature it shows great potential for other medical use cases, where black-box classifiers are not desirable, but powerful models like ANNs are required.

Physiological couplings, that differ in strength between OSA patients and healthy controls, have been revealed by the interpretation techniques. Most of the findings are in line with other studies, but the potential PN phenotypes for OSA patients during N3 still need confirmation from the sleep physiology community.

9 Acknowledgements

The work was supported by the German Federal Ministry of Economic Affairs and Energy (ZIM project BeCRF, grant number KF3470401BZ4), the German Federal Ministry of Education and

Research (project deep.TEACHING, grant number 01IS17056 and project deep.HEALTH, grant number 13FH770IX6) and the HTW Berlin doctoral program grant (HTW Berlin Booster).

10 References

[1] S. Al-Janabi, A. Huisman und P. J. Van Diest, „Digital pathology: current status and future perspectives," *Histopathology,* Bd. 61, pp. 1-9, 2012.

[2] P. Bandi, O. Geessink, Q. Manson, M. Dijk, M. Balkenhol, M. Hermsen, B. E. Bejnordi, B. Lee, K. Paeng, A. Zhong und others, „From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge," *IEEE transactions on medical imaging,* 2018.

[3] B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. Ginneken, N. Karssemeijer, G. Litjens, J. A. W. M. Laak und CAMELYON16 Consortium, „Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer," *JAMA,* Bd. 318, pp. 2199-2210, 12 2017.

[4] P. Anderer, G. Gruber, S. Parapatics, M. Woertz, T. Miazhynskaia, G. Klösch, B. Saletu, J. Zeitlhofer, M. J. Barbanoj, H. Danker-Hopfe, S. L. Himanen, B. Kemp, T. Penzel, M. Grozinger, D. Kunz, P. Rappelsberger, A. Schlogl und G. Dorffner, „An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24× 7 utilizing the Siesta database," *Neuropsychobiology,* Bd. 51, pp. 115-133, 2005.

[5] J. B. Stephansen, A. N. Olesen, M. Olsen, A. Ambati, E. B. Leary, H. E. Moore, O. Carrillo, L. Lin, F. Han, H. Yan, Y. L. Sun, Y. Dauvilliers, S. Scholz, L. Barateau, B. Hogl, A. Stefani, S. C. Hong, T. W. Kim, F. Pizza, G. Plazzi, S. Vandi, E. Antelmi, D. Perrin, S. T. Kuna, P. K. Schweitzer, C. Kushida, P. E. Peppard, H. B. D. Sorensen, P. Jennum und E. Mignot, „Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy," *Nature communications,* Bd. 9, p. 5229, 2018.

[6] J. Schmidhuber, „Deep learning in neural networks: An overview," *Neural Networks,* Bd. 61, pp. 85-117, 2015.

[7] J. T. Springenberg, A. Dosovitskiy, T. Brox und M. A. Riedmiller, „Striving for Simplicity: The All Convolutional Net," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings,* 2015.

[8] A. Krizhevsky, I. Sutskever und G. E. Hinton, „Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems,* 2012.

[9] S. Hochreiter und J. Schmidhuber, „Long short-term memory," *Neural computation,* Bd. 9, pp. 1735-1780, 1997.

[10] F. A. Gers, J. Schmidhuber und F. Cummins, „Learning to forget: continual prediction with LSTM," in *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470),* 1999.

[11] H. Danker-Hopfe, D. Kunz, G. Gruber, G. Klösch, J. L. Lorenzo, S. L. Himanen, B. Kemp, T. Penzel, J. Röschke, H. Dorn, A. Schlögl, E. Trenker and G. Dorffner, "Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders," *Journal of Sleep Research,* vol. 13, pp. 63-69, 3 2004.

[12] G. James, D. Witten, T. Hastie und R. Tibshirani, „Tree-Based Methods," in *An Introduction to Statistical Learning: with Applications in R*, New, York: Springer New York, 2013, pp. 303-335.

[13] R. Abbasi-Asl, Interpretable Machine Learning with Applications in Neuroscience, Berkeley, USA: University of California, 2018.

[14] M. A. Ahmad, C. Eckert und A. Teredesai, „Interpretable Machine Learning in Healthcare," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB 2018, Washington, DC, USA, August 29 - September 01, 2018*, 2018.

[15] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter und C. S. Greene, „Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface,* Bd. 15, 2018.

[16] P. R. Rosenbaum, „Overt Bias in Observational Studies," in *Observational Studies*, New, York: Springer New York, 2002, pp. 71-104.

[17] C. V. Senaratna, J. L. Perret, C. J. Lodge, A. J. Lowe, B. E. Campbell, M. C. Matheson, G. S. Hamilton und S. C. Dharmage, „Prevalence of obstructive sleep apnea in the general population: A systematic review," *Sleep Medicine Reviews,* Bd. 34, pp. 70-81, 2017.

[18] K. Simonyan, A. Vedaldi und A. Zisserman, „Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Workshop Track Proceedings*, 2014.

[19] M. D. Zeiler und R. Fergus, „Visualizing and Understanding Convolutional Networks," in *Computer Vision -- ECCV 2014*, Cham, 2014.

[20] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller und W. Samek, „On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation," *PLOS ONE,* Bd. 10, pp. 1-46, 7 2015.

[21] A. Shrikumar, P. Greenside und A. Kundaje, „Learning Important Features Through Propagating Activation Differences," in *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, 2017.

[22] S. M. Lundberg und S.-I. Lee, „A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan und R. Garnett, Hrsg., Curran Associates, Inc., 2017, pp. 4765-4774.

[23] M. Ancona, E. Ceolini, C. Öztireli und M. Gross, „Towards better understanding of gradient-based attribution methods for Deep Neural Networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[24] N. Pörner, H. Schütze und B. Roth, „Evaluating neural network explanation methods using hybrid documents and morphosyntactic agreement," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, 2018.

[25] J. Van Hulse, T. M. Khoshgoftaar, A. Napolitano und R. Wald, „Threshold-based feature selection techniques for high-dimensional bioinformatics data," *Network Modeling Analysis in Health Informatics and Bioinformatics,* Bd. 1, pp. 47-61, 01 6 2012.

[26] I. Guyon, J. Weston, S. Barnhill und V. Vapnik, „Gene Selection for Cancer Classification using Support Vector Machines," *Machine Learning,* Bd. 46, pp. 389-422, 01 1 2002.

[27] J. Kim, J. R. Wozniak, B. A. Mueller, X. Shen und W. Pan, „Comparison of statistical tests for group differences in brain functional networks," *NeuroImage,* Bd. 101, pp. 681-694, 2014.

[28] L. Breiman, „Random Forests," *Machine Learning,* Bd. 45, pp. 5-32, 01 10 2001.

[29] J. H. Friedman, „Stochastic gradient boosting," *Computational Statistics & Data Analysis,* Bd. 38, pp. 367-378, 2002.

[30] M. B. A. McDermott, S. Wang, N. Marinsek, R. Ranganath, M. Ghassemi und L. Foschini, „Reproducibility in Machine Learning for Health," in *Reproducibility in Machine Learning, ICLR 2019 Workshop, New Orleans, Louisiana, United States, May 6, 2019,* 2019.

[31] B. Grüning, J. Chilton, J. Köster, R. Dale, N. Soranzo, M. Beek, J. Goecks, R. Backofen, A. Nekrutenko und J. Taylor, „Practical Computational Reproducibility in the Life Sciences," *Cell Systems,* Bd. 6, pp. 631-635, 6 2018.

[32] M. Hutson, „Artificial intelligence faces reproducibility crisis," *Science,* Bd. 359, pp. 725-726, 2018.

[33] S. N. Goodman, D. Fanelli und J. P. A. Ioannidis, „What does research reproducibility mean?," *Science translational medicine,* Bd. 8, pp. 341ps12--341ps12, 2016.

[34] R. Tatman, J. VanderPlas und S. Dane, *A Practical Taxonomy of Reproducibility for Machine Learning Research,* 2018.

[35] T. Glatard, R. Da Silva, N. Boujelben, R. Adalat, N. Beck, P. Rioux, M.-E. Rousseau, E. Deelman und A. Evans, „Boutiques: an application-sharing system based on Linux containers.," *Frontiers in Neuroscience,* Bd. 9, 2015.

[36] L. Field, S. Suhr, J. Ison, W. Los, P. Wittenburg, D. Broeder, A. Hardisty, S. Repo und A. Jenkinson, *Realising the full potential of research data: common challenges in data management, sharing and integration across scientific disciplines,* Zenodo, 2013.

[37] D. S. Marcus, T. R. Olsen, M. Ramaratnam und R. L. Buckner, „The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data," *Neuroinformatics,* Bd. 5, pp. 11-34, 2007.

[38] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Lei, E. Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao und B. Mons, „The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data,* Bd. 3, p. 160018, 3 2016.

[39] N. Wilkins-Diehr, „Special Issue: Science Gateways—Common Community Interfaces to Grid Resources," *Concurrency Computat.: Pract. Exper.,* Bd. 19, pp. 743-749, 2007.

[40] S. Shahand, A. Benabdelkader, M. M. Jaghoori, M. a. Mourabit, J. Huguet, M. W. A. Caan, A. H. C. Kampen and S. D. Olabarriaga, "A data-centric neuroscience gateway:

design, implementation, and experiences," *Concurrency and Computation: Practice and Experience,* 4 2014.

[41] R. Ferreira Da Silva, T. Glatard und F. Desprez, „Self-Management of Operational Issues for Grid Computing: The Case of the Virtual Imaging Platform," in *Emerging Research in Cloud Distributed Computing Systems,* Bd. Chapitre 6, 2015, pp. 187-221.

[42] D. Krefting, S. Canisius, A. Hoheisel, H. Loose, T. Tolxdorff und T. Penzel, „Grid based Sleep Research — Analysis of Polysomnographies using a Grid infrastructure," *Future Generation Computer Systems,* Bd. 29, pp. 1671-1679, 2013.

[43] T. Simko, L. Heinrich, H. Hirvonsalo, D. Kousidis und D. Rodríguez, „REANA: A System for Reusable Research Data Analyses," *EPJ Web Conf.,* Bd. 214, p. 06034, 2019.

[44] B. K. Beaulieu-Jones und C. S. Greene, „Reproducibility of computational workflows is automated using continuous analysis," *Nat. Biotechnol.,* Bd. 35, pp. 342-346, 4 2017.

[45] R. Bekkerman, M. Bilenko und J. Langford, Scaling up machine learning: Parallel and distributed approaches, Cambridge University Press, 2011.

[46] C. Auffray, Z. Chen und L. Hood, „Systems medicine: the future of medical genomics and healthcare," *Genome Medicine,* Bd. 1, p. 2, 2009.

[47] A. C. Ahn, M. Tewari, C.-S. Poon und R. S. Phillips, „The Limits of Reductionism in Medicine: Could Systems Biology Offer an Alternative?," *PLoS Medicine,* Bd. 3, p. e208, 5 2006.

[48] P. C. Ivanov, K. K. L. Liu und R. P. Bartsch, „Focus on the emerging new fields of network physiology and network medicine," *New Journal of Physics,* Bd. 18, p. 100201, 2016.

[49] J. K. Aronson, „Biomarkers and surrogate endpoints," *British journal of clinical pharmacology,* Bd. 59, pp. 491-494, 2005.

[50] R. Heinzer, S. Vat, P. Marques-Vidal, H. Marti-Soler, D. Andries, N. Tobback, V. Mooser, M. Preisig, A. Malhotra, G. Waeber, P. Vollenweider, M. Tafti und J. Haba-Rubio, „Prevalence of sleep-disordered breathing in the general population: the HypnoLaus study," *The Lancet Respiratory Medicine,* Bd. 3, pp. 310-318, 2015.

[51] M. A. Carskadon, W. C. Dement und others, „Normal human sleep: an overview," in *Principles and practice of sleep medicine,* 6 Hrsg., Elsevier, 2017, pp. 15-24.

[52] V. K. Kapur, D. H. Auckley, S. Chowdhuri, D. C. Kuhlmann, R. Mehra, K. Ramar und C. G. Harrod, „Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline," *Journal of Clinical Sleep Medicine,* Bd. 13, pp. 479-504, 2017.

[53] P. Lévy, M. Kohler, W. T. McNicholas, F. Barbé, R. D. McEvoy, V. K. Somers, L. Lavie und J.-L. Pépin, „Obstructive sleep apnoea syndrome," *Nature reviews Disease primers,* Bd. 1, p. 15015, 2015.

[54] Y. Subramani, M. Singh, J. Wong, C. A. Kushida, A. Malhotra und F. Chung, „Understanding phenotypes of obstructive sleep apnea: applications in anesthesia, surgery, and perioperative medicine," *Anesthesia and analgesia,* Bd. 124, p. 179, 2017.

[55] M. Riedl, A. Müller, J. F. Kraemer, T. Penzel, J. Kurths and N. Wessel, "Cardio-respiratory coordination increases during sleep apnea," *PloS One,* vol. 9, p. e93866, 2014.

[56] H. P. Landolt and A. A. Borbély, "Age-dependent changes in sleep EEG topography," *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology,* vol. 112, pp. 369-377, 2 2001.

[57] A. Porta, L. Faes, V. Bari, A. Marchi, T. Bassani, G. Nollo, N. M. Perseguini, J. Milan, V. Minatel, A. Borghi-Silva, A. C. M. Takahashi and A. M. Catai, "Effect of age on complexity and causality of the cardiovascular control: comparison between model-based and model-free approaches," *PloS One,* vol. 9, p. e89463, 2014.

[58] S. Nemati, B. A. Edwards, J. Lee, B. Pittman-Polletta, J. P. Butler and A. Malhotra, "Respiration and heart rate complexity: effects of age and gender assessed by band-limited transfer entropy," *Respiratory Physiology & Neurobiology,* vol. 189, pp. 27-33, 10 2013.

[59] G. Klösch, B. Kemp, T. Penzel, A. Schlögl, P. Rappelsberger, E. Trenker, G. Gruber, J. Zeitlhofer, B. Saletu, W. M. Herrmann, S. L. Himanen, D. Kunz, M. J. Barbanoj, J. Röschke, A. Värri und G. Dorffner, „The SIESTA project polygraphic and clinical database," *IEEE Eng Med Biol Mag,* Bd. 20, pp. 51-57, 2001.

[60] I. G. Campbell, „EEG recording and analysis for sleep research," *Current protocols in neuroscience,* Bd. 49, pp. 10-2, 2009.

[61] A. Rechtschaffen und A. Kales, A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects., Bethesda, Md: U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968.

[62] C. Iber, S. Ancoli-Israel, A. L. J. Chesson und S. F. Quan, „The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. 1st Edition," 2007.

[63] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, R. M. Lloyd, C. L. Marcus, B. V. Vaughn und American Academy of Sleep Medicine, „The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Version 2.4.," 2017.

[64] B. Kemp, A. Värri, A. C. Rosa, K. D. Nielsen und J. Gade, „A simple format for exchange of digitized polygraphic recordings," *Electroencephalography and clinical neurophysiology,* Bd. 82, pp. 391-393, 1992.

[65] B. Kemp und J. Olivan, „European data format `plus' (EDF+), an EDF alike standard format for the exchange of physiological data," *Clinical Neurophysiology,* Bd. 114, pp. 1755-1761, 9 2003.

[66] A. Bashan, R. P. Bartsch, J. W. Kantelhardt, S. Havlin and P. C. Ivanov, "Network physiology reveals relations between network topology and physiological function," *Nature Communications,* vol. 3, p. 702, 2012.

[67] R. P. Bartsch, K. K. L. Liu, A. Bashan und P. C. Ivanov, „Network Physiology: How Organ Systems Dynamically Interact," *PLOS ONE,* Bd. 10, p. e0142143, 11 2015.

[68] P. Amstutz, M. R. Crusoe, N. Tijanić, B. Chapman, J. Chilton, M. Heuer, A. Kartashov, D. Leehr, H. Ménager, M. Nedeljkovich und others, *Common Workflow Language, v1.0,* 2016.

[69] D. P. Kingma und J. Ba, „Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings,* 2015.

[70] V. Nair und G. E. Hinton, „Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel,* 2010.

[71] J. Han und C. Moraga, „The Influence of the Sigmoid Function Parameters on the Speed of Backpropagation Learning," in *From Natural to Artificial Neural Computation, International Workshop on Artificial Neural Networks, IWANN '95, Malaga-Torremolinos, Spain, June 7-9, 1995, Proceedings,* 1995.

[72] J. S. Bridle, „Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition," in *Neurocomputing*, Berlin, 1990.

[73] J. S. Bergstra, R. Bardenet, Y. Bengio und B. Kégl, „Algorithms for Hyper-Parameter Optimization," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira und K. Q. Weinberger, Hrsg., Curran Associates, Inc., 2011, pp. 2546-2554.

[74] J. M. Bland und D. G. Altman, „Statistics notes: matching," *BMJ*, Bd. 309, p. 1128, 1994.

[75] R. Kohavi, „A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal Québec, Canada, August 20-25 1995, 2 Volumes*, 1995.

[76] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever und R. Salakhutdinov, „Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, Bd. 15, pp. 1929-1958, 2014.

[77] J. Friedman, T. Hastie und R. Tibshirani, The elements of statistical learning, 2 Hrsg., Springer Series in Statistics, 2009, pp. 241-245.

[78] J. L. Fleiss, „Measuring nominal scale agreement among many raters.," *Psychological bulletin*, Bd. 76, p. 378, 1971.

[79] J. Cohen, „A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, Bd. 20, pp. 37-46, 1960.

[80] M. Beier, C. Jansen, G. Mayer, T. Penzel, A. Rodenbeck, R. Siewert, M. Witt, J. Wu and D. Krefting, "Multicenter data sharing for collaboration in sleep medicine," *Future Generation Computer Systems*, vol. 67, pp. 466-480, 2 2017.

[81] C. Jansen, M. Beier, M. Witt, J. Wu und D. Krefting, „Extending XNAT towards a Cloud-based Quality Assessment Platform for Retinal Optical Coherence Tomographies," *Scalable Computing: Practice and Experience*, Bd. 16, 7 2015.

[82] C. Jansen, M. Witt und D. Krefting, „Employing Docker Swarm on OpenStack for Biomedical Analysis," in *Computational Science and Its Applications : ICCSA 2016: 16th International Conference, Beijing, China, July 4-7, 2016, Proceedings, Part II*, O. Gervasi, B. Murgante, S. Misra, A. C. A. M. Rocha, M. C. Torre, D. Taniar, O. B. Apduhan, E. Stankova und S. Wang, Hrsg., Cham, Springer International Publishing, 2016, pp. 303-318.

[83] C. Jansen, M. Beier, M. Witt, S. Frey und D. Krefting, „Towards Reproducible Research in a Biomedical Collaboration Platform following the FAIR Guiding Principles," in *Companion Proceedings of the10th International Conference on Utility and Cloud Computing - UCC 17 Companion*, 2017.

[84] M. Witt, C. Jansen, S. Breuer, M. Beier und D. Krefting, „Artefakterkennung über eine cloud-basierte Plattform," *Somnologie*, Bd. 21, pp. 311-318, 01 12 2017.

[85] C. Jansen, B. Schilling, K. Strohmenger, M. Witt, J. Annuscheit und D. Krefting, „Reproducibility and Performance of Deep Learning Applications for Cancer Detection in Pathological Images," in *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, 2019.

[86] D. Krefting, C. Jansen, T. Penzel, F. Han and J. Kantelhardt, "Age and gender dependency of physiological networks in sleep," *Physiological Measurement*, 2017.

[87] C. Jansen, S. Hodel, T. Penzel, M. Spott und D. Krefting, „Feature relevance in physiological networks for classification of obstructive sleep apnea," *Physiological Measurement*, Bd. 39, p. 124003, 2018.

[88] R. Fielding, „Representational state transfer (REST). Chapter 5 in Architectural Styles and the Design of Networkbased Software Architectures,“ 2000.

[89] B. K. R. Vangoor, V. Tarasov und E. Zadok, „To FUSE or Not to FUSE: Performance of User-Space File Systems,“ in *15th USENIX Conference on File and Storage Technologies (FAST 17)*, Santa, 2017.

11 Eidesstattliche Versicherung

„Ich, Christoph Jansen, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: Towards Interpretability and Reproducibility of Machine Learning Applications in Clinical Research – Assessment of the Physiological Network in Sleep Apnea selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.
Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Betreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass mir die Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis bekannt ist und ich mich zur Einhaltung dieser Satzung verpflichte.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum _____ Unterschrift

12 Anteilserklärung an den erfolgten Publikationen

Christoph Jansen hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Maximilian Beier, Christoph Jansen, Geert Mayer, Thomas Penzel, Andrea Rodenbeck, René Siewert, Michael Witt, Jie Wu und Dagmar Krefting, Multicenter data sharing for collaboration in sleep medicine, Future Generation Computer Systems, 2017
Beitrag im Einzelnen: Beteiligt an Idee, Design und Implementierung der Kollaborationsplattform; Hauptverantwortlich für die Implementierung der Back-End-Infrastruktur der Plattform und des Job Management Service; Beteiligung am gesamten Publikationsprozess, inklusive des Entwurfs, der Berichtigungen nach Gutachterbewertung und der Finalisierung des Manuskripts; Abbildung 3 erstellt; Durchführung der Messungen für Abbildung 10; Beschreibung von OpenStack und Docker im Text

Publikation 2: Dagmar Krefting, Christoph Jansen, Thomas Penzel, Fang Han und Jan W. Kantelhardt, Age and gender dependency of physiological networks in sleep, Physiological Measurement, 2017
Beitrag im Einzelnen: Beteiligt an Idee, Design, Durchführung und Auswertung der Experimente; Hauptverantwortlich für alle ANN-basierten Experimente; Beteiligung am gesamten Publikationsprozess, inklusive des Entwurfs, der Berichtigungen nach Gutachterbewertung und der Finalisierung des Manuskripts; Beschreibung der ANN-basierten Methoden und Ergebnisse im Text; Abbildung 6 erstellt

Publikation 3: Christoph Jansen, Stephan Hodel, Thomas Penzel, Martin Spott und Dagmar Krefting, Feature relevance in physiological networks for classification of obstructive sleep apnea, Physiological Measurement, 2018
Beitrag im Einzelnen: Hauptverantwortlich für Idee, Design und Auswertung aller Experimente; Durchführung der ANN-basierten Experimente; Vergleich der ANN-Ergebnisse mit den Methoden der Co-Autoren (Decision Trees, Random Forests und statistische Tests); Hauptverantwortlich für den gesamten Publikationsprozess, inklusive des Entwurfs, der Berichtigungen nach Gutachterbewertung und der Finalisierung des Manuskripts; Verfassen großer Teile des Textes, der durch die Co-Autoren in den Bereichen der Methoden (Statistik, Decision Trees, Random Forests, TDS), des medizinischen Hintergrundwissens und der medizinischen Einordnung ergänzt wurde; Tabellen 2 und 3, sowie Abbildungen 1 und 3 erstellt

_____
Unterschrift, Datum und Stempel des betreuenden Hochschullehrers/der betreuenden Hochschullehrerin

_____
Unterschrift des Doktoranden/der Doktorandin

13 Publikation: Multicenter Data Sharing for Collaboration in Sleep Medicine

M. Beier, C. Jansen, G. Mayer, T. Penzel, A. Rodenbeck, R. Siewert, M. Witt, J. Wu and D. Krefting, "Multicenter data sharing for collaboration in sleep medicine," *Future Generation Computer Systems*, vol. 67, pp. 466-480, 2 2017.

https://doi.org/10.1016/j.future.2016.03.025

Seiten in der Dissertation: 35 - 49

14 Publikation: Age and Gender Dependency of Physiological Networks in Sleep

D. Krefting, C. Jansen, T. Penzel, F. Han and J. W. Kantelhardt, "Age and gender dependency of physiological networks in sleep," *Physiological Measurement,* 2017.

https://doi.org/10.1088/1361-6579/aa614e

Seiten in der Dissertation: 51 - 67

15 Publikation: Feature Relevance in Physiological Networks for Classification of Obstructive Sleep Apnea

C. Jansen, S. Hodel, T. Penzel, M. Spott and D. Krefting, "Feature relevance in physiological networks for classification of obstructive sleep apnea," *Physiological Measurement*, vol. 39, p. 124003, 2018.

https://doi.org/10.1088/1361-6579/aaf0c9

Seiten in der Dissertation: 69 - 78

16 Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

16.1 Persönliche Angaben

16.2 Bildung

16.3 Arbeitserfahrung

16.4 Sprachen

17 Publikationsliste

17.1 Ausgewählte Publikationen

- M. Beier, **C. Jansen**, G. Mayer, T. Penzel, A. Rodenbeck, R. Siewert, M. Witt, J. Wu and D. Krefting, "Multicenter data sharing for collaboration in sleep medicine," *Future Generation Computer Systems,* vol. 67, pp. 466-480, 2 2017. Journal Impact Factor 2018: **5.768**
- D. Krefting, **C. Jansen**, T. Penzel, F. Han and J. W. Kantelhardt, "Age and gender dependency of physiological networks in sleep," *Physiological Measurement,* 2017. Journal Impact Factor 2018: **2.246**
- **C. Jansen**, S. Hodel, T. Penzel, M. Spott and D. Krefting, "Feature relevance in physiological networks for classification of obstructive sleep apnea," *Physiological Measurement,* vol. 39, p. 124003, 2018. Journal Impact Factor 2018: **2.246**

17.2 Zusätzliche Publikationen

- **C. Jansen**, M. Beier, M. Witt, J. Wu and D. Krefting, "Extending XNAT towards a Cloud-based Quality Assessment Platform for Retinal Optical Coherence Tomographies," *Scalable Computing: Practice and Experience,* vol. 16, 7 2015.
- **C. Jansen**, M. Witt and D. Krefting, "Employing Docker Swarm on OpenStack for Biomedical Analysis," in *Computational Science and Its Applications : ICCSA 2016: 16th International Conference, Beijing, China, July 4-7, 2016, Proceedings, Part II*, O. Gervasi, B. Murgante, S. Misra, A. C. A. M. Rocha, M. C. Torre, D. Taniar, O. B. Apduhan, E. Stankova and S. Wang, Eds., Cham, Springer International Publishing, 2016, pp. 303-318.
- **C. Jansen**, M. Beier, M. Witt, S. Frey and D. Krefting, "Towards Reproducible Research in a Biomedical Collaboration Platform following the FAIR Guiding Principles," in *Companion Proceedings of the10th International Conference on Utility and Cloud Computing - UCC 17 Companion*, 2017.
- M. Witt, **C. Jansen**, S. Breuer, M. Beier and D. Krefting, "Artefakterkennung über eine cloud-basierte Plattform," *Somnologie,* vol. 21, pp. 311-318, 01 12 2017.
- **C. Jansen**, B. Schilling, K. Strohmenger, M. Witt, J. Annuscheit and D. Krefting, "Reproducibility and Performance of Deep Learning Applications for Cancer Detection in Pathological Images," in *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, 2019.

18 Danksagung

Ich danke Herrn Prof. Dr. Thomas Penzel für die Möglichkeit meine Promotion an der Charité – Universitätsmedizin Berlin durchführen zu können, für die Betreuung der Promotionsarbeit und für die Unterstützung und Expertise in medizinischen Fragen, die für die Durchführung der Arbeit essentiell waren.

Mein besonderer Dank gilt Frau Prof. Dr. Dagmar Krefting, da sie mich schon während meines Studiums an der HTW Berlin in ein wissenschaftliches Projekt und gemeinsame Veröffentlichungen miteinbezog. Mit der anschließenden Anstellung als Wissenschaftlicher Mitarbeiter eröffnete sie mir einen beruflichen Werdegang in einem fordernden und innovativen Umfeld. Ich danke ihr, dass sie mich als Mentorin gefördert und in meinem Promotionsvorhaben angespornt hat. Nur durch ihr Engagement konnte ich dieses Vorhaben erfolgreich umsetzen.

Ich danke meinen ehemaligen Kolleginnen und Kollegen an der HTW Berlin, sowie den Co-Autoren anderer Institute für eine großartige Arbeitsumgebung und viele erfolgreiche Kooperationen.

Ich danke meinen Freunden, meiner Familie und meiner Lebensgefährtin Verena Schmitt für ihre liebevolle Unterstützung und ihren Rückhalt.