Pilot 2: Kidney Disease

1. Key Information

1.1 Involved Partners

- Charité - Universitätsmedizin Berlin (Charité)
- German Research Center for Artificial Intelligence (DFKI)
- AOK Nordost (AOK)
- Universitätsklinikum Essen (Essen)
- Hasso Plattner Institute (HPI)

1.2 Involved Countries

- Germany

1.3 Keywords

- Telemedicine
- Adherence monitoring
- Risk prediction
- Prediction of rejection, graft loss, infections
- (Complex event processing)

1.4 Task Description

The aim of this pilot is to support clinical staff in the kidney transplant center which treats patients before and after a kidney transplantation. Main data sources are the TBase system, an electronic medical record (EMR) system as well as a modular data storage system, containing patient data of the last 20 years, and a patient app which provides remote vital parameters and daily drug intake of the patients. Using this data, the pilot addresses the following tasks: a) establishing a telemedicine approach, b) adherence monitoring, and c) risk predictions. All tasks aim to detect critical patients to reduce complications (thereby reduce hospitalizations), reduce costs, and improve the quality of life of the patients.

2. Building Blocks

2.1 Architecture

The digital infrastructure of the medical department of nephrology at Charité contains the TBase, an app, the necessary interfaces and a security concept.

2.1.1 System Architecture

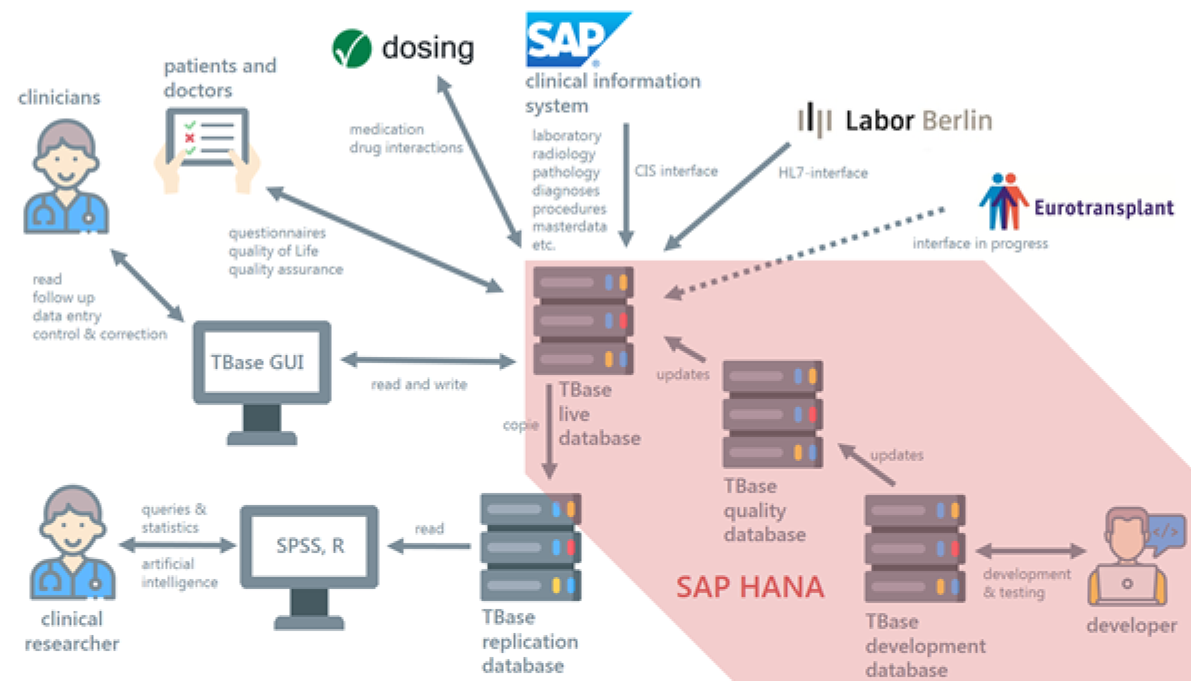


Figure1: Overview of architecture

Data Access: Different services are able to access the data, such as the Graphical User Interface (GUI) or an analytical component which is responsible for a statistical analysis. Moreover, a data cleansing component runs on the patient data in order to provide clean training data for the risk prediction models. Patient data from the hospital and mobile app is processed via Apache Flink and stored within a SAP HANA platform. The data can be accessed via authorization within the hospital only.

Analytics: The analytics component is a backend service to provide a statistical analysis for particular patients or particular patient groups. The service can be triggered via GUI. The components are implemented in python and rely mostly on methods provided by Scikit-learn.

User Interface: Our GUI has several functionalities: a) visualizing patient data, b) visualizing critical patients within a dashboard (ranking patient), and c) an analysis of single patients in context of a risk prediction score. The GUI is a web-based front-end, which sends requests via REST to the backend functionalities.

2.1.2 Data Flow & Interoperability of services

The digital infrastructure of the working group "Digital Nephrology '' at Charité includes an Applicationg Programming Interface (API) between the EMR TBase and the hospital information system of Charité - Universitätsmedizin Berlin (Charité). The following data are provided continuously:

- Patient master data
- Laboratory values
- All hospital stays
- Treatments
- Diagnostic results

In addition, the following APIs to software or databases are part of the digital infrastructure:

AiDKlinik by Dosing GmbH Heidelberg: Information on drug-drug interactions, which the MIA SERVICE® provides everytime when a drug is changed (deleted, newly prescribed, or the dosage changed).

LaborBerlin: LaborBerlin provides via API (Health Level 7, HL7; Version 2, V2) laboratory values, authentication by OAuth2. TBase extracts some laboratory results by the above mentioned API between TBase and the Charité hospital information system.

Eurotransplant: Currently, the required information from Eurotransplant has to be entered manually into TBase by physicians, nurses, and secretary staff. It is planned that data on the patient about Human Leukocyte Antigens (HLA) and the transplantable status should be shared automatically via a future API between TBase and Eurotransplant.

Home-measured data of patients after kidney transplantation: A smartphone app, called EASY app (by comjoodoc business solutions GmbH (comjoo)) sends data via HL7 Fast Healthcare Interoperability Resources (FHIR) standard, which are stored on a FHIR server (by the company medworxs.io), which is also called MACCS ("Medical Assistant for Chronic Care Services")-platform. TBase downloads this data once per minute. The identification of single values of the data takes place via a patient ID known by Charité. Personal data are currently not transferred, perspectively and in order to provide medical letter, chat data, and test results, personal data should be pseudonymised. Currently, only home-measured vital signs, blood sugar, well-being and the actual intake behaviour of medication is provided. In return, the patient receives from TBase via the EASY-app updated medication plans and laboratory values.

Nephro7: The home-measured data are also transmitted to the EMR of nephrologists in private practice, called Nephro7, which is programmed and hosted by Medvision AG. Nephro7 is connected using the same infrastructure described above via the HL7 FHIR standard, the FHIR server, and TBase.

There is also the possibility to enter vital signs, laboratory values, and medical results (e.g. biopsies) into TBase manually.

2.1.3 Necessary Hardware

All outpatient care centers of the Medical Departments at Charité (e.g. Nephrology, Neurology) use the same hardware and same resources, therefore only statements on the whole server system can be made:
HD: 1TB, ~270GB used
RAM: 750TB, ~310GB used
CPU: unknown, estimated >6 cores, MHz not known

Additional items for prediction model training and analytics with Apache Flink :
1 x MONTAGE UND KONFIGURATION
2 x Zotac 11GB D5X GTX 1080 Ti AMP
1 x Asus TUF X299 MARK 1
1 x Intel Core i9-7900X 3300 2066 BOX
1 x be quiet! Shadow Rock 2 1 x be quiet! Dark P. Pro P11 1200W
1 x SSD 1TB 1.9/3.2G 960 EVO PCIe M.2 SAM
1 x D464GB 3000-15 Predator
1 x WD 6TB WD60EFRX Red SA3
1 x be quiet! PURE BASE 600 srbk

2.1.4 Software Components

The software components of TBASE include:

Operating system: Virtual Machine (VM) SUSE Linux Enterprise 15, SAP HANA Multi Tenant Installation (HANA 2 mit SPS05 und XSA)
SAP HANA XSA - (XS advanced): Application level at which the app runs

Database: SAP HANA

Backend: Web Server with NodeJs including Node Packages, which are easy to access (npmjs.com)

Frontend: SAPUI5, JobScheduler Dashboard to configure the data regularly

2.3 Data Processing

2.3.1 Processing of large structured / unstructured data sources

2.3.1.1 Data Sources

The data stream of patient data is stored in TBase via various APIs, amongst others LaborBerlin and the hospital information system of Charité. But also the documentation of routine care of doctor visits, every call of patients to the outpatient care center is entered (e.g. change of appointment, need for prescriptions), and the home-measured data.

Data Source	**Description**	**Acquisition**	**Characteristic (Size, Patients, Years, Origin/Region)**
TBase	Patient data of kidney transplant recipients that were legitimately collected, are stored in the patient documentation system of the Berlin Kidney Transplant Centre, called TBase. TBase was designed in the 90´s and since then the patient data is collected. The access to the data is granted for the scientific and medical staff of the Department of Nephrology and Internal Intensive Care Medicine. It should be pointed out that the partners of this pilot considered the European Regulation 2016/679. Data includes structured and unstructured data and is stored in a RDBMS.	TBase was developed in the 90´s in cooperation with the Humboldt University of Berlin. In 2019 a second release of TBase in SAP HANA language took place including a telemedicine dashboard	28500 (donors, recipients (3100), deceased patients, waiting list patients, and those with rare kidney disease, but no transplantation)
MACCS	Stands for “Medical Assistant for Chronic Care Service” and is a digital generic service to improve regular treatment of patients after kidney transplantation. It includes an app which provides remote vital signs, remote well-being, regularly updated medication plan which can be clicked off, alert function and the possibility	The cooperation of Charité, comjoo, and Medvision AG a self programmed platform to share patient information (called MACCS platform). The project had been funded publicly by German Federal Ministry for Economy	In February 2021 ca. 330 patients after kidney transplantation are included and supported by a telemedicine team. The project started in February 2020 in the Kidney Transplantation Center of Charité, Berlin.

	to exchange messages and video consultation with medical staff. A telemedicine team evaluates the incoming data and takes action if necessary. The data are visualized in the telemedicine dashboard of the TBase.	and Energy (BMWi). The cooperation started in 2018. This project builds on a prior project, also publicly funded by BMWi, which ran from 2016 to 2019 and aimed to build up a secure platform to share pseudonymised patient data between several different EMRs.	

Pilot	**Multiple sources**	**Integration to data warehouse**	**Data access**	**Data stored in cloud**	**Multi-party architecture**	**Secure environment**	**Transform raw / unstructured data**
2	yes	no	Data is processed within hospital	no	no	yes	yes

2.3.1.2 De-Identification and anonymisation

In most cases data does not need to be de-identified and anonymised, as it always stays within the hospital and is processed by their staff. In case of risk prediction, the model is trained on a copy of TBase which does not include information such as name or address. In the context of training models for natural language processing, a smaller subset of text was extracted, and patient related information removed, such as reference to patient ID, patient name, address etc. This has been done with a commercial tool in a previous project (Averbis GmbH).

2.3.1.3 Acquisition

TBase is a relational database, based on SQL. To explore and extract the data for the risk prediction model, we use a python environment to connect with the database and apply queries to create a target cohort. This cohort needs to be created in a way so that later the machine learning method can easily and fastly read in the data. Therefore, we came up with a JSON format which unifies the patient related information across the multiple tables in the database.

Our JSON cohort is constructed in a way to address the needs of our problem: In the risk prediction we would like to predict each time new patient information is being stored how likely it is that a particular endpoint will occur within a certain time frame. Therefore, the JSON cohort includes multiple lines per patient, each time new data is logged in visit, hospitalization, and laboratory values, a new line is created. Each of those lines include all relevant information of that given patient of the last year, up to this given date. Rather static information such as demographics or medications (up to the given date) are of course always present.

2.3.1.4 Cleansing

TBase includes noisy and missing information (examples given below). We use a python script in order to gather the data and create the cohort, which is then used to train the risk model. Before that we apply a large range of cleaning steps. The script for cleaning the data is built in such a way that multiple configuration options of how data is cleaned (e.g. applying imputation of missing values), can be turned on and off.

Examples of noisy data:

1) Fields, which are actually supposed to store a number, store a string. Therefore, the field contains much noise, such as a) value including unit - although the unit is supposed to be stored somewhere else, b) or just other information, the doctor thinks it might be relevant.

2) For each visit (regular checkup) the database includes a table for 5 different vital parameters (e.g. weight, blood pressure). However, those fields are not always filled. Only those parameters are recorded for which the treating physician thinks it is currently important to measure it.

3) The part of the database, which stores the regular visits, can be als “misused”, to store a general note about a patient even though the patient was not in the hospital. This means that this table (was used in the past) is also used if there was a phone call with the patient or a different treating physician (e.g. GP). This can result in multiple problems. For instance in many cases various vitals are empty, and for the risk prediction we need to see how to deal with that. Moreover, it can happen that we have entries in the visit table, even though the patient has already passed away.

2.3.1.5 Data Integration

Data of MACCS is stored together with the regular EMR data in the TBase system. Historically, TBase did not include telemedicine data. However, during the BigMedilytics project this database has been updated and extended so that this external data can be also stored.

2.3.1 Multi-velocity processing of heterogeneous data streams

Does not apply

2.3.5 Complex real-time event detection

The prototype for complex event processing (CEP) aims to monitor patient adherence by analyzing the stream data produced by them remotely through web applications. The data is formed mainly of the home-measured vital signs. The system consists of a stream processing pipeline based on the data collected and stored in the TBase database from the Charité. The system is implemented using the Apache Flink framework for Big Data analysis and makes use of the CEP library provided by Flink. The streaming communications is handled by the message broker Apache Kafka and the output data analysis is stored in the time series database InfluxDB.

2.3.5.1 Notifications

The prototype was designed to be closely connected to the MACCS-platform for telemedicine. This platform is partially located in the Charité servers. When the stream analysis process takes place, the output will be sent in real-time to the SAP web application and displayed in the UI for the physicians. The notifications are provided in 3 categories: normal, warning, and alarm status.

2.3.5.2 Situations of Interest

The CEP model is designed to analyze whether an event is trespassing a given threshold (simple event), plus the development of this specific event in the timeline (complex events). For chronic patients, it is important to monitor the vital signs along some specific time range, as well as the medication intake habits of the patient. The adherence behaviours of the chronic patients is crucial to avoid long term graft loss.

2.3.5.3 Event Processing

The CEP model is following simple complex event rules (the so called patterns) and algorithms triggered in case a complex event pattern matches. Patterns based on the Apache Flink Pattern API that allows us to define complex pattern sequences. We implement in real time selection, projection, aggregations and joins. The implementation of watermarks and windowing for event consistency has been applied. The timestamp used to keep consistency is based on event processing time instead of event time.

2.3.5.4 Event Sources

The data to analyse with CEP is a combination of data from TBase (e.g. gender, age, time since transplantation) and from the EASY-app (vital signs, intake protocol).

2.3.5.5 Evaluation

The data has been collected from February 2020 onwards. After one year, the number of patients contributing to the application reached more than 400. This number has been increasing throughout the time, starting from a few patients in the beginning during the beginning of the pandemic. The patients enter data roughly one time a day.

The contribution of data for each patient is different depending on the vital sign category. For instance, for pulse-monitoring, the number of patients contributing is approximately 400, but for blood sugar monitoring only 100 patients are contributing. Something similar happens with the other signals (blood pressure, body weight, etc..), but the numbers are more alike.

	critical	suspicious	normal	suspicious	critical
Systolic blood pressure	<90mmHg	<100mmHg	100 - 129mmHg	130 - 180mmHg	>180mmHg
Diastolic blood pressure	<50mmHg	50 - 59mmHg	60 - 89mmHg	90 - 100mmHg	>100mmHg
Heart rate	<50bpm	50 - 59bpm	60 - 89bpm	90 - 120bpm	>120bpm
Temperature	<33.5°C	33,5 -36,2°C	36,3 - 37,4°C	37.5 - 38,0°C	>38.0°C
Change in weight 1 day	>(-1,5)kg	(-1,5) - (-0,5)kg	+/-0.5kg	0.5 - 1.5kg	>1,5kg
Change in weight 3 days	>(-2,5)kg	(-2,5) - (-1,0)kg	+/-1.0kg	1.0 - 2.5kg	>2.5kg
Change in weight 8 days	>(-3.0)kg	(-3,0) - (-1,5)kg		1.5 - 3.0kg	>3.0kg
Well-being	Very good (5) - green	Good (4) – light green	Medium (3) - yellow	bad (2) - orange	Very bad (1) - red
Medication adherence	<70%	70 – 89%	>=90%	/	/
Therapy adherence	<70%	70 – 89%	>=90%	/	/

Figure 2: Ranges für vital signs

Although the CEP system is realized as an online streaming application, due to a missing gateway for real-time transfer from the database, the experiment has been simulated with a batch sample of data from a 12 months period. For the simulation, we have used a python application to query the TBase and generate a streaming pipeline with a Kafka producer and a Flink system that consumes and processes the data by applying the CEP algorithms. The final output is stored in an Influx time-series database.

The preliminary output analysis contains information related with the warnings and alarms that have been triggered if any of the input data doesn't perform as expected by the thresholds designed by the doctors. Around 5% of the output has triggered an alarm event, while 10% are warnings. It is important to highlight that these results are preliminary results and need further examinations. The behavior of the human body is complex and including more variables in the automatic monitoring process may result in an improved outcome of the system and thereby help the doctors better in taking decisions. For instance, it would be beneficial to include into the monitoring process the adherence behaviour of the patient to detect how the medication intake affects the vital parameters over time.

In conclusion, the roll-out of the EASY-app in combination with a live monitoring process is an important step into telemedicine care. The effort from the Charité and the different partners in the recent projects has made it already possible to connect patients remotely with the physician. But there is still a long path to go to acquire a reliable monitoring system. It is, for example, necessary to develop intelligent systems that can adapt each threshold in the monitoring to each specific patient, instead of having global thresholds for all patients. Also it will be necessary to perform experiments that could search for a correlation between the medicament intake behaviours with the vital signs collected by the patients.

2.4 AI Components

In the following sections the AI components applied in pilot 2: kidney disease are presented and discussed. Pilot 2 uses AI methods for Natural Language Processing and prediction. Image analytics is not part of pilot 2.

2.4.1 Deep learning for multilingual NLP and image analytics

2.4.1.1 Natural Language Processing

In order to access information in German clinical text, we apply methods of natural language processing (NLP). We mainly focus on the extraction of named entities (e.g. medical conditions, body part), detection of relations between them (e.g. symedical_condition <*is-located*> body_part), and the detection of negated medical conditions. The extraction of such information can serve as input for more complex applications, such as risk prediction.

Langua ge	**How will NLP support your pilot?**	**How will NLP help you to reduce costs?**
German	In electronic health records, text is an additional source of information which often includes other information that is not directly expressed in the structured data. Thus, it is also important to extract relevant information from text in order to get a better understanding of a patient.	The idea is to combine structured and unstructured information in the prediction models in the long-term. As text includes additional information we assume that models will provide better results using text. Improved results will also mean that patient outcomes can be refined and hospitalizations reduced. The reduction of hospitalizations reduces costs.

Which NLP	**Method**	**Software framewor**	**Vocabularies/c orpus used**	**Describe your method in a few**	**Describe your corpus/training**

tasks do you address?		ks		sentences.	data?
Named Entity Recognition (NER)	Bi-LSTM	Flair	Corpus of German nephrology reports (clinical notes & discharge summaries)	Character-level bidirectional LSTM which reads in training data and is trained for each single concept.	Fine grained annotation of 12 different concepts in German nephrology reports which include discharge summaries and clinical notes.
Relation Extraction	CNN	Tensorflow	Corpus of German nephrology reports (clinical notes & discharge summaries)	Convolutional neural network for binary relation extraction, which uses as input information beside words and related concepts, also the distance between the two concepts.	Annotated relations between concepts in German nephrology reports (see NER).
Negation Detection	rule-based	NegEx	Small manually annotated corpus of negations in German clinical reports and a modified set of negation triggers	Simple negation detection of medical conditions	No training required

One of the outcomes of the project is a toolbox to process German clinical text. The developed methods will be made available for free usage. While NLP models to process clinical text exist for English, they do not exist for German. Methods are generally language independent, which means, with the corresponding data, models can be easily also trained for other languages. More information can be found in our paper [1].

2.4.1.2 Image Processing

Does not apply.

2.4.2 Prediction Algorithms

One of the tasks of the Kidney Disease Pilot was the development of a risk prediction model targeting the forecast of infection, rejection and graft failure within the next 90 days. Our model was developed using retrospective TBase data, and then later evaluated within a study, in which we compared the performance of medical doctors with and without the support of our risk prediction system. The study showed that 1) the risk prediction system tends to outperform medical doctors according to AUC-ROC, 2) medical doctors did not necessarily improve their AUC-ROC performance together with the risk prediction system, only junior physicians did, and 3) medical doctors and risk prediction partially found different patients at risk.

2.4.2.1 Task

The task is to predict the risk of (1) an infection, (2) a rejection, and (3) a graft failure within the following 90 days since the last contact of a patient with the system. The aim ist to find patients at risk earlier than usual and support them with information (empowerment and self-assessment). This in turn may hopefully reduce hospitalizations, duration of the stay and intensity of diagnostics and therapy, which are often more expensive or reduce the quality of life of patients more than routine procedures (more often invasive).

2.4.2.2 Data, Data Modelling

To train our model we use retrospective data of TBase from Charité Mitte and Virchow Klinikum from the years 2008 - 2020. The data includes more than 1400 different patients with altogether more than 100000 data points (DP). The following data points have been filtered out beforehand: a) DPs of patients below 18 (at that point in time), b) if target endpoint is currently active, c) DP within a time frame of less than a week after an infection, and a rejection, d) DPs within a time frame of less than two weeks after a transplant, and e) DPs of patients which do not have a follow-up datapoint in the next 15 to 180 days. Those filter criterions have been defined together with medical experts.

Size: >100,000 data points
Patients: > 1400 patients

2.4.2.3 Features

The models uses retrospective data from TBase, particularly we use for the model the following data:

- Socio-demographics (e.g. gender, age, smoking status)
- Vitals (e.g. blood pressure, weight)
- Primary disease
- Medications
- Laboratory values
- Transplant/donor information
- Previous diagnoses
- Stays at hospital

Information does not come at high speed. Instead the information is updated infrequently at different times, depending on the current situation change of the patient, and comes in different formats. Some data like socio-demographics or transplant information are rather static.

Using those information new features are generated:

- Mean values of vitals and lab values within a certain time frame
- Indication of increase or decrease of values in comparison to previous values

Overall more than 1200 features are used for the prediction model. During the development of the model, also text data has been integrated. However, as the improvement was marginal at that point we excluded it for the retrospective study.

2.4.2.4 Model

The risk prediction is implemented in python using scikit-learn and uses Gradient Boosted Regression Trees (GBRT). It showed promising results and can be trained on less performant computer infrastructures within the hospital and thereby be calculated quickly. The model includes 300 trees and is used for retrospective study.

In parallel to that, we also explored other architectures using a neural architecture (see [2]) and evaluated it on different data (MIMIC-III), also showing promising results. However, as the manual evaluation took many months, we will report only results on Charité data, using the GBRT.

2.4.2.5 Evaluation

Description of Evaluation Setup: The evaluation is carried out as retrospective study and evaluated within a 5-fold cross validation. In addition to that we explore how well a medical doctor can solve the task of predicting one of the endpoints in the next 90 days. Then, we compare this to our method and examine if a doctor can achieve better results including the prediction of our model.

Size of Data: Data includes more than 1400 patients with overall more than 100,000 data points, and then split into 70% training, 15% development and 15% test data.

Evaluation Method: The model is evaluated using AUC-ROC.

Results: Graft Failure: ROC: 0.92, Rejection: 0.80, Infection: 0.80

Comparison to other related work: A comparison to other related work is difficult, as results might differ if cohort and data varies.

2.5 Security and privacy of data access and processing

TBase: The patient data are in general only accessible with a medical confidentiality agreement. Data scientists did not work on the productive TBase, however, they signed a confidentiality agreement. To make this possible, a separate replication server with extra access for data scientists had been set up in the past, where only the pseudonymized basic data were provided and no personal data of patients. As mentioned above, the completed prediction models are only conducted on TBase data (and not on home-measured data). Therefore, the information on security and privacy of data access and process are described for TBase.

The home-measured data: Patient data sent via EASY-app are transferred by HL7 FHIR standard. In addition, the patient data is transferred encrypted (PKI-infrastructure) and the data are also pseudonymised by using a secure user ID. The trust service provider (here: medworx.io) is responsible for the transit, including a pseudonymisation key. There is a consent management procedure within the EASY app implemented, which also allows patients to withdraw the consent at any time and thereby stop sending data. The FHIR server is only a transit server. The master data is eventually stored at Charité behind a firewall (central server) and in the EMR Nephro7 by MedVision AG. The data is only stored in the app (and not in the smartphone itself). The connection between the personal smartphone and the FHIR server is enabled pseudonymised and the full identity of patients is only known by the telemedicine team. Regarding the communication data, the master data is stored at comjoo.

Guest scientist contracts with Charité have been made with all colleagues from DFKI. The Team from DFKI received access to the patient documentation system of Berlin Kidney Transplantation Centre, called TBase, by 2-step authentication and they only can use an operational frontend without access to the pure patient data.

2.5.1 Access Control

TBase: The data is stored on a virtual machine of the IT business unit. Access to the server rooms is only permitted to registered persons. Access is logged by the security service.

Home-measured data: Medworx.io uses Keycloak to identify the user roles via single sign on (SSO).

2.5.1.1 Authentication

TBase: The database is secured by passwords. Access to the database is described as follows:

- User must sign privacy policy
- Patients must sign consent form
- SAP User Rights Management Tool: (responsible: Key-User of the Clinic)
- User rights levels: read, write, delete, administrators
- Users can be limited to patient group
- User passwords expire after 183 days (controlled via GB IT policy)

Home-measured data: Users identify themselves by login data received during the onboarding process into the system (see below).

2.5.1.2 Authorization

TBase: There are five database user roles for user management: (1) read, (2) write, (3) delete, (4) administrators, and (5) interfaces. These roles have correspondingly limited access to the data tables at the database level and correspondingly limited access to the web interface. The administrators of the TBase application have full access to the data and all functions on the web interface. In addition to reading, entering, modifying and deleting the data, the TBase admin can create, block or delete users via a separate mask, which only an admin can call up, and, if necessary, assign certain patients (restricted patient view) separately to a user. Users with change or deletion rights can call up "Edit" masks in addition to the display masks in order to be able to change or delete data. Users with read rights have read-only access and cannot change data. Interface users generally do not have access to the web interface, but only limited access to certain interface tables on the database page. Each change in each data table is automatically transferred to so-called "TBase-History" tables via triggers. This is completely identical to the TBase tables. In addition, each data table has the additional time, user, action columns that are continuously populated with the timestamp, database users, and database operation (insert, update, delete). Thus, all database operations are logged (audit trail). The data cannot be changed or deleted by anyone else except the system administrator ("superuser"). TBase is only accessible within the Charité intranet and not from outside (Internet). The FHIR server itself is passive and does not have access to other systems. Conversely, the comjoodoc EASY and MyTherapy apps, as well as the TBase and Nephro7 patient documentation system, can send requests to the FHIR server and upload or download data.

Home-measured data: While the patient is included into the system, the identity is confirmed by the telemedicine team. The staff initiates the digital inclusion into the system: TBase now asks medworx.io, if it knows the patient-ID. If not, medworx.io sends to TBase an empty user ID, which is filled out by TBase using the encrypted ID. Now the user exists and the patient has to login into the medworx.io system with the pseudonymised login data to set the consent in an medworx.io frontend. Now the connection between the EASY-app and FHIR transit is established.

2.5.2 Data Protection

Using the SAP HANA in-memory database platform, a secure database technology that protects data with strategies such as application-level authorization, SSO, MIT-Kerberos protocol, and Security Assertion Mark-up Language (SAML) is used. The platform secures communication, data storage, and application services using the latest encryption and testing techniques. All developments on the database (DB) are controlled by authorizations. This ensures the security of data by design at a high level. In addition, all data are kept behind the certified Charité firewall. In compliance with the latest European Union General Data Protection Regulation (EU

GDPR) a robust data protection concept was implemented, including data flow diagrams, data protection risk assessment (DSFA) and authorization concept. All documents are kept in a procedure directory of the Charité Data Protection Office.
In particular regarding the home-measured data, a legal opinion was sought, which concluded that the digital infrastructure containing the connection between TBase, EASY-app, and the APIs and the particular content, which is transmitted, adheres to GDPR.

2.5.2.1 Data at rest

See 2.5.2, data protection.

2.5.2.2 Data in transit

The transport processes are carried out by a secure standard encryption (https) or HL7 FHIR standard (medication) or comjoodoc (communication data).

2.5.3 Auditory and logs

2.5.3.1 System Auditory

System logs are recorded and can be viewed as required. Otherwise simple table structure is used for the audit trail. This is an exact copy of the data tables, which additionally contain the data columns "user", "user action", and "timestamp". All changes ("insert", "update", and "delete") in the data tables are entered into these “shadow” tables in the audit trail via insert trigger. Thus, we receive an audit trail of all changes per user as well as a differentiated database backup, which allows us to recover only special data/changes per user action.

2.5.3.2 Services Auditory

The data can be viewed by dashboard and can be created and viewed via the JobSchedulerDashboard. The jobs run regularly (1x per minute). The success or failure of the run is logged as a status message. Further the individual steps of the jobs are additionally logged.

2.5.4 Privacy measurements

2.5.4.1 Data Privacy Impact Assessment (DPIA)

A DPIA has been performed and is available by Charité - Universitätsmedizin Berlin.

2.5.4.2 Legal/Ethical process

All legal and ethical processes have been adhered to:

- Votum of Ethic commission of Cahrité
- EU-DSGVO
- State hospital law of Berlin

2.5.4.3 Processes for complying with the current legislation

A periodic review is carried out following a data protection management process. There is a common steering body.
It is ensured that personal data are processed in accordance with the instructions of the client.

- Powers of instruction are defined
- On-site checks by the client are possible
- The data protection agreement is drawn up in accordance with the requirements of Section 11 of the EU GDPR
- Sampling is carried out
- Control rights are defined
- Consent of client is obtained before inclusion into telemedicine concept
- Consent can be withdrawn at any time by client within the EASY-app
- Patient data must be stored at Charité for at least 10 years, but in the comjoodoc and medworx.io the data are deleted after a month regularly and if wished by the patient

2.6 Trustworthy AI

2.6.1 technology/user adoption and establishing trust

As described above in broad detail, we stuck to the GDPR and let our concept be evaluated regarding the legal aspects (by lawyers specialised on patient rights) and ethical aspects (by the ethical commission of Charité).
Basically, the telemedicine team consists of nurses and physicians, who have an education on ethical theory and bring those principles to life in daily work life as well. The telemedicine team also designs the concept and is in close loop with patients so that their needs are heard and implemented.
The telemedicine team evaluates any “advice” given by the prediction model before changing the treatment of a patient.

2.6.2 ethical principles

- **Respect for human authority:** The system only stores patient data after giving signed consent. Patients can withdraw consent at any time without stating a reason. The data will be deleted from the system, however, as TBase is part of the hospital information system, data must be kept for at least 10 years by law.
- **Prevention of harm:** Before the concept aired, it proved to be safe (data protection impact assessment, see above). In addition, the concept was evaluated for data safety and data protection by an independent law firm specialised for patient rights.
- **Fairness:** The effort put into fairness is high, e.g. the app has been translated into several languages (German, English, Polish, Mandarin, Turkish, Vietnamese..). However, the app is available in English and German. In addition, for partially blind and blind patients the option to enter data by audio function has been prepared, but is not available currently.

2.6.3 key requirements

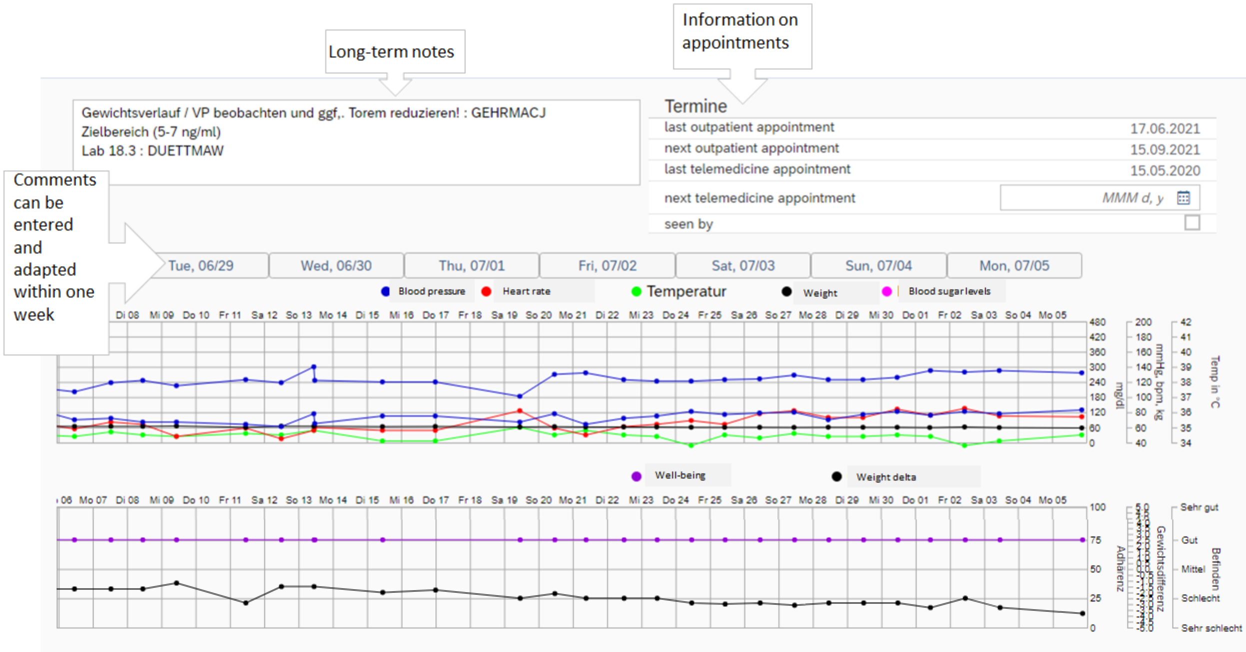
- **Human agency and oversight:** Explanations on why the prediction model came to a certain solution are provided along with the result of the prediction model, e.g. which laboratory values or which signs and symptoms decides to flag a patient as critical
- **Technical robustness and safety:** Before the prediction models are used, the usefulness is evaluated in the frame of studies. The first study compares the assessments of physicians and the prediction models themselves with each other. In a second step, the prediction models should be implemented into TBase, which is kept for future work.. In case a patient will be flagged with a particular risk, the model explains why and the telemedicine team watches the scene (observational study design). In the end, the sensitivity, specificity, positive and negative predictive values can be assessed, which makes statements on the usefulness of the prediction models.
- **Privacy and data governance:** In a long lasting process the data were cleansed and standardised (more information above). The primary process, where the computer scientist provides an overview of the type of data, interdisciplinary discussions were led, so that the computer scientist gained deep understanding.
- **Transparency:** Transparency is ensured at any time.
- **Diversity, non-discrimination and fairness:** As humans play an important role in evaluating the prediction models and the explainability of decisions (telemedicine team), vulnerable and marginalized groups were protected at any time.
- **Societal and environmental well-being:** The telemedicine concept saves resources as double laboratory testing is reduced and paper can be saved.
- **Accountability:** Currently, no therapeutic decision is made upon AI constipated prediction models. As explained above in detail in a first step they are evaluated in the frame of clinical studies before implemented into life systems.

2.7 System-Interaction

2.7.1 Human-Machine Interface / GUI

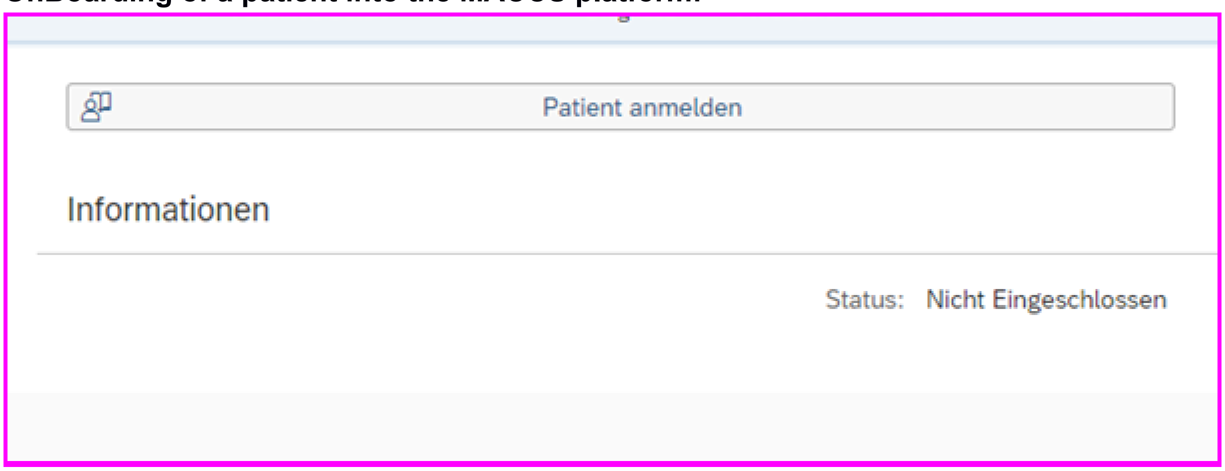
The interface for the hospital doctor is part of the TBase and screenshots emphasise the high usability. The EASY app was improved over time with help of users (patients) and now has a highly intuitive design.

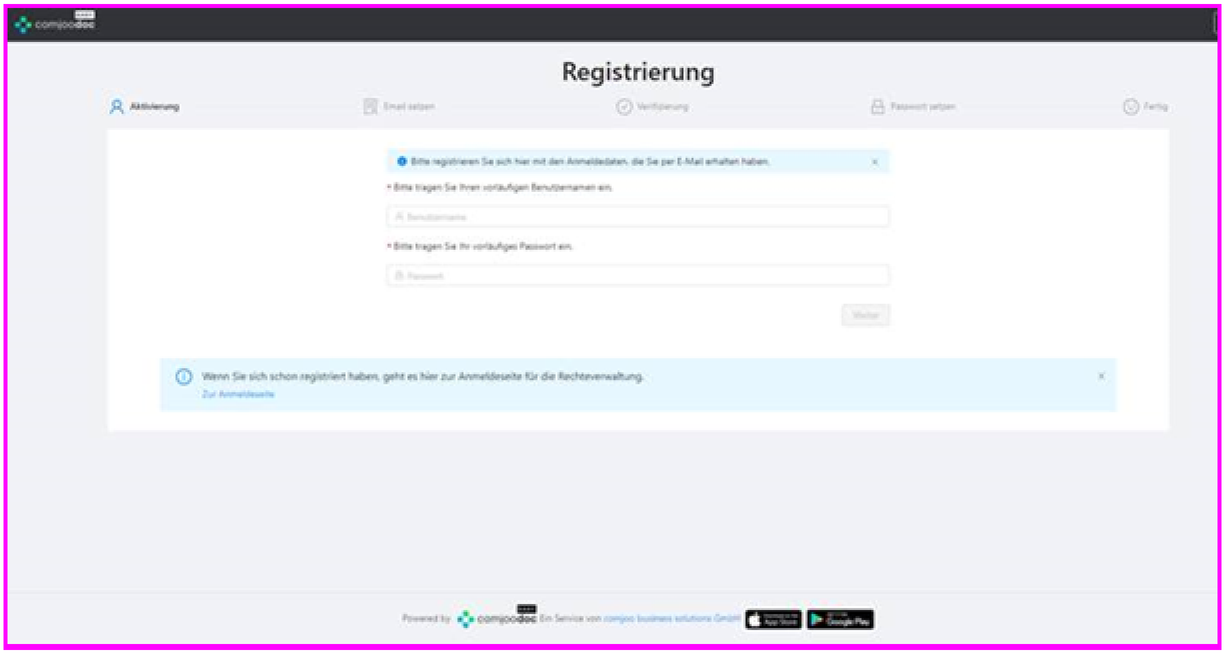
Screenshots User Interface TBase:



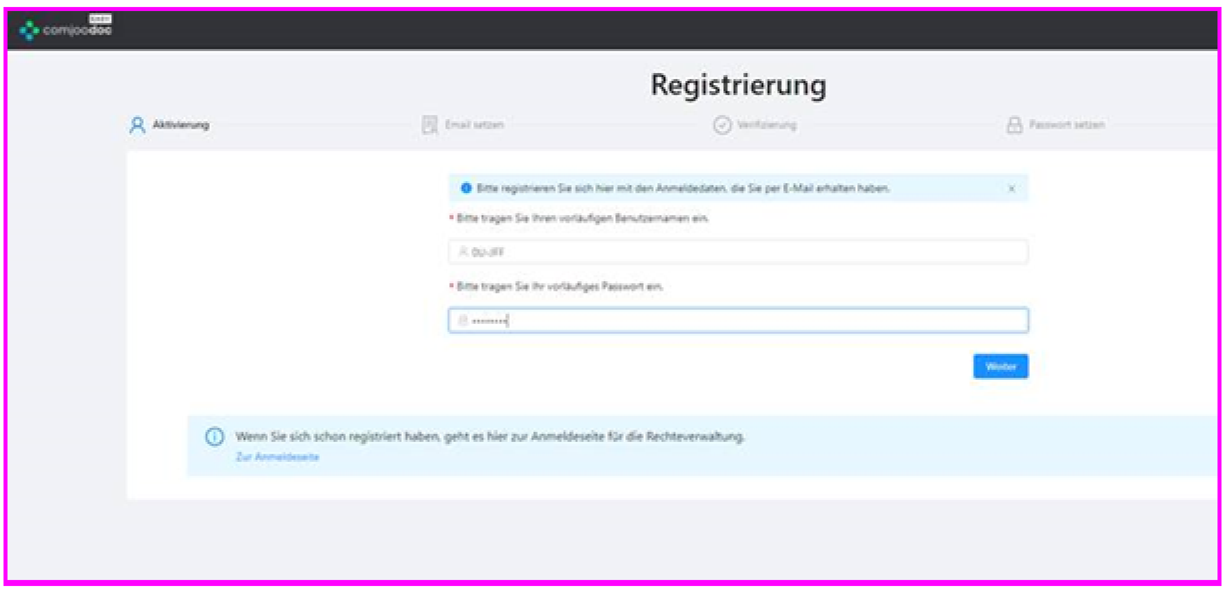
Telemedicine dashboard with tooltip guidelines explaining functionalities

OnBoarding of a patient into the MACCS platform:

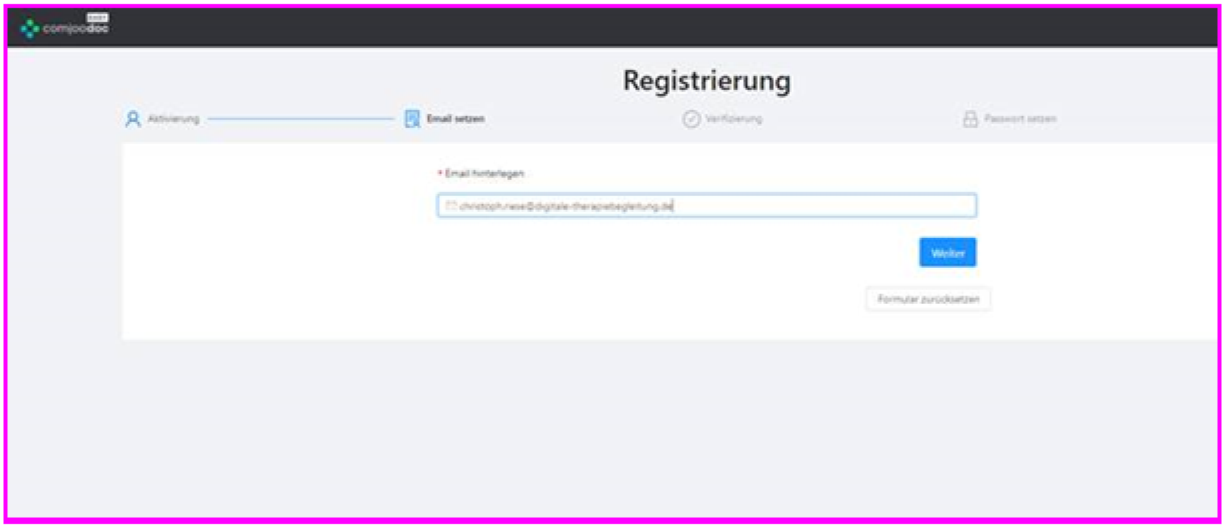


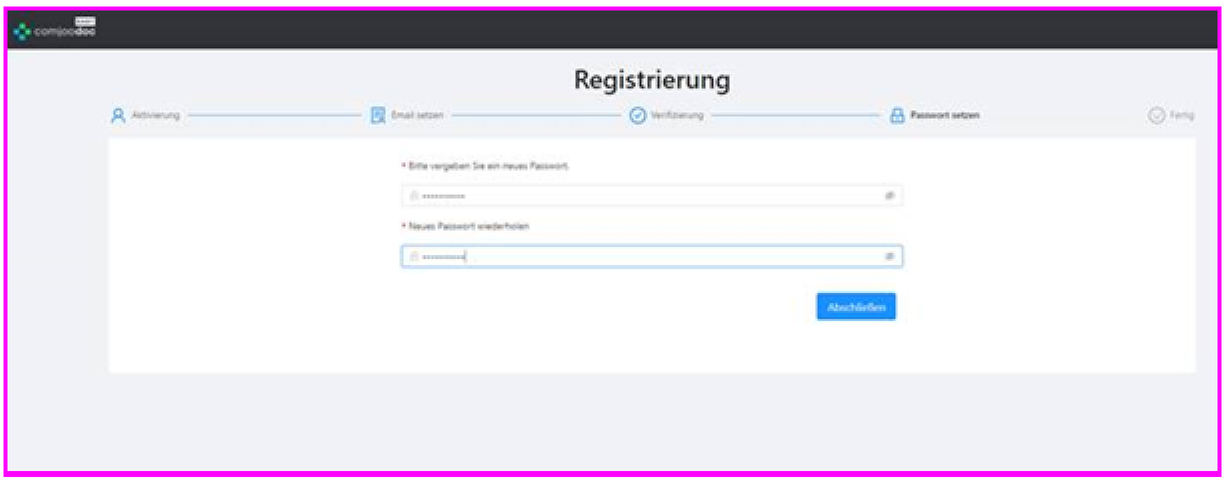


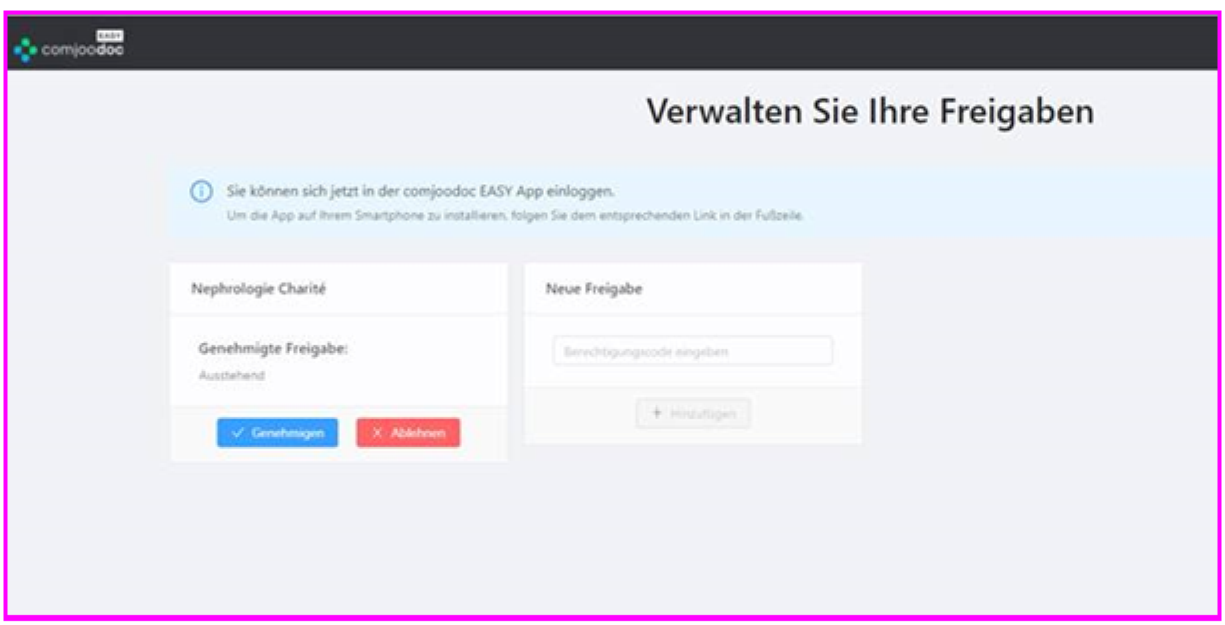
Registrierung
Aktivierung
Email setzen
Verifizierung
Passwort setzen
Fertig
Bitte registrieren Sie sich hier mit den Anmeldedaten, die Sie per E-Mail erhalten haben.
* Bitte tragen Sie Ihren vorläufigen Benutzernamen ein.
* Bitte tragen Sie Ihr vorläufiges Passwort ein.
Wenn Sie sich schon registriert haben, geht es hier zur Anmeldeseite für die Rechteverwaltung.
Zur Anmeldeseite



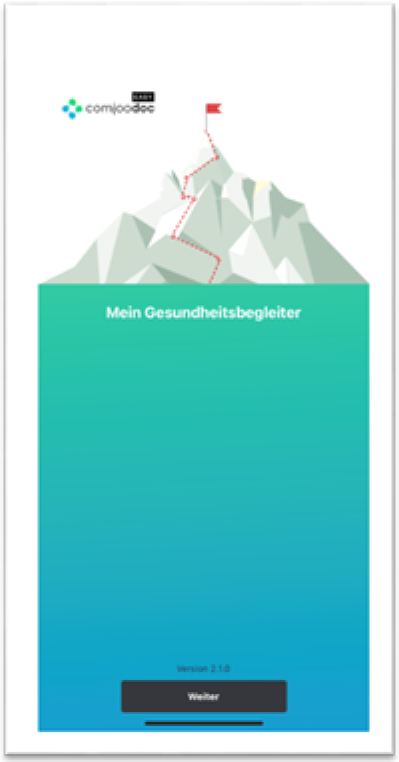
Registrierung
Aktivierung
Email setzen
Verifizierung
Passwort setzen
Bitte registrieren Sie sich hier mit den Anmeldedaten, die Sie per E-Mail erhalten haben.
* Bitte tragen Sie Ihren vorläufigen Benutzernamen ein.
* Bitte tragen Sie Ihr vorläufiges Passwort ein.
Weiter
Wenn Sie sich schon registriert haben, geht es hier zur Anmeldeseite für die Rechteverwaltung.
Zur Anmeldeseite

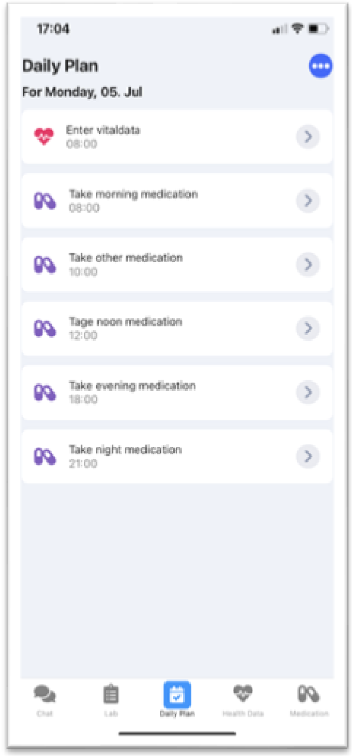

comjoodoc EASY
Registrierung
Aktivierung
Email setzen
Verifizierung
Passwort setzen
* Email hinterlegen
Weiter
Formular zurücksetzen

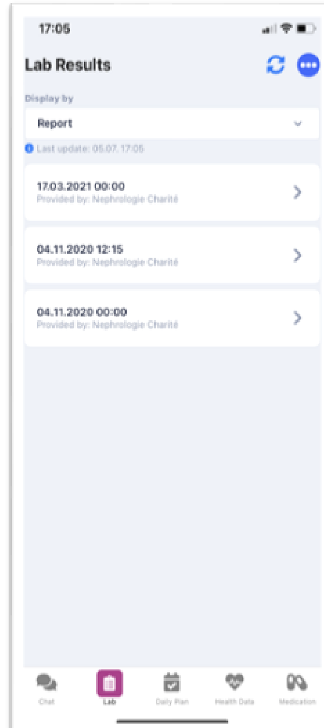

comjoodoc EASY
Registrierung
Aktivierung
Email setzen
Verifizierung
Passwort setzen
Fertig
* Bitte vergeben Sie ein neues Passwort
* Neues Passwort wiederholen
Abschließen

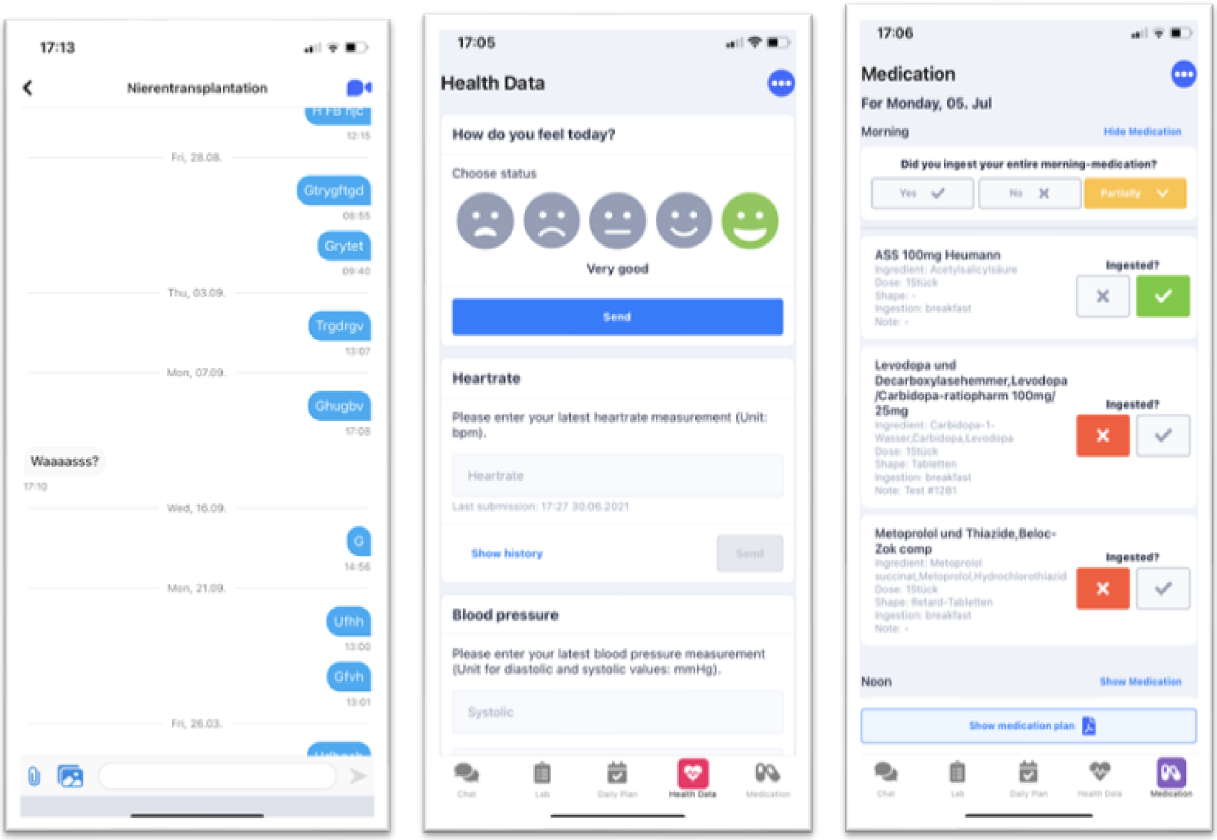

comjoodoc EASY
Verwalten Sie Ihre Freigaben
Sie können sich jetzt in der comjoodoc EASY App einloggen.
Um die App auf Ihrem Smartphone zu installieren, folgen Sie dem entsprechenden Link in der Fußzeile.
Nephrologie Charité
Genehmigte Freigabe:
Ausstehend
Genehmigen
Ablehnen
Neue Freigabe
Berechtigungscode eingeben
Hinzufügen

Screenshots User Interface EASY-app:









Screenshots of the user interface of the patient app (EASY-app).

2.7.2 Education

TBase is designed to be usable so that users do not need to be trained in particular. Once per month an update is released that contains bug fixes and new features, which are tested thoroughly before releasing. As an introduction an email to all TBase users is sent explaining the new functions and how they work.

Patients that after the kidney transplantation agree on using the EASY-app, were trained intensively by the telemedicine nurse, and if needed the training will be repeated.

3. Learnings

3.1 Challenges & Barriers

- **Architecture.** To build up the architecture of the digital infrastructure at our medical unit, time and resources (meaningly manpower and budget) were needed. Also, the bureaucracy of the hospital and federal state, especially regarding data protection, slowed down the process. The fortune was that the digital infrastructure could build on the results of a previous project, which objected to the data protection part mainly. The effort to put into data protection of a telemedicine concept can not be underlined enough and should be supported by a lawyer and ethics commission as well as a data protection unit.

- **Commitment of the patients.** One of the most important key challenges is to establish the commitment of the patients with the adherence process.
- In our setup we have followed up the interest of the kidney patients that were using the telemedicine application. The preliminary results showed that in one year of monitoring the interest in the telemedicine approach was varying in time. For instance, in the first quadrant in the first year there was a hype in the use of the application, followed by a small decrease. Then in the second and third quadrant there was again an increase of interest but at the end of the year the number of inputs decreased by 20%. It is important to point out that the technology is far from being mature. The patients are adapting to the new application and the results from the analysis are not yet sent back to the patients, motivating them to keep working with the telemedicine approach.
- **Bad data quality and missing data.** Even in the healthcare sector in which high standards are applied to keep data quality high, errors can happen. The reasons for these errors are manifold. For example, data is entered together with the unit, although the interface only requires the value, typos may result in wrong entries, or data is missing because a sensor measuring vital signs got disconnected. For the processing of the data this creates challenges. So, either the data needs to be pre-processed in order to correct data or filter out faulty data. In a streaming data context this is even more challenging if raw data is processed, because these data cleansing processes need to be applied online that means as part of the data analysis process and in a streaming context usually only part of the stream of data is visible at a time.

- **Natural Language Processing**
    - If you want to process clinical text, hardly any NLP tools and dataset are available. This is especially true when we process texts that are not written in English.
    - Access to clinical data is difficult. Training NLP tools is usually based on large amounts of labelled data. However, access to clinical data is for research partners often only possible as part of a concrete research project with a hospital. Collecting data from multiple hospitals and departments is nearly impossible due to different regulations (hospital, department, state, federal state).

Labeling of this data is also challenging, since it often requires not only a NLP understanding but also medical domain knowledge.

- **Prediction Algorithms**
    - **Cohort selection.** The modelling of the cohort got more and more complex over time. Discussing design decisions with peers and particularly with medics, in order to get some feedback (also sanity checks) is quite difficult, as the data is complex.
    - Modelling the cohort bear many challenges:
        - TBase has grown historically, therefore many particularities about how the data is modelled or why something is missing etc. cannot be easily understood without frequent discussions with the people building and maintaining the DB. (-> data after death, visit entries without visit)
        - It probably makes sense that a computer scientist who is going to apply machine learning later on, creates the cohort. This person however, does not necessarily have the medical expertise to understand everything. So it can happen that things are modelled in a wrong way. Errors are difficult to find, as data is very complex, and particular things have to be discussed with medics, who first need to understand what you are doing. For instance for our use case, ICD10 codes were not a good choice to model the endpoint, as they a) did not properly describe the endpoint according to our definition, and b) they were partially used infrequently in older parts of the data.
    - **Comparability**: When developing machine learning models it is normally very useful to have a) training & test data already available and a b) reference/baseline system for that data. In this case you can develop your own system with some improvements, just insert the data and directly compare to the baseline. If the results are below the baseline, you can do some engineering, optimization or just change the architecture. However, as clinical data is not publicly available the aforementioned setup normally doesn't work. For this reason the very first step is to use a well-established method which has been successful on a similar task and re-implement it on your data. The main problem here is: When is your system, your baseline good enough? What is good enough? Something similar as reported on a similar task on different data? As data is different, so can be your results. In some cases we experience in papers, that people re-implement a system as a baseline, and then introduce their “new” method. However, an improvement can be also achieved, as you put more effort into your new method rather than the baseline. Also, achieving similar results as reported in other papers can be difficult as a) data differs, and b) people might not report each single mini step of their implementation, but at the same time something not so obvious might have led to significant improvements.
    - **Different languages, different worlds**: Interdisciplinary work between at least computer scientists and medical staff (and possibly others) is necessary to successfully develop machine learning models in medicine. However, in each

domain the (scientific) working style might differ significantly. Besides the terminology, the background and the understanding differs. In order to be successful, frequent meetings are necessary, and so is the understanding of the other domain. Being able to explain complex topics in a very simple way is a very important skill. If the computer scientist is not able to explain what he is doing, the medical colleagues are not able to support him/her in the best possible way - or problems are found later.

- **To carry out a study in healthcare**: Carrying out a study in medicine with test subjects often differs from studies in computer science (machine learning) where you ideally "just" use your data and insert it into your model. At least in this way, studies can be carried out quickly - while studies including human subjects can be very time consuming: Before a study can start, an ethics commission must give permission. To avoid statistical biases, a power calculation must be performed and, to reach appropriate numbers of cases, some studies. Also a manual evaluation by human subjects is time consuming.

- **Security and privacy of data access and processing:** This topic combines many aspects and is quite interdisciplinary. A sketch of the concept should be discussed with all professions over and over again. The implementation of all programm parts lasts long and should be evaluated regularly. An external evaluation is very expensive, which is not easy to find in underfunded systems such as the Health Care System.

- **Trustworthy AI:** To find the key information on what medical professionals listen to, the programmer has to implement these into the prediction model and, if an aspect is the reason to flag a patient as being at risk for a complication of disease, it should be visible for the medical expert.

3.2 Lessons Learned

- **Architecture:**

  First of all, if one plans to work with sensitive patient data, the data protection unit should be contacted and the concept should be highlighted in every detail. Only after the planned digital architecture is secure, interfaces can be activated. In Germany, a contract including clear rules for data processors should be signed. The data should be stored in Germany at best or in Europe. No services from Chinese or American providers should be used, especially not to monitor health activities via app. We recommend always rely on standardization (e.g. HL7 FHIR standard, FHIR server, SNOMED CT, LOINC, PROM) and interdisciplinary teams, who work together at eye level.

- **Processing of large structured / unstructured data sources**

  As the medical data "grew" over time and mostly without standardization, this should be the groundwork in any big data project: standardization of data. If one plans on working in the field of medical data, one has to understand that medical experts often enter information without taking classes before. Even diagnosis or procedures have not been standardized yet, although attempts are on its way (LOINC, SNOMED CT).

Another advice is to not expect whole datasets and that all data always have to be complete for medical doctors to draw a conclusion. So if values are not there, it does not automatically mean that the dataset is incomplete and must be deleted for the better good. The intensive cooperation with doctors is the key to understanding and accomplishing the processing of large and unstructured data.

- **Complex real-time event detection**

Complex event processing (CEP) is a technology for aggregating, processing, and analyzing massive streams of data in order to gain real-time insights. CEP can, for example, be used as a technique to constantly monitor data streams in order to identify critical health conditions of patients as early as possible. Defining patterns that the stream of data is monitored against, to create warning or alarm events, require certain domain knowledge, such as normal ranges of bloodpressure depending on age and gender. So developing such systems make a close collaboration between data science and medical team necessary.

- **Natural Language Processing:**
    - Make models available, so the community can benefit from it
    - Character-based approaches might be beneficial for clinical text, in comparison to word based approaches, as you might face a lot of typos and domain specific words.
    - Domain-adaptation can help -> pretrain on large other data, and then optimize on small in domain data

- **Prediction Algorithms:**
    - Imputation did not lead to improvements - we tried different techniques and it seemed to be a waste of time for our use case
    - **Comparability:** The development of publicly available datasets/cohorts, such as MIMIC-III, can make your life much easier.
    - **ROC score:** Even though the ROC score appears to be commonly used and very popular in the medical domain, we learnt that it is a very unreliable score if positive and negative examples are strongly unbalanced. We recommend AUC-PR (precision/recall) instead.
- **Security and privacy of data access and processing:** The requirements of medical experts (nurse, physician at hospital or in private practice, patients) should be ascertained at first and combined with the already established infrastructure. The existing infrastructure must be analysed and reassessed with new features. The datastreams should be sketched and let be evaluated by data protection experts from the beginning. Including their advice into the concept and then starting to program any software. In addition, the interdisciplinarity of the team is highly important. We can recommend agile programming. Our team consisted of nurses, physicians in various positions, computer scientists and medical informatics. An independent expert for patient data issues is expensive though, but a help to find pitfalls and is not professionally blinkered.

- **Trustworthy AI:** Before using prediction models for a wide range of patients, they must be evaluated in the frame of clinical studies. In these studies, medical experts should be in charge, because they are designed differently than informatic trials. The usefulness and explainability should be proved by medical experts.

3.3 Main (quantifiable) achievements

1. Prediction models regarding rejection, graft failure, and infection in patients after kidney transplantation, model evaluated on retrospective data and within a study with medical doctors
2. Telemedicine dashboard, directly implemented into the EMR TBase
3. Digital “Fever chart”, also directly implemented into the EMR TBase
4. Laboratory values, text messages, and updated medication plans can be pushed into an app
5. Patient-app: Design considering UX that enables to provide home-measured vital signs and send text messages which are presented in the telemedicine dashboard
6. Data are provided by HL7 FHIR standard
7. Data are stored on a FHIR server, located at Charité
8. Digital infrastructure evaluated for data safety and data protection
9. Contract with insurance companies in order to refund expenditures (personnel, development of concept)
10. Development of NLP models to process German clinical text

4. Output

4.1 Papers

[1] Roland Roller, Laura Seiffe, Ammer Ayach, Sebastian Möller, Oliver Marten, Michael Mikhailov, Christoph Alt, Danilo Schmidt, Fabian Halleck, Marcel Naik, Wiebke Düttmann and Klemens Budde. Information Extraction Models for German Clinical Text. In 2020 IEEE International Conference on Healthcare Informatics (ICHI). Oldenburg, 2020.

[2] Nils Rethmeier, Necip Oguz Serbetci, Sebastian Möller and Roland Roller. EffiCare: Better Prognostic Models via Resource-Efficient Health Embeddings In AMIA Annual Symposium Proceedings. American Medical Informatics Association, 2020.

[3] Wiebke Düttmann, Marcel G Naik, Bianca Zukunft, Bilgin Osmonodja, Friederike Bachmann, Mira Choi, Roland Roller, Manuel Mayrdorfer, Fabian Halleck, Danilo Schmidt, Klemens Budde. eHealth in Transplantation In Transplant International: Official Journal of the European Society for Organ Transplantation. 2020.

[4] Laura Seiffe, Oliver Marten, Michael Mikhailov, Sven Schmeier, Sebastian Möller and Roland Roller. From Witch's Shot to Music Making Bones - Resources for Medical Laymen to Technical Language and Vice Versa In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, 2020

[5] Roland Roller, Gaurav Vashisth, Philippe Thomas, He Wang, Michael Mikhailov and Mark Stevenson. Graph-KD: Exploring Relational Information for Knowledge Discovery In Proceedings of the ISWC 2019 Posters & Demonstrations, Auckland, New Zealand, 2019

[6] Nils Rethmeier, Barbara Plank. MoRTy: Unsupervised Learning of Task-specialized Word Embeddings by Autoencoding. In: Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019). Association for Computational Linguistics 7/2019.

4.2 Open Source & Resources (refer to ELG)

- The models to process German clinical text are made publicly available and can also be applied to other biomedical texts outside the nephrology domain.
- TBase of Charité will be made open access perspectively by publishing the source code.

4.3 Demos

Demo video of the home-measured data system
Version including photos (in German):
https://vimeo.com/tidefilm/download/515512959/421b719083